ZESZYTY
NAUKOWE
POLITECHNIKI
ŚLĄSKIEJ

TADEUSZ CZACHÓRSKI

PROCES POWROTÓW ELEMENTARNYCH W DOSKONALENIU APROKSYMACJI DYFUZYJNEJ JAKO METODY MODELOWANIA PRACY SYSTEMÓW KOMPUTEROWYCH

P. 4193/87

INFORMATYKA

Z. 9
GLIWICE
1987

POLITECHNIKA ŚLĄSKA

ZESZYTY NAUKOWE

Nr 1005

TADEUSZ CZACHÓRSKI

P. 4193/87

PROCES POWROTÓW ELEMENTARNYCH W DOSKONALENIU APROKSYMACJI DYFUZYJNEJ JAKO METODY MODELOWANIA PRACY SYSTEMÓW KOMPUTEROWYCH

GLIWICE 1987

OPINIODAWCY
Prof. dr hab. inż. Jan Węglarz
Doc. dr hab. inż. Adam Wolisz

KOLEGIUM REDAKCYJNE
REDAKTOR NACZELNY — Prof. dr hab. inż. Wiesław Gabzdyl
REDAKTOR DZIAŁU — Dr inż. Stanisław Kozielski
SEKRETARZ REDAKCJI — Mgr Elżbieta Stinzing
CZŁONKOWIE KOLEGIUM — Prof. dr hab. inż. Adolf Maciejny
— Prof. dr inż. Stanisław Malzacher
— Prof. dr hab. inż. Bronisław Skinderowicz

OPRACOWANIE REDAKCYJNE
Alicja Nowacka

Wydano za zgodą
Rektora Politechniki Śląskiej

PL ISSN 0208-7286

Dział Wydawnictw Politechniki Śląskiej
ul. Kujawska 3, 44-100 Gliwice

Nakł. 160+85 Ark. wyd. 12,25 Ark. druk. 10,125 Papier offset. kl III 70x100, 70 g
Oddano do druku 5.02 87 Podpis. do druku 29.06.87 Druk ukończ. we wrześniu 1987
Zam. 116/87 L-24 Cena zł 245.—

Skład, fotokopie, druk i oprawę
wykonano w Zakładzie Graficznym Politechniki Śląskiej w Gliwicach

P6/88

SPIS TREŚCI

Str.

WPROWADZENIE 9

Rozdział I. ANALITYCZNE I OBLICZENIOWE PROBLEMY PROBABILISTYCZNYCH MODELI SYSTEMÓW KOMPUTEROWYCH 15

1.1. Sformułowanie problemu i notacja 15
1.2. Łańcuchy Markowa jako podstawa modelowania sieci stanowisk obsługi 18
1.3. Obliczanie dokładnych i przybliżonych rozwiązań dla sieci dekomponowalnych 24
1.4. Rozwiązania numeryczne niedekomponowalnych sieci markowskich 29
1.5. Modele wykorzystujące przybliżoną dekompozycję sieci 32
1.6. Podsumowanie 37

Rozdział II. APROKSYMACJA DYFUZYJNA - POJĘCIA PODSTAWOWE 39

2.1. Aproksymowane procesy N(t), U(t) 39
2.2. Aproksymujący proces X(t) 41
2.3. Twierdzenia graniczne dla stanowiska G/G/1 44
2.4. Warunki brzegowe dla procesu dyfuzji 46

Rozdział III. APROKSYMACJA DYFUZYJNA LICZBY KLIENTÓW W POJEDYNCZYM STANOWISKU OBSŁUGI 49

3.1. Aproksymacja dyfuzyjna stanowisk G/G/1/N i G/G/1 - stan ustalony 50
3.2. Czas pierwszego przejścia w procesie dyfuzji i okres czynny stanowiska G/G/1 52
3.3. Stan nieustalony i dynamika stanowisk G/G/1/N i G/G/1 60

Rozdział IV. APROKSYMACJA DYFUZYJNA WIRTUALNEGO CZASU CZEKANIA W STANOWISKU G/G/1 76

4.1. Rezultaty dokładne dotyczące procesu U(t) 77
4.2. Aproksymacja Gavera 78
4.3. Aproksymacja procesem dyfuzji z powrotami elementarnymi - stan ustalony 80
4.4. Aproksymacja U(t) z uwzględnieniem trzeciego momentu rozkładu czasu obsługi - stan ustalony 85
4.5. Aproksymacja stanów nieustalonych i dynamiki zmian średniego czasu czekania 94

Str.

Rozdział V. STANOWISKO OBSŁUGI ZE SKOŃCZONĄ POPULACJĄ KLIENTÓW - APROKSYMACJA DYFUZYJNA MODELU KONSERWATORA 110

5.1. Analityczny opis stanów nieustalonych i ustalonych w modelu konserwatora 111

5.2. Aproksymacja dyfuzyjna podstawowego modelu konserwatora - stan ustalony 116

5.3. Wyniki numeryczne 120

5.4. Model z wieloma klasami klientów 126

5.5. Aproksymacja stanu nieustalonego 136

UWAGI KOŃCOWE 140

LITERATURA 142

STRESZCZENIA 157

СОДЕРЖАНИЕ

Стр.

Введение 9

Глава I. АНАЛИТИЧЕСКИЕ И РАСЧЁТНЫЕ ПРОБЛЕМЫ ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ КОМПЬЮТЕРНЫХ СИСТЕМ 15

1.1. Формулировка проблемы и обозначения 15

1.2. Цепи Маркова как основа моделирования сети обслуживающих приборов 18

1.3. Вычисление точных и приближённых решений для декомпонированных сетей 24

1.4. Машинные решения для недекомпонированных марковских сетей 29

1.5. Модели использующие приближённую декомпозицию сети 32

1.6. Подведение итогов 37

Глава II. ДИФФУЗИОННАЯ АППРОКСИМАЦИЯ - ОСНОВНЫЕ ПОНЯТИЯ 39

2.1. Аппроксимированные процессы N(t), U(t) 39

2.2. Аппроксимирующий процесс X(t) 41

2.3. Передельные теоремы для обслуживающих приборов типа G/G/1 44

2.4. Граничные условия для процесса диффузии 46

Глава III. ДИФФУЗИОННАЯ АППРОКСИМАЦИЯ ЧИСЛА КЛИЕНТОВ ДЛЯ ОДНОГО ОБСЛУЖИВАЮЩЕГО ПРИБОРА 49

3.1. Диффузионная аппроксимация приборов типа G/G/1/N и G/G/1 - установившееся состояние 50

3.2. Время первого прохода в процесс диффузии и активный период прибора типа G/G/1 52

3.3. Неустановившееся состояние и динамика приборов типа G/G/1/N и G/G/1 60

Глава IV. ДИФФУЗИОННАЯ АППРОКСИМАЦИЯ ВИРТУАЛЬНОГО ВРЕМЕНИ ОЖИДАНИЯ ДЛЯ ОДИНОЧНОГО ОБСЛУЖИВАЮЩЕГО ПРИБОРА 76

4.1. Точные результаты для процесса U(t) 77

4.2. Аппроксимация Гавера 78

4.3. Диффузионная аппроксимация с элементарными возвращениями - установившееся состояние 80

4.4. Аппроксимация U(t) с учётом третьего момента распределения времени обслуживания - установившееся состояние 85

4.5. Аппроксимация неустановившихся состояний и динамики изменения среднего времени ожидания 94

Стр.

Глава V. ОБСЛУЖИВАЮЩИЙ ПРИБОР С КОНЕЧНЫМ МНОЖЕСТВОМ КЛИЕНТОВ – ДИФФУЗИОННАЯ АППРОКСИМАЦИЯ МОДЕЛИ КОНСЕРВАТОРА 110

5.1. Аналитическое описание неустановившихся и установившихся состояний в модели консерватора 111

5.2. Диффузионная аппроксимация основной модели консерватора – установившееся состояние 116

5.3. Машинные результаты 120

5.4. Модель с многими классами клиентов 126

5.5. Аппроксимация неустановившегося состояния 136

ВЫВОДЫ 140

ЛИТЕРАТУРА 142

РЕЗЮМЕ 157

CONTENTS

Page

INTRODUCTION 9

Chapter I. ANALYTICAL AND NUMERICAL PROBLEMS OF PROBABILISTIC MODELS OF COMPUTER SYSTEMS 15

1.1. Problem formulation and notation 15

1.2. Markov chains in modelling queueing networks 18

1.3. Exact and approximate solutions of decompoeable networks 24

1.4. Numerical solutions of nondecomposable Markovian networks 29

1.5. Models with an approximate network decomposition 32

1.6. Recapitulation 37

Chapter II. DIFFUSION APPROXIMATION - BASIC NOTIONS 39

2.1. Approximated processes N(t), U(t) 39

2.2. An approximating process X(t) 41

2.3. Limit theorems for G/G/1 station 44

2.4. Boundary conditions for a diffusion process 46

Chapter III. DIFFUSION APPROXIMATION OF THE CUSTOMER NUMBER IN A SINGLE SERVICE SYSTEM 49

3.1. Diffusion approximation of the stations G/G/1/N and G/G/1 - steady state solution 50

3.2. The first passage time of diffusion process and busy period of G/G/1 station 52

3.3. A transient state and dynamics of G/G/1/N and G/G/1 stations 60

Chapter IV. DIFFUSION APPROXIMATION OF VIRTUAL WAITING TIME IN A SINGLE SERVICE STATION 76

4.1. Exact results concerning the process U(t) 77

4.2. Gaver's approximation 78

4.3. Diffusion approximation with an elementary return process - steady state solution 80

4.4. A third-order approximation of U(t) - a steady state solution 85

4.5. Transient state approximation and dynamics of the mean waiting time changes 94

Page

Chapter V. SERVICE STATION WITH A FINITE CUSTOMER POPULATION - DIFFUSION APPROXIMATION OF THE REPAIRMAN MODEL 110

5.1. An analytical description of transient and steady states of the repairman model 111

5.2. A diffusion approximation of the basic repairman model - steady state solution 116

5.3. A multiclass model 126

5.4. A transient state approximation 136

FINAL REMARKS 140

LITERATURE 142

SUMMARY 157

WPROWADZENIE

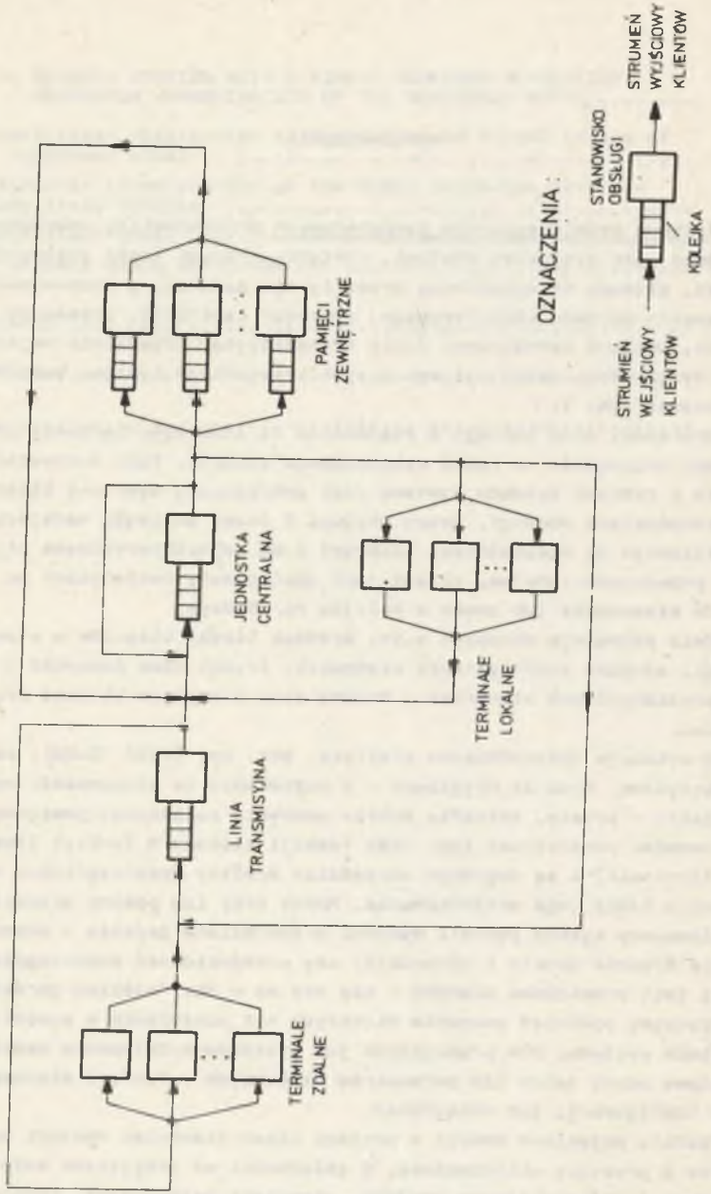
Kolejkowe modele systemów komputerowych przedstawiają modelowany system jako sieć stanowisk obsługi, pomiędzy którymi krążą obsługiwani klienci. Klienci to wykonywane przez system zadania, a stanowiska obsługi to elementy systemu komputerowego: procesor centralny, procesory komunikacyjne, pamięci zewnętrzne, linie transmisyjne, urządzenia wejścia-wyjścia. Przykładowy model ujętego w tych kategoriach systemu komputerowego przedstawia rys. 1.

Jednorazowy czas obsługi w stanowisku to czas nieprzerwanej pracy modelowanego urządzania na rzecz wykonywanego zadania. Fakt korzystania przez zadanie z różnych zasobów systemu jest zobrazowany wędrówką klienta pomiędzy stanowiskami obsługi. Czasy obsługi i czasy pomiędzy nadejściem kolejnych klientów są wielkościami losowymi i są scharakteryzowane przez rozkłady prawdopodobieństwa. Klient jest obsługiwany natychmiast po nadejściu do stanowiska lub czeka w kolejce na obsługę.

Modele pozwalają obliczyć m.in. średnią liczbę klientów w stanowiskach obsługi, stopień wykorzystania stanowisk, średni czas czekania w kolejkach poszczególnych stanowisk i średni czas przejścia klienta przez całą ich sieć.

Jak wykazuje dotychczasowa praktyka, por. np. [PERF 72-84], modele probabilistyczne, mimo iż wyjątkowo - w porównaniu ze złożonością opisywanego obiektu - proste, potrafią dobrze uchwycić zależności pomiędzy jego podstawowymi parametrami (np. czas reakcji systemu w funkcji liczby aktywnych terminali) i są dogodnym narzędziem analizy poszczególnych wariantów systemu w fazie jego projektowania. Można przy ich pomocy przewidzieć, czy planowany system potrafi wykonać przewidziane zadania w odpowiednio krótkim średnim czasie i sprawdzić, czy przepustowość poszczególnych jego części jest prawidłowo dobrana - czy nie ma w nim "wąskich gardeł", które są przyczyną opóźnień znacznie większych niż opóźnienia w pozostałych częściach systemu. Dla pracujących już systemów modelowanie umożliwia ilościową ocenę zmian ich parametrów użytkowych w funkcji planowanych zmian konfiguracji lub obciążenia.

Prostota pojęciowa modeli w postaci sieci stanowisk obsługi nie idzie w parze z prostotą obliczeniową. W zależności od przyjętych założeń dotyczących rozkładów czasów obsługi, strumieni wejściowych, regulaminów szeregowania klientów - modele prowadzą do rozwiązania analitycznego, numerycznego lub wymagają zastosowania metod symulacyjnych. Dlatego też wiele uwagi poświęca się numerycznej implementacji modeli, tworząc pakie-



Rys. 1. Przykładowy model systemu komputerowego
Fig. 1. An exemplary model of a computer system

ty odpowiednich programów, np. [BEST 78, Sane 82, Whit 83b,c, Vera 84] oraz opracowując metody przybliżone tam, gdzie brak dokładnego rozwiązania lub jest ono zbyt pracochłonne.

Aproksymacja dyfuzyjna jest jedną z przybliżonych metod modelowania. W mocno obciążonym stanowisku obsługi, tj. w stanowisku, którego współczynnik wykorzystania jest bliski jedności, odpowiednio unormowane procesy N(t) i U(t)

N(t) - liczba klientów, obsługiwanych lub czekających, obecnych w stanowisku w chwili t,

U(t) - zapas pracy, czas potrzebny do zakończenia obsługi wszystkich obecnych w stanowisku w chwili t zadań,

są - niezależnie od typu rozkładu odstępu czasu między klientami w strumieniu wejściowym i rozkładu czasu obsługi - słabo zbieżne do jednowymiarowego procesu dyfuzji z prądem, ograniczonego do dodatniej półosi liczb rzeczywistych i posiadającego jako warunek brzegowy umieszczoną w punkcie x = 0 barierę odbijającą. Pozwala to wykorzystać rezultaty dotyczące stosunkowo dobrze poznanego procesu dyfuzji, np. [Levy 48, Cox 65], do przybliżonego opisu zachowania stanowiska obsługi, w którym prawdopodobieństwo, iż brak w nim klientów i stanowisko pozostaje bezczynne, można pominąć.

Niech $f(x,t;x_0)$ oznacza gęstość procesu dyfuzji X(t)

$$f(x,t;x_0)dx = Pr[x < X(t) \leq x + dx \mid X(0) = x_0].$$

Postać $f(x,t;x_0)$ jest znana po rozwiązaniu równania dyfuzji

$$\frac{\partial f(x,t;x_0)}{\partial t} = \frac{\alpha}{2} \frac{\partial^2 f(x,t;x_0)}{\partial^2 x} - \beta \frac{\partial f(x,t;x_0)}{\partial x}$$

lub w stanie ustalonym równania

$$0 = \frac{\alpha}{2} \frac{d^2 f(x)}{dx^2} - \beta \frac{df(x)}{dx}$$

gdzie $f(x) = \lim_{t \to \infty} f(x,t;x_0)$.

W zależności od przyjętych współrzędnych procesu dyfuzji, gęstość $f(x,t;x_0)$ aproksymuje dyskretny rozkład prawdopodobieństwa liczby klientów w stanowisku lub gęstość procesu U(t).

Dobierając odpowiednio współczynniki β, α można opisywać stanowiska wielokanałowe, stanowiska z wieloma klasami klientów, z obsługą priorytetową, z grupowymi nadejściami klientów - a więc stanowiska, których analiza dokładna przy założonych ogólnych rozkładach czasu obsługi i czasu

między zgłoszeniami nie jest możliwa. Metodę tę można po obliczeniu dwóch pierwszych momentów rozkładu odstępów czasu między klientami w strumieniu wejściowym odnieść do każdego stanowiska w sieci.

Literatura dotycząca zastosowania aproksymacji dyfuzyjnej w teorii kolejek w ogóle i do modelowania systemów informatycznych w szczególności jest już obszerna - ważniejsze jej pozycje zostały wymienione w końcowej części rozdziału pierwszego.

Zakres aproksymacji dyfuzyjnej można rozszerzyć - zaproponował to Gelenbe [Gele 75] wprowadzając proces powrotów elementarnych jako warunek graniczny procesu dyfuzji: proces po osiągnięciu wartości zerowej zatrzymuje się przez czas odpowiadający jednorazowemu okresowi bezczynnemu stanowiska, po czym rozpoczyna się w wybranym punkcie na osi x. W ten sposób model opisuje zachowanie stanowiska słabiej obciążonego, w którym okresów bezczynności nie można pominąć.

Cel i zakres rozprawy

Celem rozprawy jest udoskonalenie metody aproksymacji dyfuzyjnej poprzez:

1) opracowanie metody rozwiązywania równań dyfuzji uzupełnionej o proces powrotów elementarnych (rozwiązanie przedstawione w [Gele 75] ogranicza się do stanu ustalonego, gdy równanie dyfuzji jest zwykłym, a nie cząstkowym równaniem różniczkowym),
2) sprawdzenie w kilku podstawowych w teorii kolejek zadaniach czy i w jakim stopniu wprowadzenie powrotów elementarnych zwiększa dokładność aproksymacji.

Rozdział I stanowi zwięzły monograficzny przegląd rezultatów dotyczących modeli w postaci sieci stanowisk obsługi. Rozdział II wprowadza podstawowe pojęcia związane z aproksymacją dyfuzyjną. Rozdziały III-V zawierają oryginalne wyniki pracy. W rozdziale III pokazano, że proces dyfuzji z powrotami elementarnymi można przedstawić jako superpozycję procesów dyfuzji z ekranem pochłaniającym (po dojściu do ekranu proces kończy się) i uzyskano tą drogą aproksymacje procesu $N(t)$ w stanowisku G/G/1, tj. w stanowisku z dowolnym strumieniem wejściowym klientów i dowolnym rozkładem czasu obsługi oraz w stanowisku G/G/1/N, tj. w stanowisku G/G/1 z ograniczoną do N klientów kolejką. W rozdziale IV wykorzystano tę samą metodę superpozycji procesów do aproksymacji $U(t)$ w stanowisku G/G/1. Zbadano też zasadność aproksymacji procesu $U(t)$ procesem Markowa wyższego rzędu niż proces dyfuzji. W rozdziale V stworzono dyfuzyjny model stanowiska obsługi z ograniczoną populacją klientów (stanowisko G/G/1//N, tzw. problem konserwatora), uwzględniając powroty elementarne oraz zmienne skokowo wraz z wartością procesu dyfuzji współczynniki β i α.

Trudno ocenić błąd aproksymacji dyfuzyjnej inaczej jak poprzez porównanie jej wyników z wynikami dokładnymi (gdy jest to możliwe) lub wynikami symulacji. Stąd występująca w pracy dość duża liczba przykładów obliczeniowych, umożliwiająca dyskusję błędów aproksymacji, w także pozwalająca stwierdzić, że obliczenia numeryczne związane z opracowanymi modelami są, wziąwszy pod uwagę omówione w początkowym rozdziale trudności innych metod modelowania, stosunkowo proste.

Praca została wykonana w Zakładzie Systemów Automatyki Kompleksowej PAN w ramach tematu "Rozwój statystycznych metod opisu działania, oceny wydajności i doboru struktury złożonych systemów komputerowych" w problemie węzłowym "Rozwój zintegrowanych systemów komputerowego sterowania z uwzględnieniem więzów międzysystemowych". Oprócz prac teoretycznych, temat ten obejmował przygotowanie pakietu programowego umożliwiającego wykonywanie obliczeń modelowych, stworzenie oprogramowania zbierającego dane statystyczne o pracy dużego systemu komputerowego czasu rzeczywistego, opracowanie probabilistycznych modeli dwóch konkretnych systemów komputerowych (systemu sterującego procesem przemysłowym oraz wielodostępnego systemu zarządzającego) - obejmował więc wszystkie podstawowe zagadnienia związane z opracowaniem i stosowaniem probabilistycznych modeli systemów komputerowych [ZSAK 81-85, Czac 85b]. Uzyskane w ten sposób własne doświadczenia w pełni potwierdzają, zdaniem autora, przydatność probabilistycznych modeli przy projektowaniu i rozbudowie systemów komputerowych - i uzasadniają wysiłki zmierzające do doskonalenia teorii tych modeli.

Większość zebranych w pracy w jednolitej formie wyników była opublikowana wcześniej w kwartalniku "Podstawy Sterowania".

Autor dziękuje Profesorowi Stefanowi Węgrzynowi za umożliwienie realizacji niniejszej pracy, za pomoc i zachętę oraz Kolegom z Zespołu Statystycznych Modeli Systemów Komputerowych za dyskusję i uwagi.

R o z d z i a ł I

ANALITYCZNE I OBLICZENIOWE PROBLEMY PROBABILISTYCZNYCH MODELI SYSTEMÓW KOMPUTEROWYCH

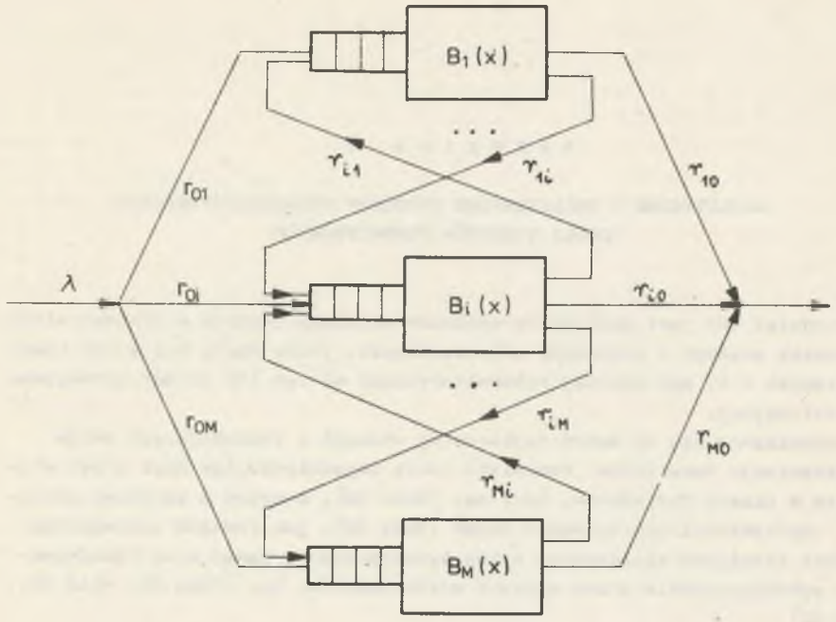
Rozdział ten jest poświęcony ogólnemu omówieniu modeli w postaci sieci stanowisk obsługi i problemów obliczeniowych, jakie wiążą się z ich wykorzystaniem - by móc później scharakteryzować na tym tle metodę aproksymacji dyfuzyjnej.

Ograniczono się do metod najbardziej znanych i posiadających swoje implementacje numeryczne. Pominięto takie zagadnienia,jak opis pracy stanowisk w czasie dyskretnym, por. np. [Bhar 80], systemy z bardziej złożonymi regulaminami szeregowania zadań [Woli 82], jak również szczególnie obecnie rozwijane zagadnienia opisu synchronizacji zadań oraz równoczesnego wykorzystywania przez zadanie wielu zasobów, np. [Thom 83, Heid 83, Iaco 82].

1.1. Sformułowanie problemu i notacja

Poniżej przedstawiono ważniejsze wyniki dotyczące modeli w postaci ogólnej, tj. w postaci sieci złożonej z dowolnej liczby stanowisk i posiadającej dowolną topologię - rys. 1.1. Omawiać będziemy kolejno modele, których wynik ma jawną postać analityczną (nie zawsze użyteczną w obliczeniach), modele wymagające użycia metod numerycznych oraz modele, przy rozwiązywaniu których czynione są w stosunku do założeń różne uproszczenia. Modele te będziemy nazywać odpowiednio modelami analitycznymi, numerycznymi i przybliżonymi.

Opis każdego stanowiska w sieci na rys. 1.1 zawiera rozkład czasów obsługi klientów oraz regulamin szeregowania w kolejce zgłoszeń czekających na obsługę. Opis sieci uzupełniają prawdopodobieństwa przejść pomiędzy każdą parę stanowisk. Sieć może być zamknięta - zawiera wówczas stałą liczbę przebywających w niej ciągle klientów lub otwarta - klienci wchodzą do niej i opuszczają ją, a liczba klientów obecnych w sieci jest zmienna. Klienci są nierozróżniani lub są zgrupowani w klasy; klienci poszczególnych klas różnią się czasem obsługi w stanowiskach i drogą w sieci, mogą też różnić się priorytetem w korzystaniu z czasu pracy stanowisk. Sieć mieszana to sieć otwarta dla pewnych klas klientów i zamknięta dla innych.



Rys. 1.1. Ogólna postać sieci stanowisk obsługi jako modelu systemu komputerowego

Fig. 1.1. A general queueing network as a model of a computer system

Wprowadźmy następujące oznaczenia

Dla sieci:

M - liczba stanowisk obsługi w sieci,

N - liczba (stała lub zmienna) klientów w sieci,

K - liczba klas klientów,

λ - liczba klientów przychodzących do sieci w jednostce czasu;

Dla stanowiska i w sieci:

$A_i(x), a_i(x)$ - dystrybuanta i funkcja gęstości rozkładu odstępu czasu między kolejnymi zgłoszeniami; wartość średnia tego rozkładu to $E[a_i] = 1/\lambda_i$, gdzie λ_i jest średnią liczbą klientów nadchodzących do stanowiska i w jednostce czasu; wariancja - σ^2_{ai}, współczynnik zmienności rozkładu A_i - $C_{ai} = \sigma^2_{ai} \lambda_i^2$.

$B_i(x), b_i(x)$ - dystrybuanta i funkcja gęstości rozkładu czasu obsługi; wartość średnia $E[b_i] = s_i = 1/\mu_i$, μ_i jest współczynnikiem obsługi - średnią liczbą klientów, jaką mogłoby obsłużyć stanowisko w jednostce czasu, gdyby pracowało bez przerwy; średni czas obsługi jest stały lub zależny od liczby klientów obecnych w stanowisku, $s_i = s_i(n_i)$, wariancja rozkładu σ^2_{bi}, współczynnik zmienności $C_{bi} = \sigma^2_{bi}\mu_i^2$

n_i - liczba klientów w stanowisku i,

$\mathcal{P}_i$ - przepustowość - średnia liczba klientów obsłużonych w jednostce czasu; w stanie ustalonym $\mathcal{P}_i = \lambda_i$.

ρ_i - współczynnik wykorzystania stanowiska - część czasu, przez który stanowisko jest czynne,

r_{ij} - prawdopodobieństwo, że klient po zakończeniu obsługi w stanowisku i przejdzie do kolejki stanowiska j; r_{o} jest prawdopodobieństwem, że bezpośrednio po zakończeniu obsługi w stanowisku i klient opuści sieć; r_{oi} jest prawdopodobieństwem, że klient rozpocznie pobyt w sieci od stanowiska i; w sieci zamkniętej $r_{oi} = r_{io} = 0$;

$\underline{R} = [r_{ij}]$ nazywamy macierzą tranzycji.

W przypadku wielu klas klientów parametr odnoszący się do klasy k posiada dodatkowy indeks, np.

N^k - liczba klientów klasy k w sieci,

n_i^k - liczba klientów klasy k w stanowisku i,

s_i^k - średni czas obsługi klientów klasy k w stanowisku i,

r_{ij}^{kl} - prawdopodobieństwo, że klient klasy k po zakończeniu obsługi w stanowisku i przejdzie do stanowiska j stając się klientem klasy l.

Przez rozwiązanie modelu rozumie się określenie stacjonarnego rozkładu prawdopodobieństwa wszystkich rozróżnialnych stanów sieci, jeżeli np. stan sieci określa liczba klientów w poszczególnych stanowiskach, to rozwiązaniem jest rozkład $p(n_1,\ldots,n_M) = p(\underline{n})$. W praktyce bardziej interesujące jest określenie kilku wynikających z rozkładu $p(\underline{n})$ wielkości, takich jak:

- rozkład $p_i(n_i)$ lub wartość średnia $E[n_i]$ liczby klientów obecnych w poszczególnych stanowiskach,
- średni czas czekania $Q_i = E[w_i]$ w kolejkach poszczególnych stanowisk lub średni czas reakcji $T_i = E[w_i] + s_i$,
- przepustowość $\mathcal{P}_i$ i wykorzystanie stanowisk ρ_i, $i = 1,\ldots,M$.

Nie ma obecnie jednej ogólnej metody rozwiązania przedstawionego powyżej problemu. Przyjęty sposób postępowania zależy od szczegółowych za-

łożeń dotyczących typu rozkładu czasu obsługi, regulaminów kolejek, a także od rozmiaru sieci - liczby stanowisk, liczby klientów i ich klas.

Istniejące modele dość często wymagają złożonych obliczeń, dlatego problem modelowania dotyczy nie tylko teorii modeli, lecz także efektywności algorytmów obliczeniowych związanych z nimi, a także uproszczonych i przybliżonych sposobów uzyskania rozwiązania. Z konieczności niniejszy rozdział zawiera opis wielu metod i stanowi bibliograficzny przegląd rozwiązań użytecznych przy projektowaniu służącego modelowaniu oprogramowania, zwracając uwagę na problemy praktycznej implementacji modeli.

1.2. Łańcuchy Markowa jako podstawa modelowania sieci stanowisk obsługi

Proces stochastyczny $N(t)$, którego wartość określa liczba klientów w pojedynczym stanowisku obsługi, przybiera wartości całkowite $n = 0,1,\ldots$; zmiana wartości procesu następuje w chwilach nadejścia klienta lub zakończenia obsługi. Jeżeli klienci nadchodzą w odstępach czasu o rozkładzie wykładniczym $a(x) = \lambda(n)e^{-\lambda(n)x}$, tj. tworzą poissonowski strumień zgłoszeń, a czas obsługi również podlega rozkładowi wykładniczemu $b(x) = \mu(n)e^{-\mu(n)x}$, to $N(t)$ jest łańcuchem Markowa z czasem ciągłym.

Stan procesu jest w pełni określony przez wartość n, a równania wiążące stacjonarne prawdopodobieństwa stanów mają postać, por. np. [Klei 75]

$$0 = -p(n)[\lambda(n) + \mu(n)] + p(n-1)\lambda(n-1) + p(n+1)\mu(n+1),$$

$$\mu(0) = 0, \qquad n = 0,1,\ldots \tag{1.1}$$

i rozwiązanie

$$p(n) = p(0)L(n)I(n)s^n, \tag{1.2}$$

gdzie $L(n) = \prod_{i=0}^{n-1} \lambda(i)$, $I(n) = \prod_{i=1}^{n} \frac{1}{C(i)}$, $C(i) = \mu(i)s$.

Warunkiem istnienia stanu stacjonarnego procesu i rozwiązania (1.2) jest

$$p(0) = \left[1 + \sum_{n=1} L(n)I(n)s^n\right]^{-1} > 0.$$

Niech do otwartej sieci M stanowisk tego typu napływa poissonowski strumień klientów, a parametr λ tego strumienia zależy od liczby klientów w całej sieci, $\lambda = \lambda(N)$. Bilans równowagi stanu $\underline{n} = [n_1,\ldots,n_M]$ sieci ma postać

$$0 = -p(\underline{n})\left[\lambda(N) + \sum_{i=1}^{M} \mu_i(n_i)(1-r_{ii})\right] +$$

$$+ \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} p\left[\underline{n}(i^+, j^-)\right] \mu_i(n_i+1)r_{ij}$$

$$+ \sum_{i=1}^{M} p\left[\underline{n}(i^-)\right] \lambda(N-1)r_{Oi} + \sum_{i=1}^{M} p\left[\underline{n}(i^+)\right] \mu_i(n_i+1)r_{iO} \tag{1.3}$$

gdzie $\underline{n}(i^-) = [n_1,\ldots,n_i-1,\ldots,n_M]$, $\underline{n}(i^+) = [n_1,\ldots,n_i+1,\ldots,n_M]$, $\underline{n}(i^+, j^-) = [n_1,\ldots,n_i+1,\ldots,n_j-1,\ldots,n_M]$ są stanami sąsiednimi do $\underline{n}$. Rozwiązaniem (1.3) jest [Jack 63]

$$p(\underline{n}) = \frac{1}{G} \prod_{i=1}^{M} \lambda(N)I_i(n_i)(e_i s_i)^{n_i}, \tag{1.4}$$

a w przypadku sieci zamkniętej, gdy $r_{Oi} = r_{iO} = 0$ dla $i = 1,\ldots,M$, [Gord 67]

$$p(\underline{n}) = \frac{1}{G} \prod_{i=1}^{M} I_i(n_i)(e_i s_i)^{n_i}. \tag{1.5}$$

G jest stałą normalizacyjną, $G = \sum_{\underline{n}} p(\underline{n})$, e_i jest rozwiązaniem układu równań, zwanych równaniami ruchu

$$e_i = r_{Oi} + \sum_{j=1}^{M} e_j r_{ji}, \qquad i = 1,\ldots,M, \tag{1.6}$$

i określa średnią liczbę wizyt w stanowisku i klienta przechodzącego przez sieć. Iloczyn $e_i s_i$ jest średnim łącznym zapotrzebowaniem klienta na pracę i-tego stanowiska. W sieci otwartej rozwiązanie (1.6) jest jednoznaczne, a $e_i \lambda(N)$ określa przepustowość i-tego stanowiska.

Jeżeli λ = const., μ_i = const., dla i = 1,...,M, to

$$p(\underline{n}) = \prod_{i=1}^{M} \rho_i^{n_i}(1-\rho_i), \quad \rho_i = \lambda_i/\mu_i = \lambda e_i/\mu_i. \tag{1.7}$$

W sieci zamkniętej e_i, jako rozwiązanie układu równań jednorodnych, są znane z dokładnością do czynnika, ale obecność stałej G w (1.5) sprawia, że wybór rozwiązania (1.6) teoretycznie nie ma wpływu na wartości $p(\underline{n})$.

Poważnym ograniczeniem opisanych modeli jest założenie o wykładniczych rozkładach A(x) i B(x). Można go uniknąć i mimo to pozostać w sferze modeli Markowa przez wprowadzenie rozkładów będących kombinacją rozkładów wykładniczych. Szczególnie użyteczny jest rozkład Coxa (tzw. rozkład prawie ogólny), por. [Klei 75], którym można aproksymować dowolne rozkłady. Przyjmijmy np., że A(x) pozostaje bez zmian, a czas obsługi podlega rozkładowi Coxa:

$$b(x) = \sum_{f=1}^{F} (g_f \mu_1 e^{-\mu_1 x} * \ldots * \mu_f e^{-\mu_f x}), \tag{1.8}$$

(gdzie $g_f = b_f \prod_{i=0}^{f-1} a_i$, * jest operacją splotu, $a_o = 1$) tzn. składa się z pewnej liczby faz (co najmniej jednej i nie więcej niż F), których czas trwania ma rozkład wykładniczy z parametrem μ_f dla fazy f. Faza f jest z prawdopodobieństwem b_f fazą ostatnią ($b_F = 1$); faza następna rozpoczyna się z prawdopodobieństwem $a_f = 1-b_f$. Stan systemu obsługi określa para (n,f), gdzie f jest aktualnie wykonywaną fazą obsługi, a bilans prawdopodobieństw stanu (n,f) ma postać

$$0 = -p(n,f)[\lambda(n) + \mu_f] + p(n-1,f)\,\lambda(n-1)$$

$$+ p(n,f-1)\mu_{f-1}a_{f-1}(1-\delta_{f,1}) + \delta_{f,1}\sum_{j=1}^{F} p(n+1,j)\mu_j b_j \tag{1.9}$$

gdzie δ_{ij} jest deltą Kroneckera.

Równanie (1.9) nie ma rozwiązania w prostej, podobnej do (1.2) formie (por. [Perr 83]), nie ma również sieć stanowisk tego typu. Prosta forma rozwiązania równań bilansu prawdopodobieństw pojedynczego stanowiska oraz iloczynowa, jak w przypadku (1.4) i (1.5), forma rozwiązania dla sieci stanowisk jest uwarunkowana ich dodatkową własnością [Chan 72, Chan 77]: intensywność wejścia w dany stan związana z nadejściem klienta

do stanowiska jest równa intensywności wyjścia z tego stanu związanej z odejściem klienta ze stanowiska. Równania (1.1) można rozłożyć na równania

$$0 = -p(n)\lambda(n) + p(n+1)\mu(n+1),$$
$$0 = -p(n)\mu(n) + p(n-1)\lambda(n-1), \tag{1.10}$$

a każde z równań (1.3) rozłożyć na równania

$$0 = -p(\underline{n})\mu_i(n_i) + e_i\lambda(N-1)p[\underline{n}(i^-)], \quad i = 1,\ldots,M, \tag{1.11}$$

natomiast podobnej operacji z równaniami (1.9) zrobić nie można. Równania (1.1), (1.3), (1.9), uwzględniające wszystkie dopuszczalne przejścia między stanami, nazywa się równaniami równowagi globalnej, równania (1.10), (1.11) związane z ruchem jednego klienta - równaniami równowagi lokalnej. Wśród modeli markowowskich jest więc klasa modeli posiadających analityczne rozwiązanie równań równowagi lokalnej w postaci jawnej

$$p(\underline{n}) = \frac{1}{G}\sum_{i=1}^{M} h_i(n_i), \tag{1.12}$$

w której postać $h_i(n_i)$ zależy tylko od własności stanowiska i. Modele posiadające rozwiązanie w postaci (1.12) nazywa się sieciami dekomponowalnymi (separowalnymi) lub sieciami o rozwiązaniu iloczynowym. Dla pozostałych modeli markowowskich trzeba szukać rozwiązania równań równowagi globalnej drogą numeryczną. Stanowiska z rozkładem czasów obsługi typu Coxa pozostają w równowadze lokalnej tylko w przypadku, gdy obsługa klienta rozpoczyna się natychmiast po jego nadejściu. Wśród wielu możliwych stanowisk w równowadze lokalnej [Noet 79] i prawie ogólnym rozkładem B(x) zastosowania znalazły następujące:

(i) stanowisko posiadające dostatecznie wiele identycznych równoległych kanałów obsługi, by równocześnie i z tym samym B(x) obsłużyć wszystkich klientów. Stan tego stanowiska wyznacza wektor $\underline{\vartheta} = [\vartheta_1,\ldots,\vartheta_F]$, ϑ_f jest liczbą klientów w fazie f obsługi; równanie równowagi globalnej dla stanu $\underline{\vartheta}$ ma postać

$$0 = -p(\underline{\vartheta})\Big[\lambda(n) + \sum_{f=1}^{F} \vartheta_f\mu_f\Big] + p[\vartheta(1^-)]\,\lambda(n-1)$$

$$+ \sum_{f=1}^{F-1} p\big[\underline{\vartheta}(f^+,(f+1)^-)\big]\vartheta_f\mu_f a_f + \sum_{f=1}^{F} p[\vartheta(f^+)](\vartheta_{f+1})\mu_f b_f.$$

$$0 = -p(\underline{\varrho})\lambda(0) + \sum_{f=1}^{F} p[\underline{\varrho}(f^+)]\mu_f b_f, \qquad (1.13)$$

a równanie równowagi lokalnej dla tego stanu związane z wejściem i wyjściem jednego klienta z fazy f obsługi, to

$$0 = -p(\underline{\varrho})\varrho_f\mu_f + e_f\lambda(n-1)p[\underline{\varrho}(f^-)], \quad f = 1,\ldots,F, \qquad (1.14)$$

$$e_f = a_1 \ldots a_f.$$

Rozwiązaniem (1.13) lub (1.14) jest

$$p(\underline{\varrho}) = p(0)L(n)I(n)n! \prod_{f=1}^{F} \frac{1}{\varrho_f!}(e_f s_f)^{\varrho_f}, \quad s_f = \frac{1}{\mu_f}, \qquad (1.15)$$

a prawdopodobieństwo, że w systemie jest n klientów

$$p(n) = \sum_{\underline{\varrho} : \sum_{f=1}^{F}\varrho_f = n} p(\underline{\varrho}) = p(o)L(n)I(n)s^n, \qquad (1.16)$$

gdzie

$$s = \sum_{f=1}^{F} e_f s_f.$$

Do modelu stanowiska można wprowadzić klasy klientów, zakładając, że klient klasy k, k=1,...,K, występuje w strumieniu wejściowym z prawdopodobieństwem a^k, $\sum_{k=1}^{K} a^k = 1$, i oznaczając stan systemu wektorem $\underline{\varrho} = [\underline{\varrho}^1,\ldots,\underline{\varrho}^k,\ldots,\underline{\varrho}^K]$, gdzie $\underline{\varrho}^k = [\varrho^k_1,\ldots,\varrho^k_f,\ldots,\varrho^k_F]$, ϱ^k_f oznacza liczbę klientów klasy k w fazie f obsługi. Po uwzględnieniu w (1.13), (1.14) klas klientów otrzymujemy [Koba 78]

$$p(n^k) = p(0)L(n)I(n)n! \sum_{k=1}^{K} \frac{1}{n^k!}(a^k s^k)^k \qquad (1.17)$$

oraz

$$p(n) = p(0)L(n)I(n)s^n, \quad s = \sum_{k=1}^{K} a^k s^k, \qquad (1.18)$$

(ii) stanowisko, którego moc rozdziela się równo pomiędzy wszystkich klientów,

(iii) stanowisko z regulaminem priorytetowym dającym pierwszeństwo klientowi, który przyszedł najpóźniej: w chwili nadejścia nowego klienta bieżąca obsługa zostaje przerwana i jest kontynuowana od miejsca przerwania po zakończeniu obsługi zgłoszeń, które pojawiły się później.

Dla stanowisk typu (ii) i (iii) równania równowagi i ich rozwiązania mają postać analogiczną do (1.14)-(1.18).

Zachowujące równowagę lokalną stanowiska z prawie ogólnym rozkładem czasu obsługi oraz stanowiska z regulaminem naturalnym i wykładniczym rozkładem b(t) można połączyć w sieć [Bask 75], tworząc najogólniejszy ze stosowanych obecnie modeli w formie sieci dekomponowalnej.
Sieć może być otwarta dla jednych klas klientów, a zamknięta dla innych; w stanowiskach z regulaminem naturalnym klasy klientów nie są rozróżniane. Stan sieci opisuje wektor $\underline{n} = [\underline{\varrho}_1,\ldots,\underline{\varrho}_M]$, $\underline{\varrho}_i$ dotyczy stanowiska i, i jest zdefiniowane jak poprzednio dla stanowisk z rozkładem Coxa; dla stanowisk z regulaminem naturalnym $\underline{\varrho}_i = n_i$. Rozwiązanie tej sieci jest naturalnym uogólnieniem (1.15) [Bask 75, Koba 78]

$$p(\underline{n}) = \frac{1}{G}\prod_{i=1}^{M} L(N)I_i(n_i)n_i! \sum_{k=1}^{K} \prod_{f=1}^{F_i^k} \frac{1}{\varrho^k_{if}!}(e^k_{if}\, s^k_{if})^{n^k_{if}}, \qquad (1.19)$$

gdzie $e^k_{if} = (\prod_{l=1}^{f-1} a^k_{il})e^k_i$

lub po agregatyzacji

$$p(n^k_i) = \frac{1}{G}\prod_{i=1}^{M} L(N)I_i(n_i)n_i! \prod_{k=1}^{K} \frac{1}{n^k_i!}(e^k_i\, s^k_i)^{n^k_i}, \qquad (1.20)$$

$$p(n_i) = \frac{1}{G}\prod_{i=1}^{M} L(N)I_i(n_i)s_i^{n_i}, \quad s_i = \sum_{k=1}^{K} e^k_i\, s^k_i. \qquad (1.21)$$

W (1.19) zakłada się, że strumień zewnętrzny, jeżeli istnieje, jest funkcją łącznej liczby N klientów w sieci, tj. $\lambda = \lambda(N)$.

1.3. Obliczanie dokładnych i przybliżonych rozwiązań sieci dekomponowalnych

Mimo istnienia rozwiązania w jawnej postaci (1.19-1.21) jego praktyczne wykorzystanie nie jest wcale proste. Przede wszystkim trzeba wyznaczyć stałą normalizacyjną G. Liczba stanów, które może przyjmować sieć rośnie bardzo szybko wraz ze wzrostem wartości M, N, K, F_i, np. sieć zamknięta stanowisk z regulaminem naturalnym ($K = F_i = 1$) ma $\binom{N+M-1}{N}$ stanów, czyli dla $M = N = 10$ liczba stanów przekracza 92 000 - tyle wyrażeń typu (1.5) trzeba by obliczyć, zsumować, by wyznaczyć stałą normalizacyjną, przeskalować, by móc dopiero przystąpić do obliczania wartości średnich $E[n_i]$, $E[w_i]$. Dla dużej sieci, biorąc pod uwagę czas obliczeń i błędy zaokrągleń, zadanie staje się niewykonalne. Aby ominąć część tych trudności stosuje się zazwyczaj jedną z dwóch metod obliczeniowych: metodę splotową lub metodę wartości średnich.

Metoda splotowa obliczania stałej normalizacyjnej opiera się na spostrzeżeniu [Buze 73], że dla sieci o rozwiązaniu iloczynowym stałę normalizacyjną, której wartość jest między innymi funkcją rozmiaru sieci i liczby klientów, $G = G(M,N)$, można wyrazić poprzez stałe dla sieci zmniejszonych o jedno stanowisko i zawierających mniej klientów. Dla sieci zamkniętej z jedną klasą klientów

$$G(M,N) = \sum_{n=0}^{N} G_{i^-}(M-1,\ N-n)\ h_i(n) \tag{1.22}$$

indeks i^- wskazuje, że stała odnosi się do sieci, z której usunięto stanowisko i. Obliczenie stałej $G(M,N)$ sprowadza się więc do rekurencyjnego obliczania splotu (N+1) - wymiarowych wektorów

$$\underline{G}_m = \underline{G}_{m-1} \circledast \underline{h}_m \qquad m = 2,\ldots,M \tag{1.23}$$

gdzie $\underline{G}_m = [G(m,n)]_{n=0,\ldots,N}$, $\underline{h}_m = [h_m(n)]_{n=0,\ldots,N}$,

$\circledast$ jest operacją splotu wektorowego; z warunkiem początkowym $\underline{G}_1(n) = \underline{h}_1(n)$, $n = 0,\ldots,N$.

Stała $G(M,N)$ jest ostatnią składową wektora $\underline{G}_M$. Ponadto rozkłady brzegowe $p_i(n_i)$ nie wymagają uprzedniego wyznaczenia $p(\underline{n})$, lecz wyrażają się przez wymienione stałe:

$$p_i(n) = \sum_{\underline{n}:n_i=n} p(\underline{n}) = h_i(n)\,\frac{G_{i^-}(M-1,\ N-n)}{G(M,N)}, \tag{1.24}$$

co dodatkowo upraszcza obliczanie miar pracy stanowisk, np. przepustowość wyznaczamy jako

$$\mathcal{P}_i = \sum_{n=1}^{N} p_i(n)\ _i(n) = e_i\,\frac{G_{i^-}(M,N-1)}{G(M,N)}\ . \tag{1.25}$$

W przypadku wielu klas klientów [Reis 75] (załóżmy, że klienci nie mogą zmieniać klasy, $N^k = \text{const}$) w miejscu splotu wektorowego występują sploty macierzy K-wymiarowych:

$$G(m,N^1,\ldots,N^k) = \sum_{n^1=0}^{N^1} \cdots \sum_{n^k=0}^{N^k} G(m-1,n^1,\ldots,n^k) \circledast H(m,N^1-n^1,\ldots,N^k-n^k), \qquad m=1,\ldots,M, \tag{1.26}$$

$$G(1,n^1,\ldots,n^k) = H(1,n^1,\ldots,n^k)\ .$$

Jeżeli klienci mogą zmieniać klasy, to zbiór K klas można podzielić na $C \leqslant K$ podzbiorów, zwanych łańcuchami, pomiędzy którymi nie ma wymiany klientów i transformować rozważaną sieć w sieć z C klasami klientów bez możliwości zmiany klas [Brue 80]. Dalsze oszczędności obliczeniowe dla sieci o rzadkich macierzach $\underline{R}$ i łańcuchach, w których ruch klientów jest ograniczony do części sieci daje odpowiednie uporządkowanie kolejności obliczeń (1.23), (1.26) [Lam 81]. (1.23) i (1.26) upraszczają się też znacznie, jeżeli średnie czasy obsługi s_i^k są stałe, tj. niezależne od n_i. Implementacje algorytmów obliczeniowych w poszczególnych przypadkach sieci zamkniętych omawia dokładnie [Brue 80], implementację dla sieci mieszanych - przy ograniczeniach zapewniających zbieżność algorytmu (stały strumień wejściowy, stałe, począwszy od pewnej liczby klientów w stanowiskach, czasy obsługi) zawiera [Mere 78].

Wartość obliczonej stałej normalizacyjnej zależy od przyjętego rozwiązania równań ruchu [Lam 82]

$$G = G(M,\underline{N},\underline{e}) = (e_1^1)^{N_1} \cdots (e_1^k)^{N^k}\, G(M,\underline{N},\underline{1}), \tag{1.27}$$

gdzie $\underline{N} = [N^1,\ldots,N^k]$

i w zależności od wyboru $\underline{e}$ może w trakcie obliczeń szybko rosnąć lub maleć, przekraczając dopuszczalny w maszynie cyfrowej zakres liczb zmiennoprzecinkowych. Postać zależności (1.27) pozwala na łatwe przeskalowanie G w trakcie obliczeń

$$G(M,\underline{N},\underline{e}') = G(M,\underline{N},\underline{e}) \prod_{k=1}^{K} \left(\frac{e'^k}{e^k}\right)^{N^k} \tag{1.28}$$

i utrzymanie jej wartości w dopuszczalnych granicach.

W metodzie wartości średnich [Reis 79, 80], obliczanie stałej normalizacyjnej jest niepotrzebne. Metoda wykorzystuje fakt, iż rozkład dowolnej kolejki obserwowany przez nadchodzącego klienta jest stacjonarnym rozkładem tej kolejki po usunięciu z sieci wspomnianego klienta [Sevc 79, Lave 80]:

$$p_i(\underline{n}_i, \underline{N}) \Big|_{\text{klienta klasy } k}^{\text{tuż przed nadejściem}} = p_i\left[\underline{n}_i, \underline{N}(k^-)\right], \tag{1.29}$$

gdzie $\underline{N}(k^-) = [N^1,\ldots,N^k-1,\ldots N^K]$,

a więc średni czas przejścia T_i^k przez stanowisko i (przy niezależnym od liczby klientów czasie obsługi) wynosi

$$T_i^k(\underline{N}) = \begin{cases} s_i^k & \text{dla stanowisk z większą od } N \text{ liczbą kanałów obsługi} \\ s_i^k\left[1 + Q_i\ \underline{N}(k^-)\right] & \text{dla pozostałych typów stanowisk sieci dekomponowalnych.} \end{cases} \tag{1.30}$$

Metoda korzysta też z prawa Little'a, por. np. [Klei 75], wiążącego średnią liczbę klientów nadchodzących w jednostce czasu do stanowiska λ, średnią liczbę klientów obecnych w stanowisku Q i średni czas reakcji T

$Q = \lambda T$, stosując je do całej sieci

$$\lambda_i^k(\underline{N}) = \frac{N^k}{\mathcal{T}_i^k(\underline{N})} \tag{1.31}$$

i do pojedynczego stanowiska

$$Q_i^k(\underline{N}) = \lambda_i^k(\underline{N}) T_i^k(\underline{N}) \qquad i = 1,\ldots,M. \tag{1.32}$$

$\mathcal{T}_i^k$ jest średnim czasem obiegu sieci, poczynając od stanowiska i, klientów klasy k. Ponieważ pomiędzy wizytami w stanowisku i klient przchodzi przez stanowisko j średnio e_j^k/e_i^k razy, więc

$$\mathcal{T}_i^k(\underline{N}) = \sum_{j=1}^{M} T_j^k(\underline{N}) \frac{e_j^k}{e_i^k}. \tag{1.31a}$$

Zakładając $T_i^k(\underline{0}) = 0$ i zaczynając od populacji $\underline{0}$, a kończąc na $\underline{N}$ oblicza się, wykonując iteracyjnie (1.30)-(1.32), wielkości $Q_i^k(\underline{N})$, $T_i^k(\underline{N})$, a z nich dalsze miary pracy systemu. W przypadku zależnego od liczby klientów czasu obsługi rozkład długości zastanej w momencie nadejścia kolejki nie określa czasu reakcji, bo późniejsze zgłoszenia mogą zmienić czas obsługi; (1.30) przybiera wówczas postać

$$T_i^k(\underline{N}) = \frac{n_i^k(\underline{N})}{\lambda_i^k(\underline{N})} = \sum_{\underline{0}}^{\underline{N}} \left[\sum_{k=1}^{k} n_i^k s_i^k(n_i^k) p_i\left[\underline{n}_i(k^-), \underline{N}(k^-)\right]\right], \tag{1.33}$$

gdzie $\underline{n}_i(k^-) = [n_1,\ldots,n_i^k-1,\ldots,n_I^k]$,

co wymaga również iteracyjnego obliczania rozkładów brzegowych

$$p_i(\underline{n}^k,\underline{N}) = \sum_{k=1}^{K} s_i^k(n_i^k)\lambda_i^k(\underline{N}) p_i\left[\underline{n}_i(k^-)\underline{N}(k^-)\right], \tag{1.34}$$

$$p_i(\underline{0}, \underline{N}) = 1 - \sum_{\underline{n}_i\,:\,\sum_{k=1}^{K} n_i^k \geq 1} p_i(\underline{n}^k, \underline{N}). \tag{1.35}$$

Jeżeli wykorzystanie ρ_i stanowiska jest duże, (1.35) może stać się źródłem numerycznej niestabilności algorytmu i znacznych błędów [Chan 80], dlatego [Reis 81] proponuje obliczanie $p_i(\underline{0},\underline{N})$ poprzez agregację sieci poza stanowiskiem i (co będzie omówione w podrozdz. 5)

$$p_i(\underline{0},\underline{N}) = p_i\left[\underline{0}, \underline{N}(k^-)\right] \lambda^k(\underline{N}) / \lambda_{z(i)}^k(\underline{N}), \tag{1.36}$$

gdzie $\lambda_{z(i)}^k(\underline{N})$ jest przepływem przez zredukowaną sieć.

Istnieje różne odmiany algorytmu wartości średnich, wśród nich algorytmy łączące metodę średnich z metodą splotową [Chan 80, Reis 81].

Przedstawiony powyżej algorytm odnosi się do stałych populacji N^k, w przypadku klientów zmieniających klasy trzeba transformować sieć podobnie jak w metodzie splotowej [Brue 80]. Rozszerzenie metody średnich na sieci otwarte i mieszane zawierają: [Zaho 81] w przypadku stałych s_i^k i [Tucc 82] dla czasów obsługi zależnych od obciążenia. Szczegółowe porównanie złożoności obliczeniowej metod splotowej i wartości średnich zawiera [Zaho 80]. W przypadku stałych, niezależnych od liczby klientów średnich czasów obsługi, jest ona podobna, proporcjonalna do KM $\prod_{k=1}^{K} (N^k+1)$.

W ograniczonych przypadkach złożoność metody splotowej można zmniejszyć kosztem znacznie większego zapotrzebowania pamięci [Chan 80]. Stosunek rozmiaru wymaganej pamięci w metodzie splotowej i metodzie średnich wynosi $\frac{2 \max_k N^k + 1}{KM+K+M}$, względna efektywność tych metod zależy więc od parametrów sieci. Łatwość implementacji, poglądowość, brak potrzeby skalowania stałej normalizacyjnej i większa stabilność przemawiają za metodą wartości średnich. W przypadku zależnych od obciążenia czasów obsługi metoda średnich wymaga znacznie większej pamięci, a obliczenia mogą być niestabilne [Drix 82, Reis 79], jeżeli wartość $p_i(0)$ znajdzie się poniżej zakresu liczb zmiennoprzecinkowych w maszynie cyfrowej. Jeżeli niektóre stanowiska mają średnie czasy obsługi stałe, a inne zmienne, można sieć rozdzielić na dwie części i każdą z nich rozwiązać inną metodą [Reis 79]. Obie metody, dokładne w przypadku sieci separowalnych, dały początek kilku metodom przybliżonym. Z metody splotowej wywodzi się redukcja stanów dużych sieci poprzez zmniejszenie liczby klas, klientów i stanowisk obsługi [Zaho 80] - przybliżenia dużej sieci separowalnej mniejszą, również separowalną, rozwiązywalę w standardowy sposób. Metody przybliżone oparte na metodzie średnich są albo ekstrapolacją tej metody na przypadki sieci nieseparowalnych albo redukują, kosztem dokładności, czas obliczeń dla dużych sieci. Uwzględniono w sposób przybliżony m.in. stanowiska z naturalnym regulaminem kolejki i dowolnym, różnym dla poszczególnych klas rozkładam czasów obsługi, priorytety poszczególnych klas, równoczesna korzystanie przez klientów z wielu stanowisk obsługi, blokowanie (stanowisko jest wolne, ale może obsługiwać klientów dopiero po spełnieniu pewnych dodatkowych warunków) [Bard 79, Balb 79, Reis 80, Szcz 82].

W metodzie średnich trzeba przeprowadzić kolejno obliczenia dla sieci o populacji od $\underline{0}$ do $\underline{N}$. Dla uproszczenie obliczeń można założyć, że zmiana populacji $\underline{N}$ o jednego klienta klasy k niewiele zmienia rozkład kolejek [Bard 79]

$$Q_i^k[\underline{N}(k^-)] \approx Q_i^k[\underline{N}], \tag{1.37}$$

lub, że liniowo zmienia liczbę klientów klasy k w stanowisku [Schw 79]

$$Q_i^k[\underline{N}(k^-)] \approx \frac{N^k-1}{N^k} Q_i^k(\underline{N}). \tag{1.38}$$

Równania (1.37) lub (1.38) uwzględnione w (1.30) powodują, że pętla działań (1.30)-(1.32) jest wykonywana od razu dla pełnej populacji $\underline{N}$, a jej wielokrotne użycie służy iteracyjnemu wyznaczeniu $Q_i^k(\underline{N})$. Jako punkt startowy można np. przyjąć $Q_i^k = \frac{N^k}{M}$.

W równaniu (1.38) zakłada się, że zmiana wartości N^k odbija się tylko na kolejce klientów tej klasy, co może być źródłem poważnych błędów, jeżeli wśród pozostałych klas są klasy o małej liczbie klientów. Dlatego [Zaho 79] proponuje jako podstawę rowziązania iteracyjnego

$$T_i^k[\underline{N}(1^-)] \approx T_i^k(\underline{N}) - \frac{s_i^k}{N^1} Q_i^1(\underline{N}), \tag{1.39}$$

a [Chan 82] następująco rozbudowuje (1.37):

$$Q_i^k[\underline{N}(1^-)] \approx N^k\left[F_i^k(\underline{N}) + D_i^{k1}(\underline{N})\right]. \tag{1.40}$$

gdzie $F_i^k(\underline{N}) = \frac{Q_i^k(\underline{N})}{N^k}$ jest częścią zadań klasy k pozostających w kolejce i, a $D_i^{k1}(\underline{N}) = F_i^k[\underline{N}(1^-)] - F_i^k(\underline{N})$ jest zmianą F_i^k spowodowaną usunięciem z populacji jednego klienta klasy 1. Nie ma co prawda dowodu zbieżności opisanych iteracji, nie wiadomo czy uzyskane rozwiązanie jest jedyne, nieznany jest też maksymalny błąd tych aproksymacji, niemniej zaczynają one w praktyce wypierać dokładne algorytmy metody wartości średnich [Chan 82, Zaho 80] w przypadku stałych średnich czasów obsługi, do którego się odnoszę. Można również traktować równania (1.29)-(1.32) oraz jedno z uzupełnień (1.38)-(1.40) jako układ równań nieliniowych i poszukiwać jego rozwiązanie innymi metodami, zapewniającymi zbieżność i jednoznaczność [Chow 83].

1.4. Rozwiązania numeryczne niedekomponowalnych sieci markowowskich

Modele markowowskie pozwalają opisać przypadki znacznie ogólniejsze od sieci separowalnych - można w nich uwzględnić dowolne zależności między zdarzeniami - mają jednak w praktyce bardzo poważne ograniczenie: wymiar układu równań równowagi globalnej, które trzeba rozwiązać, a których liczba jest równa liczbie etanów modelu, szybko przekracza możliwości obliczeniowe dużych maszyn cyfrowych. Postępowanie obejmuje następujące etapy:

1. Odwzorowanie przestrzeni stanów $\underline{n}$ w wektor, $[n_1,\ldots,n_M] \rightarrow n$, tak, by system mógł być opisany dwuwymiarową macierzą $\underline{Q}$ przejść między etanami, i wyznaczenie tej macierzy.

Odwzorowanie musi być wzajemnie jednoznaczne. W przypadku sieci otwartych, których nieskończoną przestrzeń stanów trzeba ograniczyć, jest waż-

na, by odwzorowanie zachowywało sąsiedztwo stanów; może to być np. odwzorowanie [King 80]

$$[n_1,\ldots,n_M] \rightarrow \sum_{i=1}^{M} \left[\prod_{j=0}^{i-1} \sum_{r=1}^{i} (n_r + j)/i! \right]. \qquad (1.41)$$

Trzeba też wziąć pod uwagę, że przy odwzorowaniu odwrotnym $n \rightarrow \underline{n}$ mogą być generowane stany wzbronione i zabezpieczyć się przed tym, dobierając elementy macierzy $\underline{Q}$ tak, by stacjonarne prawdopodobieństwa stanów n generujących wzbronione stany $\underline{n}$ były zerowe [King 80].

2. Dobór rozmiaru macierzy $\underline{Q}$ w przypadku nieskończonej liczby stanów. Można posłużyć się tu następującą zależnością [Twee 73]

$$\frac{cof(i,j)}{cof(j,j)} < \frac{p(i)}{p(j)} < \frac{cof(i,i)}{cof(j,i)}, \qquad (1.42)$$

gdzie $cof(i,j)$ jest dopełnieniem algebraicznym elementu q_{ij} macierzy $\underline{Q}^{(n)}$, powstałej przez pozostawienie w macierzy $\underline{Q}$ n pierwszych wierszy i kolumn (i skreślenie pozostałych), by dobrać n, dla którego $p(n)/p(0)$ jest wystarczająco małe.

3. Rozwiązanie otrzymanego układu równań

$$\underline{Q}^T \underline{P} = \underline{0}. \qquad (1.43)$$

Element q_{ij} macierzy $\underline{Q}$ jest prawdopodobieństwem przejścia ze stanu i do stanu j, $q_{ii} = -\sum_{\substack{j=1 \\ j \neq i}}^{n} q_{ij}$, $\underline{P}$ jest wektorem prawdopodobieństw stanów. Macierz $\underline{Q}$ jest z reguły macierzą dużą i rzadką, dlatego do rozwiązania (1.43) wygodnie stosować metody iteracyjne, nie zmieniające jej zawartości i umożliwiające przez to pamiętanie jej w formie upakowanej. Jednakże powszechnie używane metody iteracyjne - Jacobiego, Gaussa-Seidela czy nadrelaksacji - nie gwarantują w przypadku macierzy $\underline{Q}$ zbieżności i w praktyce są często wolnozbieżne [Stew 77]. Dlatego przekształca się (1.43) do postaci

$$\underline{W}^T \underline{P} = \underline{P} \qquad (1.44)$$

gdzie $\underline{W}^T = \underline{Q}^T \Delta t + \underline{I}$.

$\underline{I}$ jest macierzą jednostkową, a $\Delta t \leq (\max_i |q_{ii}|)^{-1}$ jest dobrane tak, by $\underline{W}$ była macierzą stochastyczną. Największa wartość bezwzględna wartości własnych macierzy stochastycznej jest równa jedności i poszukiwanie wektora $\underline{P}$ staje się poszukiwaniem lewego wektora własnego odpowiadającego dominującej wartości własnej macierzy $\underline{W}$ [Stew 78]. Niech $\underline{W}$ posiada wartości własne $\lambda_1,\ldots,\lambda_n$; $|\lambda_1| > \ldots > |\lambda_n|$ i niech $\underline{x}_1$ jest lewym wektorem własnym odpowiadającym λ_1:

$$\underline{W}^T \underline{x}_1 = \lambda_1 \underline{x}_1, \qquad \underline{x}_1 \equiv \underline{P}.$$

Metoda potęgowa, polegająca na kolejnym przemnażaniu przez macierz $\underline{W}$ dowolnie wybranego wektora początkowego $\underline{P}_{(0)}$, dopóki kolejne rezultaty nie są do siebie proporcjonalne, jest zbieżna do $\underline{x}_1$:

$$\underline{P}_{(k)} = \underline{W}^T \underline{P}_{(k-1)} = (\underline{W}^T)^k \underline{P}_{(0)} = \sum_{i=1}^{n} \alpha_i \lambda_1^k \underline{x}_1 =$$

$$= \lambda_1^k \left[c_1 \underline{x}_1 + \sum_{i=2}^{n} \left(\frac{\lambda_1}{\lambda_1}\right)^k \underline{x}_i \right]. \qquad (1.45)$$

a o jej szybkości decyduje wartość $\frac{|\lambda_2|}{|\lambda_1|}$. Ponieważ często $|\lambda_2| \approx |\lambda_1|$, dla przyspieszenia metody można ją równocześnie stosować do m wektorów próbnych, co daje pozytywne rezultaty, o ile $|\lambda_{m+1}| / |\lambda_2|$ jest ponad m-krotnie mniejsze od $|\lambda_2| / |\lambda_1|$, czas bowiem każdej iteracji wzrasta m razy [Stew 78]. Można też stosować inne metody, których zestawienie i porównanie, wskazujące na zalety metody Arnoldiego [Arno 51] zawiera [Cach 81]. [Stew 77] pokazuje, że skrócenie obliczeń można uzyskać zastępując macierz $\underline{W}$ innymi macierzami, powstałymi przez tożsamościowe przekształcenia (1.44). Ma to szczególne znaczenie w przypadku sieci prawie dekomponowalnych, gdy $\underline{W}$ ma szereg wartości własnych bliskich jedności. Sieci prawie dekomponowalne to sieci, w których macierz przejść $\underline{Q}$ można, przestawiając jej wiersze i kolumny, przedstawić jako $\underline{Q} = \underline{Q}^* + \varepsilon \underline{C}, \varepsilon << 1$, gdzie $\underline{Q}$ jest blokową diagonalną macierzą stochastyczną. Wśród stanów sieci istnieją takie ich grupy, wewnątrz których można w przybliżeniu niezależnie rozpatrywać przejścia między stanami, redukując w ten sposób znacznie wymiar problemu. Trzeba jednak wpierw znać i przeanalizować macierz $\underline{Q}$, dlatego zastosowanie dekompozycji w tym ujęciu, mimo dobrze opracowanych podstaw matematycznych [Cour 77] jest niewielkie, np. [Dodd 81]; dekompozycji dokonuje się raczej na podstawie macierzy $\underline{R}$, rozdzielając stanowiska, nie stany [Lazo 79, Vant 80, Zaho 83].

1.5. Modele wykorzystujące przybliżoną dekompozycję sieci

Naturalnym sposobem poszukiwania rozwiązań przybliżonych jest rozszerzenie rozwiązania iloczynowego (1.12) na przypadki, w których nie obowiązuje ono w sposób dokładny. W szczególności chodzi o sieć zawierającą stanowiska z regulaminem naturalnym i dowolnym rozkładam czasu obsługi. [Shum 76] zakłada, że dla takiej sieci zamkniętej, obsługującej jedną klasę klientów, rozwiązanie $p(\underline{n})$ jest w przybliżeniu iloczynem rozwiązań $p_i(n_i)$ dla niezależnych pojedynczych stanowisk typu M/G/1/N, tj. stanowisk z wykładniczym rozkładem A(x), dowolnym rozkładem B(x) i ograniczoną do N liczbą klientów w stanowisku. Model oparty na tym założeniu nazywa się modelem rozszerzonej formy iloczynowej. Postać $p(n)$ dla M/G/1/N można określić na podstawie jej związku z rozkładam $p'(n)$ dla stanowiska M/G/1 z nieograniczoną kolejkę [Shum 76]

$$p(n) = \frac{p'(n)/(1-\rho)}{1 + \frac{\rho}{1-\rho}\sum_{\vartheta=0}^{N-1} p'(\vartheta)}, \quad n = 0,\ldots,N-1; \quad p(N) = 1 - \sum_{n=0}^{N-1} p(n). \tag{1.46}$$

$p'(n)$ jest znane w postaci jawnej dla niektórych rozkładów B(x), dla innych może być obliczone z ogólnie znanej, por. np. [Klei 75] zależności Pollaczka-Chinczyna dla funkcji tworzącej $p'(n)$. Można również wprowadzić [Balb 79] do rozwiązania stanowiska wielokanałowe M/G/k/N. W sieci, która nie jest separowalna, wartość $p(\underline{n})$ zależy od wyboru rozwiązania równań ruchu, znanego dla sieci zamkniętej z dokładnością do czynnika. Po wyznaczeniu znormalizowanego wektora $\underline{e}$, w którym $\sum_{i=1}^{M} e_i = 1$, wektor przepływów $\underline{\lambda}$ dobierany jest tak, by po obliczeniu na jego podstawie $p_i(n_i)$, następnie $p(\underline{n})$, a stąd przepustowości

$$\frac{\rho_i}{e_i} = \frac{1}{e_i} \sum_{\underline{n}:n_i>0} p(\underline{n}) \qquad i = 1,\ldots,M, \tag{1.47}$$

otrzymać wartości spełniające możliwie dokładnie równanie ruchu, a więc by otrzymać wektor $\underline{\lambda}'$ o składowych pozostających do siebie w tych samych proporcjach co składowe $\underline{\lambda}$. Można to zrealizować np. przez minimalizację funkcji

$$f(\alpha) = \sum_{i=1}^{M} e_i s_i (\alpha_i - \alpha'_{i*})^2, \tag{1.48}$$

gdzie α określa wektor wyjściowy, $\underline{\lambda} = \alpha \underline{e}$, a $\underline{\alpha}'$ określa wektor otrzymany, $\underline{\lambda}'_i = \alpha'_i e_i$. α'_{i*} jest wybraną składową $\underline{\alpha}'$, np. odpowiadającą stanowisku o największym zatłoczeniu, gdzie wyniki są wówczas najdokładniejsze. [Chan 75] proponuje dla tej samej sieci inny, nazwany tu iteracyjną metodą równoważnego stanowiska, schemat obliczania $p_i(n_i)$ poprzez redukcję sieci do dwóch stanowisk - wyróżnionego stanowiska "i" i stanowiska zastępczego, reprezentującego resztę sieci. Jest to stanowisko z wykładniczym rozkładem $B_{z(i)}(x)$, którego wartość średnia $s_{z(i)}(n)$ jest obliczana na podstawie przepustowości sieci mającej n klientów oraz zwarte stanowisko i : $s_i = 0$. Rozwiązanie dwustanowiskowej sieci złożonej z wykładniczego stanowiska zastępczego i dowolnego stanowiska "i" jest znane [Cour 71] i dostarcza $p_i(n_i)$, $i = 1,\ldots,M$. Redukcja sieci poza stanowiskiem i do jednego stanowiska (źródła klientów w przypadku sieci otwartej) jest nazywana przez analogię z podobną procedurę dla obwodów elektrycznych twierdzeniem Nortona. Ogólniejszą wersję twierdzenia dotyczącą wielu klas klientów lub dowolnego fragmentu sieci zawierają odpowiednio [Krit 82] i [Bals 82]. Redukcja jest dokładna w przypadku sieci dekomponowalnej, dlatego [Chan 75] posługuje się przy niej pomocniczą siecią dekomponowalną o tej samej topologii co sieć badana i średnich czasach obsługi s_i korygowanych iteracyjnie po wyznaczeniu wszystkich $p_i(n_i)$ tak, by wyznaczone na podstawie $p(\underline{n})$ przepustowości stanowisk spełniały równania ruchu, a łączna średnia długość kolejek zgadzała się z rzeczywistością, $\sum_{i=1}^{M} Q_i = N$. Marie [Mari 79b] zwiększa dokładność metody obliczając, po wyznaczeniu strumienia wyjściowego $\lambda_i(N-n_i)$ ze stanowiska zastępczego, $p_i(n_i)$ jako rozkład w izolowanym stanowisku obsługi z rozkładem czasu obsługi Coxa i ograniczoną do N kolejkę. Jeżeli te same, co w [Chan 75] testy poprawności rozwiązania nie są spełnione, inaczej dobierane są czasy obsługi w sieci pomocniczej:

$$s_i^*(n_i) = \frac{p_i(n_i)}{\lambda_i(n_i-1)p_i(n_i-1)}. \tag{1.49}$$

Rozwiązaniem nie jest tu iloczyn $p_i(n_i)$, lecz rozwiązanie dla dekomponowalnej sieci pomocniczej. Poszukuje się w ten sposób równoważnej sieci dekomponowalnej o tych samych rozkładach kolejek co sieć badana. Sieć taka być może istnieje, chociaż przedstawiony w [Balb 79] dowód istnienia okazał się błędny [Rubi 84]. Opisany algorytm Marie'go do niej jednak nie prowadzi [Stew 80]. Inny wariant metody proponuje [Balb 79], a [Neus 82] zawiera jej rozszerzenie na przypadek regulaminów priorytetowych. [Sevc 77] stara się zwiększyć dokładność metody przez ocenę drugiego momentu rozkładu odstępów czasu między klientami w strumieniu wyjściowym stanowiska zastępczego, tj. w strumieniu wejściowym stanowiska i. Problemem własności strumieni klientów w sieci zajmuję się szerzej metody izolacji

stanowisk, wyznaczając w sposób przybliżony wartość drugiego momentu rozkładu $A_i(x)$ dla każdego stanowiska. Zakłada się dla uproszczenia, że strumienie wyjściowe poszczególnych stanowisk są niezależne i są procesami odnowy, tj. odstępy czasu pomiędzy kolejnymi klientami opuszczającymi dane stanowisko i są niezależne i mają ten sam rozkład $D_i(x)$. Parametry tego rozkładu można oszacować korzystając z zależności dokładnej dla stanowiska M/G/1, por. [Klei 75], przybliżonej dla stanowiska z ogólnym strumieniem wejściowym [Czac 79], więżącej gęstość $d_i(x)$ rozkładu $D_i(x)$ z gęstościami rozkładów $A_i(x)$, $B_i(x)$:

$$d_i(x) = \rho_i b_i(x) + (1-\rho_i)a_i(x) * b_i(x), \tag{1.50}$$

skąd obliczamy $E[d_i^2]$, a następnie współczynnik zmienności rozkładu

$$c_{di} = \frac{E[d_i^2] - E[d_i]^2}{E[d_i]^2} = c_{ai}(1-\rho_i) + \rho_i^2 c_{bi} + \rho_i(1-\rho_i). \tag{1.51}$$

Wartość średnia $E[d_i]$ to oczywiście $1/\lambda_i$, równania ruchu są te same dla dowolnych rozkładów $B_i(x)$. Strumień wyjściowy rozdziela się za stanowiskiem zgodnie z prawdopodobieństwami przejść r_{ij}, strumień pomiędzy stanowiskami i i j charakteryzuje rozkład

$$d_{ij}(x) = r_{ij}d_i(x) + r_{ij}(1-r_{ij})d_i(x) * d_i(x) + r_{ij}(1-r_{ij})^2 d_i(x) *$$

$$* d_i(x) * d_i(x) + \ldots, \tag{1.52}$$

skąd

$$\lambda_{ij} = r_{ij}\lambda_i \qquad c_{dij} = (c_{di}-1)r_{ij}. \tag{1.53}$$

Strumień wejściowy stanowiska j jest superpozycją niezależnych procesów $d_{ij}(t)$ i jako suma procesów odnowy ma współczynnik zmienności, por. [Kueh 79, Gele 76],

$$c_{aj} = \frac{1}{\lambda_j}\sum_{i=1}^{M} r_{ji}\lambda_i\left[(c_{di}-1)r_{ji} + 1\right] + \frac{c_{oi}r_{oi}}{j} \tag{1.54}$$

Ostatni człon (1.54) odnosi się do strumienia zewnętrznego o parametrach λ_i c_{oi} wpływającego do sieci otwartej. Równania (1.51) i (1.54) tworzą układ równań liniowych, z których można wyznaczyć c_{ai}, i = 1,...,M. Wartości λ_i określają jak zawsze równania ruchu. Znając powyższe para-

metry strumienia wejściowego można każde ze stanowisk analizować oddzielnie, posługując się jedną z wielu metod (ich przegląd i ocenę dokładności zawiera [Czac 79]) przybliżonego opisu pracy pojedynczego stanowiska G /G/1 z ogólnym strumieniem wejściowym i ogólnym rozkładem czasów obsługi, znanymi z dokładnością do dwóch pierwszych momentów. Wśród metod tych wyróżnia się aproksymacja dyfuzyjna, por. np. [Gele 75]. Przybliża ona wartość p(n) zastępując liczbę klientów n w stanowisku [zmiany n spowodowane traktowanymi niezależnie procesem wejściowym i wyjściowym mają, w dostatecznie długim okresie czasu t, rozkład normalny o średniej $(\lambda-\mu)t = \beta t$ i wariancji $(\lambda c_a + \mu c_b)t = \alpha t$, niezależnie od postaci A(x) i B(x)] wartością x procesu dyfuzji, którego infinitezymalne zmiany w czasie dt mają rozkład normalny o średniej βdt i wariancji αdt. Funkcja gęstości prawdopodobieństwa tego procesu spełnia w stanie ustalonym równanie

$$0 = \frac{\alpha}{2}\frac{d^2f(x)}{dx^2} - \beta\frac{df(x)}{dx}, \tag{1.55}$$

którego rozwiązanie przy warunku brzegowym ograniczającym proces do $x \geqslant 0$ przybliża $p(n) : f(n) \approx p(n)$.

Dobierając współczynniki β i α można też uwzględnić:

- klasy klientów o różnym czasie obsługi, lecz wspólnej kolejce z regulaminem naturalnym [Gele 77],

$$\beta = \lambda - \mu,$$

$$\alpha = \sum_{k=1}^{K}\lambda^k c_a^k + \mu^3\left[\sum_{k=1}^{K}(c_b^k + 1)\frac{a^k}{(\mu^k)^2} - \frac{1}{\mu^2}\right], \tag{1.56}$$

gdzie $\frac{1}{\mu} = \sum_{k=1}^{K}\frac{a^k}{\mu^k}$, $a^k = \lambda^k/\lambda$, $\lambda = \sum_{k=1}^{K}\lambda^k$,

a rachunek strumieni poszczególnych klas przebiega analogicznie do przedstawionego,

- L-kanałowe stanowiska obsługi [Hala 78, Pujo 78]

$$\left.\begin{aligned}\beta(x) &= \lambda - l\mu\\ \alpha(x) &= \lambda c_a + l\mu c_b\end{aligned}\right\} \quad l < x \leqslant l+1, \quad l=0,1,\ldots,L-1$$

$$\left.\begin{aligned}\beta(x) &= \lambda - L\mu\\ \alpha(x) &= \lambda c_a + L\mu c_b\end{aligned}\right\} \quad x > L \tag{1.57}$$

- grupowe zgłoszenia klientów [Gale 76]

$$\beta = \lambda\vartheta_g - \mu, \quad \alpha = \lambda\vartheta_g[C_a + C_g] + \mu C_b, \tag{1.58}$$

gdzie λ, C_b odnoszą się do strumienia grup klientów, a ϑ_g i C_g to wartość średnia i współczynnik zmienności liczby klientów w grupie,

- regulamin priorytetowy z wywłaszczaniem (nadejście klienta o priorytecie wyższym niż priorytet aktualnie obsługiwanego klienta powoduje natychmiastowe zawieszenie obsługi tego ostatniego i rozpoczęcie obsługi przybyszа) [Czac 80, Nain 81, Reim 83]. Rozwiązanie równania dyfuzji ze współczynnikami uwzględniającymi parametry strumienia wejściowego i czasu obsługi k klas klientów, k = 1,2,...,K, (przyjmujemy, że klienci klasy pierwszej mają najwyższy priorytet) [Czac 80]

$$\beta^k = \sum_{l=1}^{k} \lambda^l - \sum_{l=1}^{k} \frac{1}{R^I}\mu^l$$

$$\alpha^k = \sum_{l=1}^{k} \lambda^l c_a^l + \sum_{l=1}^{k} \frac{\rho^l}{R^I}\mu^l c_b^l,$$

gdzie $R^l = \sum_{j=1}^{l} \rho^l$

służy do wyznaczenia przybliżonego rozkładu $p^k(n)$ łącznej liczby klientów klas od 1 do k, skąd wyznaczamy następnie rozkład $v^k(n)$ liczby klientów należących do klasy k

$$v^1(n) = p^1(n),$$

$$v^k(n) = \frac{p^k(n) - \sum_{i=0}^{n-1} p^{k-1}(n-i)v^k(i)}{p^{k-1}(0)} \qquad k = 2,\ldots,K.$$

Użyta po raz pierwszy do opisu procesów skokowych w genetyce [Fell 51], wprowadzona do teorii kolejek przez Igleharta [Igle 65] i Newella [News 68, 71] aproksymacja dyfuzyjna była wielokrotnie wykorzystywana w modelach systemów komputerowych, poczynając od modeli elementów tych systemów - np. opis pracy pamięci dyskowej [Adam 77, Koga 82], badanie wykorzystania procesora centralnego [Gave 73a,b], [Bade 75] - a na modelowaniu i

optymalizacji sieci komputerowych kończąc [Koba 77b, Banh 78, Koga 80, Koga 80, Pujo 78, Taka 84, Remb 85].

Zamiast (1.51) można posłużyć się dokładną zależnością wiążącą C_{di} ze średnim czasem czekania $E[w_i]$ w stanowisku i typu GI/G/1, por. [Kueh 79]

$$C_{di} = C_{ai} + 2\rho_i^2 C_{bi} - 2\rho_i(1 - \rho_i)\frac{E[w_i]}{s_i} \tag{1.59}$$

gdzie $E[w_i]$ jest szacowane jedną ze wspomnianych metod przybliżonych. Jeżeli będzie to np. aproksymacja [Kräm 68], por. [Czac 79]

$$E[w_i] \approx s_i \frac{\rho_i}{2(1 - \rho_i)}(C_{ai} + C_{bi}), \tag{1.60}$$

to uzyskamy [Kueh 79]

$$C_{di} = C_{ai} + \rho_i^2(C_{bi} - C_{ai}). \tag{1.61}$$

Dalsze heurystyczne propozycje przybliżonego obliczania drugich momentów strumieni zawierają [Alle 80, Sevc 77, Trip 79]. Wyniki symulacji [Lest 83] wskazują na ograniczoną dokładność tych rachunków, opartych na założeniu, że procesy wyjściowe stanowisk są niezależnymi procesami odnowy. Trzeba też dodać, że pominięte tu trzecie momenty rozkładów $A_i(x)$, $B_i(x)$ mają również znaczny wpływ na zachowanie się stanowisk [Bere 81]. [Trip 79] próbuje krzyżować poszczególne warianty obliczeń strumieni z różnymi aproksymacjami stanowisk, szukając doświadczalnie kombinacji obciążonej najmniejszym łącznym błędem.

1.6. Podsumowanie

Powyżej omówiono najczęściej stosowane metody obliczeniowe dotyczące kolejkowych modeli systemów komputerowych. Omówione metody poświęcają najwięcej uwagi sieciom stanowisk z regulaminem naturalnym, stosunkowo słabiej opracowane są modele z szeregowaniem priorytetowym. Zmierzające do urealnienia modeli odstępstwa od założeń o poissonowskich strumieniach zgłoszeń i wykładniczych rozkładach czasów obsługi powodują trudności analityczne i obliczeniowe, każąc odwoływać się do metod przybliżonych. Dokładność tych metod jest różna i z reguły może być oceniona tylko doświadczalnie poprzez obliczenia dla typowych konfiguracji sieci i parametrów stanowisk i przez porównanie rezultatów z wynikami innych metod (w tym symulacji).

Brak jeszcze dostatecznego materiału, by szczegółowo wartościować istniejące metody przybliżone. Ocenę najbardziej metodyczną, choć niepełną i ograniczoną do bardzo małych zamkniętych sieci niedekomponowalnych zawiera [Balb 79], rekomendując metodę rozszerzonej formy iloczynowej w przypadku sieci z jednym stanowiskiem o rozkładzie czasu obsługi zdecydowanie róznym od wykładniczego ($C_b > 2$) oraz wariant [Mari 79] metod równoważnego stanowiska w przypadku większej liczby stanowisk z $C_b > 2$. Z reguły błąd metod rośnie wraz ze wzrostem wartości C_b poszczególnych stanowisk i jest większy przy ocenie długości kolejki i czasu czekania niż przy wyznaczaniu stopnia obciążenia stanowisk. Badaniem dokładności aproksymacji dyfuzyjnej w przypadku otwartej sieci stanowisk zajmuje się [Reis 74].

Niektóre metody wymagają nie tylko dużego nakładu obliczeń, lecz także dużego i pracochłonnego oprogramowania. Rodzi to konieczność tworzenia biblioteki programów, która ma zazwyczaj, np. [Best 78, Vera 84, Saue 82], postac zorganizowanego pakietu zawierającego moduły obliczeniowe odpowiadające poszczególnym metodom. Dane wejściowa - konfiguracja sieci, parametry stanowisk, wybór metody, instrukcje sterujące - są przedstawiane w specjalnie przystosowanym do tego celu języku i analizowane przez program wypełniający wspólną dla wszystkich modułów bazę danych. Następnie wywoływany jest odpowiedni moduł obliczeniowy. Oprogramowanie to nie tylko ułatwia praktykę modelowania, lecz stwarza sposobność do dalszych prac poznawczych, w tym - do oceny dokładności i zakresu stosowania metod przybliżonych.

Rozdział II

APROKSYMACJA DYFUZYJNA - POJĘCIA PODSTAWOWE

Teoretycznym uzasadnieniem aproksymacji dyfuzyjnej są twierdzenia graniczne dla mocno obciążonych stanowisk obsługi, stwierdzające, iż odpowiednio unormowane procesy $N(t)$, $U(t)$ są w tych stanowiskach zbieżne do procesu dyfuzji. Przytoczymy poniżej najważniejsze z tych twierdzeń, omówiwszy wcześniej podstawowe pojęcia z nimi związane. Ograniczymy się tu w zasadzie do pojedynczych stanowisk typu G/G/1.

2.1. Aproksymowane procesy $N(t)$, $U(t)$

Przypomnijmy niektóre wprowadzone w poprzednim rozdziale oznaczenia.

W chwilach t_1, t_2, t_3,..., tzn. w odstępach czasu $a_1 : t_2-t_1$, $a_2 = t_3-t_2$,..., nadchodzą do stanowiska obsługi klienci, k-ty klient jest obsługiwany przez czas b_k.

Odstępy $\{a_n, n = 1,2,...\}$ tworzą ciąg niezależnych zmiennych opisanych rozkładem $A(x)$ o średniej $\frac{1}{\lambda}$ i wariancji σ_a^2, podobnie $\{b_n, n = 1,2,...\}$ tworzą ciąg niezależnych zmiennych opisanych rozkładem $B(x)$ o średniej $\frac{1}{\mu}$ i wariancji σ_b^2. Ciągi $\{a_n\}$ i $\{b_n\}$ są niezależne.

$N(t)$ jest liczbą klientów obecnych w chwili t w stanowisku obsługi, $\{N(t), t > 0\}$ jest więc procesem przypadkowym przyjmującym wartości ze zbioru liczb całkowitych nieujemnych, zwiększającym swą wartość o 1 w momentach nadejścia nowego klienta i zmniejszającym wartość o 1 w momentach zakończenia obsługi klienta.

$U(t)$ jest wirtualnym czasem czekania, zdefiniowanym [Bene 63] następująco: jeżeli $I(t)$ jest liczbą klientów przybyłych do systemu w przedziale $[0,t)$, a $L(t)$ jest ich łącznym czasem obsługi,

$$L(t) = \sum_{i=1}^{I(t)} b_i$$

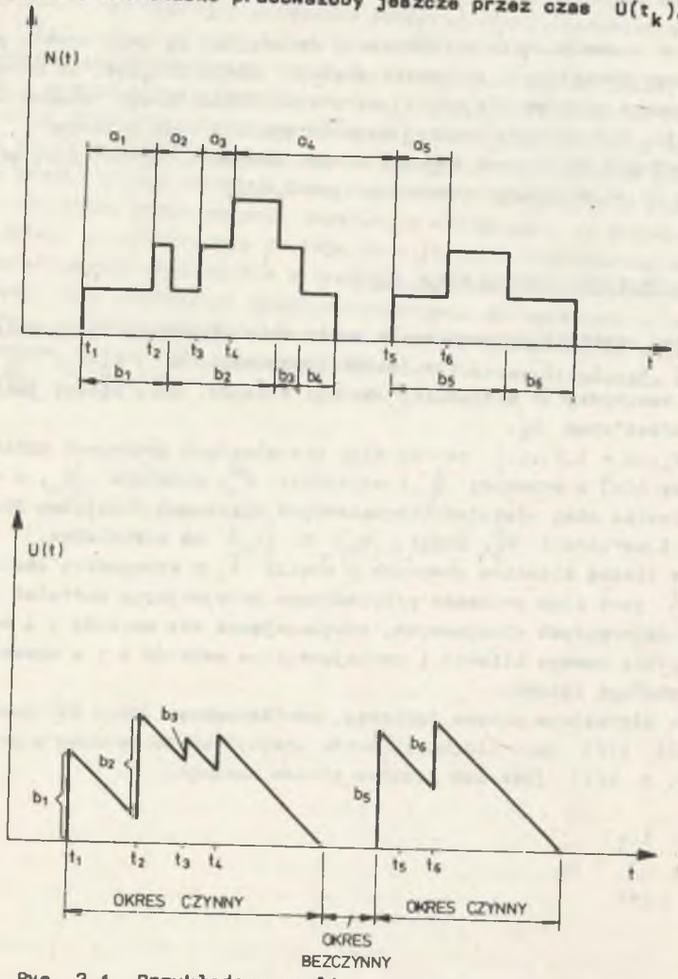
oraz

$$V(t) = I(t) - t,$$

to

$$U(t) = V(t) - \inf_{0\le s\le t}[V(s), 0].$$

W chwilach nadejścia klienta wartość procesu $\{U(t),\ t \geq 0\}$ zwiększa się skokowo o czas jego obsługi - zmienną losową podlegającą rozkładowi B(x), poza tymi momentami wartość procesu maleje liniowo, jeżeli $U(t) > 0$ lub, jeżeli $U(t) = 0$, pozostaje zerowa do momentu najbliższego skoku. U(t) nazywa się niekiedy "zapasem pracy" - gdyby w chwili t_k przestali nadchodzić klienci, stanowisko pracowałoby jeszcze przez czas $U(t_k)$. Jeżeli



Rys. 2.1. Przykładowe realizacje procesów N(t) i U(t)
Fig. 2.1. Exemplary runs of stochastic processes N(t) and U(t)

regulamin obsługi kolejki jest naturalny, klient który przychodzi do stanowiska w chwili t_s, będzie czekał na obsługę przez czas $U(t_{s-0})$.

Przykładowe przebiegi procesów N(t), U(t) przedstawia rys. 2.1.

2.2. Aproksymujący proces X(t)

Proces $\{X(t),\ t \geq 0\}$, którym zastępuje się N(t) bądź U(t) jest jednowymiarowym procesem dyfuzji - zdefiniowanym na osi rzeczywistej, a więc przyjmującym wartości w przestrzeni stanów $\mathcal{X} = \{x : -\infty < x < \infty\}$ jednorodnym procesem Markowa o ciągłym parametrze czasowym i ciągłych trajektoriach.

Teoria procesów dyfuzji, znajdująca swój początek w opracowanych przez Einsteina i Smoluchowskiego [Smol 13, 15] modelach ruchu drobin ulegających zderzeniom, rozwinęła się w dyscyplinę matematyczną dzięki pracom m.in. Wienera [Wien 23], Kołmogorowa [Kołm 31], Levy'ego [Levy 48] i Itô [Itô 65].

Poniżej przytoczymy jedynie niezbędne definicje.

Niech

$$F(x,t;\ y,\tau) = P[X(t) < x | X(\tau) = y], \qquad t > \tau$$

oznacza prawdopodobieństwo przejść procesu.

Zakładamy, że $F(x,t;y,\tau)$ jest dla ustalonego τ i y ciągłą funkcją t i że jest warunkową dystrybuantą względem x, dla której

$$\lim_{x\to-\infty} F(x,t;y,\tau) = 0$$

$$\lim_{x\to\infty} F(x,t;y,\tau) = 1 \ .$$

Ponadto

$$\lim_{t\to\tau+0} F(x,t;y,\tau) = \lim_{\tau\to t-0} F(x,t;y,\tau) = \begin{cases} 0 & x \leq y \\ 1 & x > y \end{cases} .$$

Zakładamy też, że istnieje funkcja gęstości prawdopodobieństwa

$$f(x,t;y,\tau) = \frac{\partial F(x,t;y,\tau)}{\partial x}$$

spełniająca warunki

$$F(x,t;y,\tau) = \int_{-\infty}^{y} f(z,t;y,\tau)dz \ .$$

$$\int_{-\infty}^{\infty} f(z,t;y,\tau)dz = 1 \quad ,$$

$$f(x,t;y,\tau) = \int_{-\infty}^{\infty} f(x,t;z,s)f(z,s;y,\tau)dz \quad .$$

Warunek ujmujący markowowskie własności procesu jest wyrażony przez równanie Chapmana-Kołmogorowa

$$F(x,t;y,\ ) = \int_{-\infty}^{\infty} F(x,t;z,s)d_z F(z,s;y,\tau), \quad s \in (\tau,t) \ . \tag{2.1}$$

Z równania (2.1) wynikają [Kołm 31, Gned 62] równania różniczkowe:

- odwrotne (retrospektywne) równanie Kołmogorowa

$$-\frac{\partial f(x,t;y,\tau)}{\partial \tau} = \frac{\alpha(y,\tau)}{2}\frac{\partial^2 f(x,t;y,\tau)}{\partial y^2} + \beta(y,\tau)\frac{f(x,t;y,\tau)}{\partial y} \ , \tag{2.2}$$

- proste (prospektywne) równanie Kołmogorowa

$$\frac{\partial f(x,t;y,\tau)}{\partial t} = \frac{1}{2}\frac{\partial^2[\alpha(x,t)f(x,t;y,\tau)]}{\partial x^2} - \frac{\partial[\beta(x,t)f(x,t;y,\tau)]}{\partial x} \ , \tag{2.3}$$

to ostatnie spełnione również przez $F(x,t;y,\tau)$.

Przy ich wyprowadzeniu zakłada się, że

(i) $$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-y|\geq\delta} d_x F(x,t+\Delta t;y,t) = \lim_{\Delta t\to 0}\int_{|y-z|\geq\delta} d_x F(x,t;y,t-\Delta t) = 0$$

dla $\delta > 0$, tj. że prawdopodobieństwo

$$P\left[|X(t) - X(\tau)| > \delta\right] \quad \text{przy} \quad X(\tau) = y$$

w czasie infinitezymalnego przedziału Δt jest znikome w porównaniu z Δt.

(ii) jeżeli $X(t-\Delta t) = y$, to średnie i wariancja zmian procesu w następującym potem przedziale czasu Δt są wyrażone odpowiednio przez

$$\int_{-\infty}^{\infty}(x-y)d_x F(x,t;y,t-\Delta t)$$

i

$$\int_{-\infty}^{\infty}(x-y)^2 d_x F(x,t;y,t-\Delta t).$$

Ponieważ powyższe całki mogą być rozbieżne, definiuje się obcięte momenty, zwane infinitezymalną średnią i wariancją zmian $X(t)$

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-y|<\delta}(x-y)d_x F(x,t;y,t-\Delta t) =$$

$$= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-y|<\delta}(x-y)d_x F(x,t)+\Delta t;y,t) = \beta(y,\tau),$$

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-y|<\delta}(x-y)^2 d_x F(x,t;y,t-\Delta t) =$$

$$= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-y|<\delta}(x-y)^2 d_x F(x,t+\Delta t;y,t) = \alpha(y,\tau) \ .$$

Występujące w (2.2), (2.3) funkcje β, α nazywa się odpowiednio współczynnikiem dryftu i współczynnikiem dyfuzji.

(iii) $$\frac{\partial F(x,t;y\tau)}{\partial y}, \ \frac{\partial^2 F(x,t;y,\ )}{\partial y^2}, \ \frac{\partial^2[\alpha(x,t)f(x,t;y,\tau)]}{\partial x^2},$$

$$\frac{\partial[\beta(x,t)f(x,t;y,\tau)]}{\partial x}$$

istnieją i są ciągłe.

Opartą na teorii półgrup operatorów dyskusję istnienia i jednoznaczności rozwiązań równań Kołmogorowa zawierają [Fell 52, Hill 49, Yosi 57].

W niniejszej pracy będziemy wykorzystywać równanie proste (2.3), nazywane też równaniem Fockera-Plancka, z niezależnymi od czasu funkcjami $\beta(x)$, $\alpha(x)$.

W najprostszym przypadku stałych współczynników β, α i gdy proces jest określony na całej osi liczb rzeczywistych, rozwiązanie równania

$$\frac{\partial f(x,t;\Psi)}{\partial t} = \frac{\alpha}{2}\frac{\partial^2 f(x,t;\Psi)}{\partial x^2} - \beta\frac{\partial f(x,t;\Psi)}{\partial x} \ , \tag{2.4}$$

gdzie Ψ oznacza warunek początkowy, $f(x,0;\Psi) = \Psi(x)$, za pomocą separacji zmiennych lub wykorzystując transformaty Fouriera nie nastręcza trudności i ma postać (por. np. [Mand 68])

$$f(x,t;\Psi) = \frac{1}{\sqrt{2\pi\alpha t}} \int_{-\infty}^{\infty} e^{-(y-x+\beta t)^2/2\alpha t} \Psi(y)dy.$$

2.3. Twierdzenia graniczne dla stanowiska G/G/1

Podane poniżej twierdzenia, wybrane spośród wielu jako bezpośrednio nas interesujące, zostały sformułowane przez Igleharta i Whitte'a [Igle 70, Igle 73]. Obszerną monografię tematu stanowi [Boro 80], zwięzły przegląd rezultatów zawierają [Whit 73], [Lern 77].

Przypomnijmy pojęcie słabej zbieżności miar prawdopodobieństwa i wielkości losowych.

Niech dane będą: przestrzeń metryczna $\mathbf{S}$ z borelowskimi zbiorami S oraz ciąg miar prawdopodobieństwa $\{P_n, n \geq 1\}$ i miara prawdopodobieństwa P na $(S,\mathbf{S})$. Ciąg $\{P_n, n \geq 1\}$ jest słabo zbieżny do P (co oznaczamy $P_n \Longrightarrow P$) jeżeli

$$\lim_{n\to\infty} \int_S f\, dP_n = \int_S f dP$$

dla wszystkich ograniczonych, ciągłych funkcji f przyjmujących wartości rzeczywiste w S.

Wielkość losowa X, mierzalne odwzorowanie z przestrzeni probabilistycznej (Ω,F,P) w S, określa miarę prawdopodobieństwa P na $(S,\mathbf{S})$ zdefiniowaną jako

$$P(A) = P(X \in A) \quad \text{dla} \quad A \in \mathbf{S}.$$

Ciąg zmiennych losowych $\{X_n, n \geq 1\}$ jest zbieżny do zmiennej losowej X (co oznacza się $X_n \Longrightarrow X$) jeżeli $P_n \Longrightarrow P$.
X_n i X nie muszą należeć do tej samej przestrzeni probabilistycznej, a tylko miary P_n i P muszą należeć do tej samej przestrzeni metrycznej.

Proces dyfuzji z zerowym współczynnikiem dryftu nazywa się procesem Wienera, a jeśli dodatkowo $\alpha = 1$ - standardowym procesem Wienera.

Niech $\xi(t)$ oznacza standardowy proces Wienera, a $\xi_R(t)$ oznacza proces $\xi(t)$ ograniczony do półosi $x > 0$:

$$\xi_R(t) = \xi(t) - \inf[\xi(u),\ 0 \leq u \leq t] .$$

Oznaczmy jeszcze

$$\hat{N}_n = \frac{N(nt) - (\lambda - \mu)nt}{(\lambda^3\sigma_a^2 + \mu^3\sigma_b^2)\sqrt{n}} ,$$

$$\hat{U}_n = \frac{U(nt) - (\rho - 1)nt}{\lambda(\sigma_a^2 + \sigma_b^2)\sqrt{n}} .$$

Twierdzenia 1-3 odnoszą się do stanowiska G/G/1 [Igle 70, 73]

Tw. 1

Jeżeli $\rho > 1$, to $\hat{N}_n \Longrightarrow \xi$, $\hat{U}_n \Longrightarrow \xi$.

Tw. 2

Jeżeli $\rho = 1$, to $\hat{N}_n \Longrightarrow \xi_R$, $\hat{U}_n \Longrightarrow \xi_R$.

Tw. 3

Jeżeli $\rho < 1$, to istnieją $c_1 > 0$, $c_2 > 0$, dla których

$$\frac{N^*(ut) - uL_E[N]}{c_1\sqrt{u}}$$

$$\frac{U^*(ut) - uL_E[N]}{c_2\sqrt{u}}$$

gdzie:

$$N^*(ut) = \int_0^{ut} N(s)ds,$$

$$U^*(ut) = \int_0^{ut} U(s)dx,$$

L_c oznacza liniową funkcję ct.

Tak więc procesy $N(t)$ i $U(t)$ w mocno obciążonym stanowisku G/G/1 mogą być aproksymowane procesem dyfuzji, swobodnym w przypadku $\rho > 1$, ograniczonym do $x > 0$ w przypadku $\rho = 1$, o współczynnikach

$$\beta = \lambda - \mu, \qquad \alpha = \lambda^3\sigma_a^2 + \mu^3\sigma_b^2,$$

gdy chodzi o $N(t)$ oraz

$$\beta = \rho - 1, \qquad \alpha = \lambda(\sigma_a^2 + \sigma_b^2),$$

gdy chodzi o proces $U(t)$.

Podobne twierdzenia udowodniono również w przypadku mocno obciążonych stanowisk z regulaminem priorytetowym oraz dla sieci stanowisk G/G/1, z których jedno jest mocno obciążone [Reim 84]. Można wówczas przeprowadzić redukcję wymiarowości procesów opisujących te systemy i pokazać, że procesem granicznym jest jednowymiarowy proces $\xi(t)$ lub $\xi_R(t)$.

Stanowisko, w którym $\rho \geq 1$ nie posiada stanu ustalonego - liczba klientów i zapas pracy w tym stanowisku rosną z czasem w sposób podlegający fluktuacjom wokół składowej liniowej - tak jak rośnie wartość procesu dyfuzji o dodatnim współczynniku dryftu.

Prawdopodobieństwo pustego stanowiska i zerowego czasu czekania stają się pomijalnie małe.

Inaczej jest w przypadku stanowiska, w którym $\rho < 1$ i które osiąga dla $t \to \infty$ stan równowagi stochastycznej - a właśnie takie stanowiska są z reguły elementami kolejkowych modeli systemów informatycznych. $N(t)$ i $U(t)$ są tu zbieżne do niezdegenerowanych zmiennych losowych i nie wymagają normalizacji. Nadal istnieje, jak wskazuje na to Tw. 3, podobieństwo tych procesów do procesu dyfuzji, lecz ich zastąpienie procesem dyfuzji o tych samych, co w przypadku $\rho \geq 1$ parametrach nie jest oczywiste. Poza tym w stanowisku słabo obciążonym pominięcie prawdopodobieństwa pustego systemu jest dużym błędem - trzeba więc wprowadzić warunki graniczne dla procesu dyfuzji dopuszczające zerową wartość procesu.

2.4. Warunki brzegowe dla procesu dyfuzji

Klasyfikację i dyskusję warunków brzegowych dla równań Kołmogorowa przeprowadził Feller [Fell 52], poniżej przytoczymy trzy ich rodzaje, mające praktyczne znaczenie w dalszym postępowaniu.

Zakładamy, że bariera znajduje się w punkcie $x = r$, a proces rozpoczyna się w punkcie $x_0 > r$.

(i) bariera odbijająca :

Ponieważ proces na pewno nie przekroczy wartości $x = r$, więc

$$\int_r^\infty f(x,t;\Psi)dx = 1,$$

czyli

$$\frac{\partial}{\partial t}\int_r^\infty f(x,t;\Psi)dx = \int_r^\infty \frac{\partial f(x,t;\Psi)}{\partial t}\,dx = 0,$$

skąd warunek brzegowy dla równania (2.5)

$$\lim_{x\to r}\left\{\frac{1}{2}\frac{\partial}{\partial x}\left[\alpha(x)f(x,t;\Psi)\right] - \beta(x)f(x,t;\Psi)\right\} = 0.$$

(ii) bariera pochłaniająca :

Proces kończy się po osiągnięciu wartości $x = r$.
Warunek brzegowy ma postać

$$f(r,t;\Psi) = f(x,t;r) = 0.$$

(iii) bariera z powrotami elementarnymi :

Proces po osiągnięciu wartości $x = r$ pozostaje w tym punkcie przez czas o rozkładzie wykładniczym z parametrem λ_r, a następnie jest kontynuowany w punkcie $x > r$ wybranym z rozkładem prawdopodobieństwa o funkcji gęstości $g(x)$. Przeskok procesu z punktu $x = r$ do wybranego punktu $x > r$ jest nazywany powrotem elementarnym.

Równania dyfuzji mają postać

$$\frac{\partial f(x,t;\Psi)}{\partial t} = \frac{\partial^2[\alpha(x)f(x,t;\Psi)]}{\partial x^2} - \frac{\partial[\beta(x)f(x,t;\Psi)]}{\partial x} + p_r(t)\lambda_r g(x),$$

$$\frac{dp_r(t)}{dt} = \lim_{x\to r}\left[\frac{\partial[\alpha(x)f(x,t;\Psi)]}{\partial x} - \beta(x)f(x,t;\Psi)\right] - \lambda_r p_r(t),$$

gdzie

$$p_r(t) = P[X(t) = r].$$

Ostatnie równanie przedstawia bilans masy prawdopodobieństw dla $x = r$. Zakłada się ponadto, że poza momentami powrotów elementarnych bariera zachowuje się tak samo jak bariera pochłaniająca:

$$\lim_{x \to r} f(x,t;\psi) = 0.$$

R o z d z i a ł III

APROKSYMACJA DYFUZYJNA LICZBY KLIENTÓW W POJEDYNCZYM STANOWISKU OBSŁUGI

Rozdział ten jest poświęcony aproksymacji dyfuzyjnej procesu $N(t)$ w stanowisku obsługi typu G/G/1/N i G/G/1 - a więc odpowiednio w stanowisku z ograniczoną i nieograniczoną kolejkę. Ponieważ aproksymacja dyfuzyjna stanów ustalonych dla tych stanowisk jest znana [Gele 75], skoncentrujemy się na opisie stanów nieustalonych i metodzie rozwiązywania równań dyfuzji, która także w następnych rozdziałach posłuży do analizy stanów przejściowych stanowiska.

Analityczne modele systemów obsługi w stanie nieustalonym są ograniczone do przypadków, w których ewolucję systemu można opisać procesem urodzin i śmierci i w których znane jest analityczne rozwiązanie odpowiadających temu procesowi równań Chapmana-Kołmogorowa. Niech $p(n,t;i)$ oznacza prawdopodobieństwo, że w chwili t znajduje się w systemie obsługi n klientów, pod warunkiem, że w chwili początkowej było ich i. Jeżeli czas obsługi ma rozkład wykładniczy z parametrem μ, a klienci nadchodzą w odstępach czasu o rozkładzie wykładniczym z parametrem λ i są obsługiwani według kolejności zgłoszenia, tj. jeżeli mamy do czynienia z systemem typu M/M/1, to równania wiążące prawdopodobieństwa stanu systemu mają postać

$$\frac{d}{dt} p(0,t;i) = -\lambda p(0,t;i) + \mu p(1,t;i),$$

$$\frac{d}{dt} p(n,t;i) = -(\lambda+\mu)p(n,t;i) + \lambda p(n-1,t;i) + \mu p(n+1,t;i), \qquad (3.1)$$

$$n = 1,2,\ldots .$$

Rozwiązaniem tego nieskończonego układu równań zajmowało się kilku autorów [Bail 54, Cham 56, Clar 53, Cono 58, Lede 54] stosując różnorakie metody. Wspólny ich wynik, ujęty w najprostszej formie, to

$$p(n,t;i) = e^{-(\lambda+\mu)t}\left[\rho^{\frac{n-i}{2}} I_{n-i}(at) + \rho^{\frac{n-i-1}{2}} I_{n+i+1}(at) + (1-\rho)\rho^n \sum_{k=n+i+2}^{\infty} \rho^{-\frac{k}{2}} I_k(at)\right], \qquad (3.2)$$

gdzie:

$$\rho = \lambda/\mu, \qquad a = 2\mu\sqrt{\rho};$$

funkcja

$$I_k(x) = \sum_{m=0}^{\infty} \frac{(\frac{x}{2})^{k+2m}}{m!(k+m)!}, \quad I_{-k} = I_k$$

jest zmodyfikowaną funkcją Bessela pierwszego rodzaju.

Rozszerzenie modelu M/M/1 na przypadki:

- $\mu(n) = n\mu$ (stanowisko obsługi z nieskończoną liczbą równoległych kanałów),
- $\mu(n) = n\mu$ dla $n \leqslant c$, $\mu(n) = c\mu$ dla $n \geqslant c$ (stanowisko zawierające c równoległych kanałów obsługi)

zawiera praca [Saat 61]; w [More 58] podano rozwiązanie układu (3.1) w przypadku $\lambda(n) = \lambda$ dla $n \leqslant N$, $\lambda(n) = 0$ dla $n > N$ (jednokanałowe stanowisko obsługi z ograniczoną kolejką - stanowisko typu M/M/1/N).

Nawet w tych nielicznych najprostszych przypadkach, w których rozwiązanie dokładne istnieje, ma ono postać dość złożoną i niewygodną w praktycznym wykorzystaniu; wynik (3.2) wymaga obliczania nieskończonego szeregu funkcji, z których każda sama jest nieskończonym szeregiem. Dlatego też liczne są próby zastępowania wzoru (3.2) wynikiem przybliżonym, lecz wymagającym mniej złożonych obliczeń. Kotiah [Koti 78] zastępuje transformatę Laplace'a funkcji tworzącej prawdopodobieństwa $p(n,t;i)$ wyrażeniami, które są funkcjami wymiernymi i których oryginał ma postać prostszą od (3.2). Rider [Ride 76] pokazuje, że dynamika zmian wartości długości kolejki w stanowisku M/M/1 może być w przybliżeniu opisana nieliniowym równaniem różniczkowym pierwszego rzędu. Stern [Ster 79] zastępuje (3.2) rozwiązaniem dla systemu M/M/1/N z odpowiednio dużym N. Jones i in. [Jone 80] proponują uproszczenie obliczeń w (3.2) przez zastąpienie szeregów funkcji Bessela inną, łatwiejszą do obliczenia funkcją.

3.1. Aproksymacja dyfuzyjna stanowiska G/G/1/N i G/G/1 - stan ustalony

Proces $N(t)$ w stanowisku G/G/1/N będziemy aproksymować procesem dyfuzji $X(t)$ z dwiema barierami umieszczonymi w $x = 0$ i $x = N$.

W przedziale $x \in (0,N)$ $X(t)$ jest procesem dyfuzji o współczynnikach

$$\beta = \lambda - \mu, \quad \alpha = \lambda^3\sigma_1^2 + \mu^3\sigma_b^2 = \lambda c_a + \mu c_b.$$

Po osiągnięciu punktu $x = 0$ proces pozostaje w tym punkcie przez czas, który jest zmienną losową o funkcji gęstości $l_0(t)$; bezpośrednio po tym proces rozpoczyna się w punkcie $x = 1$; podobnie, gdy proces znajdzie się w punkcie $x = N$, pozostaje w nim przez czas określony funkcją gęstości prawdopodobieństwa $l_N(t)$, a następnie rozpoczyna się w punkcie $x = N-1$.

Jeżeli $l_0(t)$ i $l_N(t)$ mają postać wykładniczą, $l_0(t) = \lambda_0 e^{-\lambda_0 t}$, $l_N(t) = \lambda_N e^{-\lambda_N t}$, to proces $X(t)$ można opisać równaniami [Fell 54, Bhar 60]

$$\frac{\partial f(x,t;\Phi)}{\partial t} = \frac{\alpha}{2}\frac{\partial^2 f(x,t;\Phi)}{\partial x^2} - \frac{\partial f(x,t;\Phi)}{\partial x} +$$

$$+ \lambda_0 p_0(t)\delta(x-1) + \lambda_N p_N(t)\delta(x-N+1),$$

$$\frac{dp_0(t)}{dt} = \lim_{x\to 0}\left[-\beta f(x,t;\Phi) + \frac{\alpha}{2}\frac{\partial(x,t;\Phi)}{\partial x}\right] - \lambda_b p_0(t), \tag{3.3}$$

$$\frac{dp_N(t)}{dt} = \lim_{x\to N}\left[\beta f(x,t;\Phi) - \frac{\alpha}{2}\frac{\partial f(x,t;\Phi)}{\partial x}\right] - \lambda_N p_N(t),$$

gdzie $f(x,t;\Phi)$ jest warunkową gęstością prawdopodobieństwa procesu:

$$f(x,t;\Phi) = P\left[x < X(t) \leqslant x + dx \middle| f(x,0;\Phi) = \Phi(x)\right].$$

$p_0(t)$, $p_N(t)$ określają prawdopodobieństwo, że w chwili t proces znajduje się odpowiednio w punkcie $x = 0$ lub $x = N$; $\delta(x)$ jest funkcją delta Diraca.

Typ rozkładów $l_0(t)$, $l_N(t)$ nie ma wpływu na rozwiązanie równań (3.3) w stanie ustalonym, które zależy tylko od wartości średnich tych rozkładów [Gele 80]. Metoda określania $f(x,t;\Phi)$, którą przedstawimy w podrozdziale 3.3, obowiązuje dla dowolnych rozkładów $l_0(t)$, $l_N(t)$.

W stanie ustalonym, gdy $f(x)$, p_0, p_N nie są funkcjami czasu, rozwiązania układu równań ma postać [Gele 75]

$$f(x) = \begin{cases} \frac{\lambda_0 p_0}{-\beta}(1 - e^{zx}) & \text{dla } 0 < x \leqslant 1 \\ \frac{\lambda_0 p_0}{-\beta}(e^{-z} - 1)e^{zx} & \text{dla } 1 < x \leqslant N-1 \\ \frac{\lambda_N p_N}{-\beta}\left[e^{z(x-N)} - 1\right] & \text{dla } N-1 \leqslant x < N, \end{cases} \tag{3.4}$$

$$z = 2\beta/\alpha.$$

Po uwzględnieniu warunku normalizacyjnego i warunku ciągłości funkcji $f(x)$ otrzymujemy

$$p_o = \lim_{t\to\infty} p_o(t) = \left\{ 1 + \rho e^{z(N-1)} + \frac{\rho}{1-\rho}\left[1 - e^{z(N-1)}\right]\right\}^{-1}$$

$$p_N = \lim_{t\to\infty} p_N(t) = \rho p_o e^{z(N-1)} .$$

Do modelowania procesu $N(t)$ w stanowisku G/G/1 z nieograniczoną kolejką potrzebna jest tylko jedna bariera, umieszczona w $x = 0$, z której powraca proces do $x = 1$; rozwiązania (3.4) przyjmuje wówczas postać

$$f(x) = \begin{cases} \frac{\lambda_o p_o}{-\beta}(1 - e^{zx}) & \text{dla } 0 < x \leq 1 , \\ \frac{\lambda_o p_o}{-\beta}(e^{-z} - 1)e^{zx} & \text{dla } x \geq 1, \end{cases} \qquad (3.5)$$

$$p_o = \frac{\beta}{\beta - \lambda_o}.$$

Jeżeli przyjąć $\lambda_o = \lambda$, tzn. że średni okres pobytu w barierze jest równy średniemu odstępowi czasu między nadchodzącymi klientami, to

$$p_o = 1 - \rho,$$

co jest wynikiem dokładnym.

3.2. Czas pierwszego przejścia w procesie dyfuzji i okres czynny stanowiska G/G/1

Rozważmy proces dyfuzji z barierą pochłaniającą w $x = 0$. Proces $X(t)$ rozpoczęty w $x_o > 0$ kończy się w momencie osiągnięcia wartości zerowej. Czas, po którym to nastąpi nazywamy czasem pierwszego przejścia z punktu x_o do punktu $x = 0$. Rozwiązanie $f(x,t;\psi)$ równania dyfuzji z warunkami brzegowymi

$$f(0,t;\psi) = 0$$

i

$$\lim_{x\to\infty} f(x,t;\psi) = 0$$

uzyskane po sprowadzeniu tego równania do postaci kanonicznej i zastosowaniu metody odbicia, por. [Cox 65], ma postać

$$f(x,t;\psi) = \begin{cases} \psi(x) & \text{dla } t = 0 \\ \frac{1}{\sqrt{2\pi\alpha t}} e^{\frac{\beta}{\alpha}(x-1) - \frac{\beta^2}{2\alpha}t} \int_0^\infty \left[e^{-\frac{(x-\xi)^2}{2\alpha t}} - e^{-\frac{(x+\xi)^2}{2\alpha t}} \right] \psi(\xi)d\xi & \text{dla } t > 0 . \end{cases} \qquad (3.6)$$

Okres czynny stanowiska obsługi to czas, przez który to stanowisko pracuje nieprzerwanie. Rozpoczyna się on nadejściem pierwszego po okresie bezczynnym klienta i trwa dopóki $N(t) \neq 0$ (por. rys. 2.1). Jeżeli proces dyfuzji aproksymuje proces $N(t)$ w G/G/1, to czas przejścia z punktu $x = 1$ do $x = 0$ odpowiada czasowi trwania okresu czynnego stanowiska. Wykorzystując rozwiązanie

$$f(x,t;1) = \begin{cases} \delta(x-1) & \text{dla } t = 0 \\ \frac{1}{\sqrt{2\pi\alpha t}} e^{\frac{\beta}{\alpha}(x-1) - \frac{\beta^2}{2\alpha t}} \left[e^{-\frac{(x-1)^2}{2\alpha t}} - e^{-\frac{(x+1)^2}{2\alpha t}} \right] & \text{dla } t > 0 \end{cases} \qquad (3.7)$$

obliczamy przybliżoną gęstość $\gamma_d(t)$ i dystrybuantę $\Gamma_d(t)$ czasu trwania okresu czynnego [Czac 81]:

$$\gamma_d(t) = \lim_{x\to 0}\left[-\beta f(x,t;1) + \frac{\alpha}{2}\frac{\partial f(x,t;1)}{\partial x}\right] = \frac{\alpha}{2}\lim_{x\to 0}\frac{\partial f(x,t;1)}{\partial x} =$$

$$= \frac{1}{\sqrt{2\pi\alpha t^3}} e^{-\frac{(\beta t+1)^2}{2\alpha t}}, \qquad (3.8)$$

$$\Gamma_d(t) = \int_0^t \gamma_d(t)dt = 1 - \phi\left(\frac{\beta t+1}{\sqrt{\alpha t}}\right) + e^{-\frac{2\beta}{\alpha}}\phi\left(\frac{\beta t+1}{\sqrt{\alpha t}}\right), \qquad (3.9)$$

gdzie

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}}dx.$$

Gęstość $\gamma_d(t)$ wyraża strumień masy prawdopodobieństwa, która wpływa z półosi $x > 0$ do bariery pochłaniającej.

Transformata Laplace'a $\gamma_d(t)$

$$\bar{\gamma}_d(s) = e^{-\frac{\beta+\sqrt{\beta^2+2\alpha s}}{\alpha}}$$

pozwala łatwo obliczyć momenty

$$E[\gamma_d] = -\left.\frac{d\bar{\gamma}_d(s)}{ds}\right|_{s=0} = \frac{1}{\beta},$$

$$E[\gamma_d^2] = \left.\frac{d^2\bar{\gamma}_d(s)}{ds^2}\right|_{s=0} = -\frac{\alpha}{\beta^3} + \frac{1}{\beta^2} \qquad (3.10)$$

oraz wariancję

$$\sigma^2_{\gamma_d} = \frac{1}{\beta^3}.$$

Wyniki (3.8), (3.9) mają prostszą postać i, jak się za chwilę okaże, są dokładniejsze niż zaproponowana wcześniej przez Heymana [Heym 74] aproksymacja $\gamma(t)$, która wykorzystuje proces dyfuzji, odnoszący się nie do N(t), lecz do procesu U(t). W początkowym momencie okresu czynnego zapas pracy to czas obsługi pierwszego klienta, a okres czynny to w przybliżeniu czas przejścia dla opisującego U(t) procesu dyfuzji z punktu x określonego rozkładem czasu obsługi B(x) do punktu x = 0. Gęstość czasu pierwszego przejścia do zera ma wówczas postać zależną od gęstości b(x) [Heym 74]

$$\gamma_H(t) = \frac{1}{\sqrt{2\pi\alpha t^3}} \int_0^\infty x\, e^{-\frac{(x+\beta t)}{2\alpha t}}\, b(x)dx \qquad (3.11)$$

i w przypadku np. wykładniczego rozkładu obsługi, $b(x) = \mu e^{-\mu x}$, (3.11) przybiera postać [Heym 74]

$$\gamma_H(t) = \frac{\mu}{\sqrt{2\pi\alpha t^3}} \int_0^\infty \exp\left\{-\mu x + t(\lambda-\mu)^2/4\lambda t + \mu x\right\}dx =$$

$$= \frac{\mu\rho}{\sqrt{\pi\rho\mu t}} \exp\left[-\tfrac{1}{4}(1-\rho)^2\mu t/\rho\right] -$$

$$- \tfrac{1}{2}(3\rho-1)\mu \exp\left[(2\rho-1)\mu t\right] \operatorname{Erfc}\left[(3\rho-1)\sqrt{\mu/\rho}\right],$$

gdzie

$$\operatorname{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du$$ jest dopełniającą funkcją błędu.

Współczynniki β, α w (3.11) mają oczywiście wartość określoną dla aproksymacji dyfuzyjnej U(t).

Powróćmy teraz do procesu dyfuzji aproksymującego N(t) i obliczmy jeszcze jeden rozkład charakterystyczny dla okresu czynnego w G/G/1.

Rozkład maksymalnej długości kolejki w okresie czynnym

Prawdopodobieństwu, że liczba klientów obecnych jednocześnie w stanowisku obsługi w czasie okresu czynnego nie osiągnie nigdy wartości m odpowiada prawdopodobieństwo, że proces dyfuzji między punktem startowym $x_0 = 1$ a ekranem pochłaniającym nie dojdzie do punktu o współrzędnej $x = m$.

Zapiszmy

$$\gamma_d(t) = g(t,x_m) + \int_0^t k(\tau,x_m)\gamma_m(t-\tau)d\tau, \qquad (3.12)$$

gdzie:

$g(t,x_m)$ - gęstość prawdopodobieństwa, że proces zakończy się w chwili t nie przechodząc przez punkt $x_m > x_0$,

$k(\tau,x_m)$ - gęstość prawdopodobieństwa, że cząstka w chwili $\tau < t$ znajduje się po raz pierwszy w punkcie x_m,

$\gamma_m(t-\tau)$ - gęstość prawdopodobieństwa, że dokładnie w czasie $t-\tau$ proces przejdzie z punktu x_m do początku układu.

Zgodnie z (3.8)

$$\gamma_m(t-\tau) = \frac{x_m}{\sqrt{2\pi\alpha(t-\tau)^3}}\, e^{-\frac{[\beta(t-\tau)+x_m]^2}{2\alpha(t-\tau)}}. \qquad (3.13)$$

Szukamy dystrybuanty $H_d(x_m)$ określającej prawdopodobieństwo, że proces nie osiągnie wartości x_m.

$$H_d(x_m) = \int_0^\infty g(t,x_m)dt = \lim_{s\to 0} \bar{g}(s,x_m) =$$

$$= \lim_{s\to 0}\left[\bar{\gamma}(s) - \bar{k}(s,x_m)\bar{\gamma}_m(s)\right] = 1 - \lim_{s\to 0} \bar{k}(s,x_m).$$

Funkcję $k(t,x_m)$ wyznaczymy po rozwiązaniu równania dyfuzji z warunkami brzegowymi w postaci dwóch barier pochłaniających, jednej w $x = 0$, drugiej w $x = x_m$ i z warunkiem początkowym $x_0 = 1$. Rozwiązanie ma postać (por. [Cox 65])

$$f(x,t;1) = \frac{1}{\sqrt{2\pi\alpha t}} \sum_{n=-\infty}^{\infty} \left\{ \exp\left[\frac{\beta x'_n}{\alpha} - \frac{(x-1-x'_n-\beta t)^2}{2\alpha t} \right] \right.$$

$$\left. - \exp\left[\frac{x''_n}{\alpha} - \frac{(x-1-x''_n-\beta t)^2}{2\alpha t} \right] \right\}, \qquad (3.14)$$

gdzie

$$\left.\begin{aligned} x'_n &= 2n\,x_m \\ x''_n &= 2(x_m-1) - 2n \end{aligned}\right\} \quad n = 0, \pm 1, \pm 2, \ldots$$

Gęstość $k(t,x_m)$ jest równa masie prawdopodobieństwa wypływającej z przedziału $(0,x_m)$ do bariery umieszczonej w punkcie $x = x_m$:

$$k(t,x_m) = -\lim_{x\to x_m} \left[-\beta f(x,t;1) + \frac{\alpha}{2} \frac{\partial}{\partial x} f(x,t;1) \right] =$$

$$= -\frac{\alpha}{2} \lim_{x\to x_m} \frac{\partial}{\partial x} f(x,t;1),$$

$$\bar{k}(s,x_m) = -\frac{\alpha}{2} \lim_{x\to x_m} \frac{\partial}{\partial x} \bar{f}(x,s;1),$$

czyli

$$H_d(x_m) = 1 + \frac{\alpha}{2} \lim_{x\to x_m} \frac{\partial}{\partial x} \lim_{s\to 0} \bar{f}(x,s;1). \qquad (3.15)$$

Uwzględniając w (3.12) postać transformaty Laplace'a $f(x,t;1)$

$$\bar{f}(x,s;1) = \frac{e^{-\frac{\beta}{\alpha}(x-1)}}{\sqrt{2\alpha s+\beta^2}} \sum_{n=-\infty}^{\infty} \left\{ \exp\left[-\frac{|x-1-x'_n|}{\alpha} \sqrt{2\alpha s+\beta^2} \right] \right.$$

$$\left. - \exp\left[-\frac{|x-1-x''_n|}{\alpha} \sqrt{2\alpha s+\beta^2} \right] \right\}, \qquad (3.16)$$

otrzymujemy

$$H_d(x_m) = \frac{1 - e^{\frac{2\beta}{\alpha}(x_m-1)}}{1 - e^{\frac{2\beta}{\alpha}x_m}}, \qquad x_m > 1 \qquad (3.17)$$

dystrybuantę przybliżonego rozkładu maksymalnej długości kolejki w okresie czynnym stanowiska G/G/1.

Porównajmy ten wynik z rezultatem dokładnym, znanym w przypadku stanowiska M/M/1 [Cohe 69]. Prawdopodobieństwo, że w czasie całego okresu czynnego liczba klientów n obecnych jednocześnie w stanowisku będzie mniejsza od n_m wynosi

$$H(n_m) = P[n < n_m] = \frac{1 - \rho^{n_m-1}}{1 - \rho^{n_m}}, \qquad n_m = 2,3,\ldots \qquad (3.18)$$

Wyniki numeryczne

Sprawdźmy na przykładach obliczeniowych dokładność powyższych aproksymacji, wykorzystujących proces dyfuzji z barierą pochłaniającą.

W przypadku poissonowskiego strumienia wejściowego, tj. dla stanowiska M/G/1, znane jest (np. [Klei 75]) równanie określające w sposób uwikłany transformatę gęstości czasu trwania okresu czynnego $\bar{\gamma}(s)$

$$\bar{\gamma}(s) = \bar{b}\left\{ \lambda[1 - \bar{\gamma}(s)] + s \right\}, \qquad (3.19)$$

z którego obliczamy

$$E[\gamma] = \frac{1/\mu}{1-\rho},$$

$$E[\gamma^2] = \frac{E[b^2]}{(1-\rho)^3}. \qquad (3.20)$$

Obliczone wcześniej (3.10) wartości $E[\gamma_d]$, $E[\gamma_d^2]$, podobnie zresztą jak wartości $E[\gamma_H]$, $E[\gamma_H^2]$, są dla stanowiska M/G/1 dokładne.

Dla wykładniczego $b(x)$ znana jest w sposób jawny czasowa postać $\gamma(t)$:

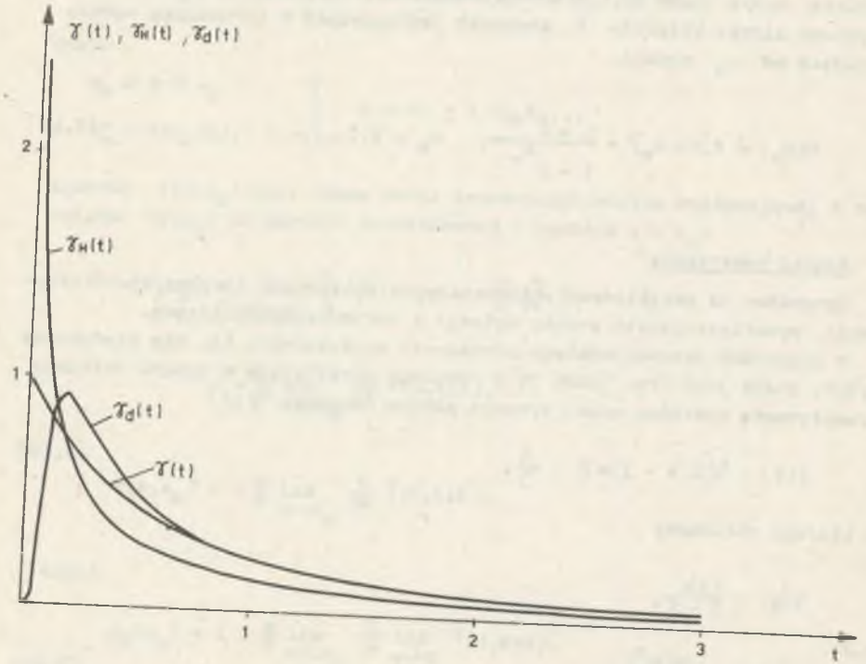
$$\gamma(t) = \frac{1}{t\sqrt{\rho}} e^{-(\lambda+\mu)t} I_1(2t\mu\rho),$$

gdzie

$$I_1(x) = \frac{x}{2} \sum_{k=0}^{\infty} \frac{1}{k!(k+1)!} \left(\frac{x}{2}\right)^{2k}$$

jest zmodyfikowaną funkcją Bessela pierwszego rzędu, pierwszego rodzaju.

Rys. 3.1 przedstawia funkcję dokładną $\gamma(t)$ oraz przybliżone gęstości $\gamma_d(t)$, $\gamma_H(t)$ rozkładu okresu czynnego w stanowisku M/M/1.



Rys. 3.1. $\gamma(t)$, $\gamma_d(t)$, $\gamma_H(t)$ w stanowisku M/M/1, $\rho = 0{,}95$, $\mu = 1$

Fig. 3.1. $\gamma(t)$, $\gamma_d(t)$, $\gamma_H(t)$ in a service station M/M/1; $\rho = 0{,}95$, $\mu = 1$

Tabela 3.I zawiera wartości dokładne dystrybuanty $\Gamma(t)$ oraz wartości błędów

$$\varepsilon_d(t) = \frac{\Gamma_d(t) - \Gamma(t)}{\Gamma(t)}, \quad \varepsilon_H(t) = \frac{\Gamma_H(t) - \Gamma(t)}{\Gamma(t)}$$

Tabela 3.I

Wartości dokładne dystrybuanty $\Gamma(t)$ dla stanowiska M/M/1, ($\lambda = \rho$, $\mu = 1$) i błędy popełnione przez jej aproksymacje $\Gamma_H(t)$, $\Gamma_d(t)$

t	$\rho = 0{,}50$			$\rho = 0{,}95$		
	$\Gamma(t)$	$\varepsilon_H(t)$	$\varepsilon_d(t)$	$\Gamma(t)$	$\varepsilon_H(t)$	$\varepsilon_d(t)$
0,01	0,0099	6,7902	-1,0000	0,0099	9,234	1,000
0,1	0,0929	1,4572	-0,8535	0,0910	1,996	-0,7346
1	0,5475	0,0607	0,0164	0,4828	0,1870	0,0068
2	0,7324	-0,0168	0,0127	0,6252	0,0683	0,0048
3	0,8197	-0,0258	0,0071	0,6946	0,0353	0,0028
4	0,8697	-0,0234	0,0042	0,7372	0,0211	0,0018
5	0,9017	-0,0191	0,0026	0,7667	0,0136	0,0013
6	0,9236	-0,0148	0,0017	0,7886	0,0091	0,0010
8	0,9511	-0,0083	0,0007	0,8196	0,0043	0,0006
10	0,9671	-0,0043	0,0002	0,8409	0,0019	0,0005
13	0,9807	-0,0010	-0,0001	0,8631	0,0002	0,0003
16	0,9881	0,0004	-0,0001	0,8788	0,0007	0,0002
	0,9934	0,0009	-0,0002	0,8939	0,0014	0,0002

jej aproksymacji dla średnio i mocno obciążonego stanowiska M/M/1. Obliczona tu wartości $\Gamma_H(t)$ nie pokrywają się z wartościami podanymi w [Heym 74] - są dokładniejsze. W tabeli 3.II zestawiono wartości dokładne i przybliżone rozkładu maksymalnej długości, jaką kolejka stanowiska M/M/1 osiąga w czasie okresu czynnego.

Dla badanych przypadków widoczna jest przewaga aproksymacji $\gamma_d(t)$, $\Gamma_d(t)$ nad $\gamma_H(t)$, $\Gamma_H(t)$ i stosunkowo duża - z wyjątkiem małych wartości t - ich dokładność.

Tabela 3.II

Dokładny i przybliżony rozkład maksymalnej długości kolejki w stanowisku M/M/1

	$\rho = 0.50$		$\rho = 0.95$	
n	H(n)	$H_d(n)$	H(n)	$H_d(n)$
2	0,66667	0,66076	0,512821	0,512818
3	0,85714	0,85166	0,683611	0,683607
4	0,93333	0,92923	0,768894	0,768890
5	0,96774	0,96494	0,819974	0,819970
6	0,98413	0,98231	0,853953	0,853949
7	0,99212	0,99100	0,878160	0,878156
8	0,99607	0,99540	0,896260	0,896255
9	0,99804	0,99764	0,910288	0,910254
10	0,99902	0,99879	0,921467	0,921463

3.3. Stan nieustalony i dynamika stanowisk G/G/1/N i G/G/1

Znamy już rozwiązania (3.6), (3.11) równania dyfuzji z jedną i dwiema barierami pochłaniającymi. Wykorzystamy obecnie te rozwiązania do opisu dyfuzji z procesami powrotów elementarnych. Dla uniknięcia niejednoznaczności procesa dyfuzji z barierami (barierą) pochłaniającymi będziemy nazywać procesem $X_p(t)$, "pojedynczym", w odróżnieniu od procesu $X(t)$ z powrotami elementarnymi, który będziemy rozpatrywać jako superpozycję procesów pojedynczych. Funkcje gęstości procesów $X_p(t)$ i $X(t)$ będziemy oznaczać odpowiednio przez $\varphi(t)$ i $f(t)$.

Stanowisko G/G/1/N

W przypadku dwóch barier pochłaniających w $x = 0$ i $x = N$ dla procesu $X_p(t)$ rozpoczętego w x_o, $X_p(0) = x_o$, rozwiązanie

$$\varphi(x,t;x_o) = \begin{cases} \delta(x-x_o) & \text{dla} \quad t = 0 \\ \dfrac{1}{\sqrt{2\pi\alpha t}} \displaystyle\sum_{n=-\infty}^{\infty} \left\{ \exp\left[\frac{\beta x'_n}{\alpha} - \frac{(x-x_o-x'_n-\beta t)^2}{2\alpha t}\right] + \right. & (3.21) \\ \left. - \exp\left[\frac{\beta x''_n}{\alpha} - \frac{(x-x_o-x''_n-\beta t)^2}{2\alpha t}\right] \right\} & \end{cases}$$

$x'_n = 2nN$, dla $t > 0$

$x''_n = 2(N - x_o) - x'_n$,

określa dla przedziału (0,N) rozkład jednostkowej masy prawdopodobieństwa, która umieszczona początkowo w całości w punkcie x_o, rozprzestrzenia się w przedziale (0,N) oraz zaczyna gromadzić się w barierach, by w końcu dla $t \to \infty$ znaleźć się w nich całkowicie. Strumień masy prawdopodobieństwa przepływający w chwili t z przedziału (0,N) do dolnej bariery to

$$\gamma_{x_o,0}(t) = -\lim_{x\to 0}\left[-\beta\varphi(x,t;x_o) + \frac{\alpha}{2}\frac{\partial}{\partial x}\varphi(x,t;x_o)\right] = -\frac{\alpha}{2}\lim_{x\to 0}\frac{\varphi(x,t;x_o)}{\partial x} \tag{3.22}$$

Podobnie strumień masy prawdopodobieństwa przepływający w chwili t z przedziału (0,N) do górnej bariery ma wartość

$$\gamma_{x_o,N}(t) = \frac{\alpha}{2}\lim_{x\to N}\frac{\partial\varphi(x,t;x_o)}{\partial x}. \tag{3.23}$$

Jeżeli przyjmiemy warunek początkowy w postaci ciągłego rozkładu $\Phi(x)$, $0 < x < N$, to

$$\varphi(x,t;\Phi(x)) = \int_{0+}^{N-} \varphi(x,t;\xi)\Phi(\xi)d\xi, \tag{3.24}$$

a strumienie prawdopodobieństwa $\gamma_{\Phi,0}(t)$, $\gamma_{\Phi,N}(t)$ mają postać analogiczną do (3.22), (3.23).

Powróćmy do procesu $X(t)$ z powrotami elementarnymi i załóżmy, że w chwili początkowej masa prawdopodobieństwa jest częściowo rozmieszczona w przedziale (0,N) zgodnie z rozkładem $\Phi(x)$, a część jej znajduje się w ekranach, tak że

$$p_o(0) + \int_{0+}^{N-} \Phi(x)dx + p_N(0) = 1.$$

Proces $X(t)$ można wyrazić poprzez superpozycję procesów $X_p(t)$:

- procesu rozpoczętego w chwili $t = 0$ z warunkiem początkowym $\Phi(x)$,
- procesów rozpoczętych w dowolnej chwili τ, $0 \leq \tau \leq t$, w punktach $x = 1$ i $x = N - 1$.

Gęstość prawdopodobieństwa $f(x,t;\Phi)$ procesu $X(t)$ to

$$f(x,t;\Phi) = \varphi(x,t;\Phi) + \int_0^t g_1(\tau)\varphi(x,t-\tau;1)d\tau + \int_0^t g_{N-1}(\tau)\varphi(x,t-\tau;N-1)d\tau \tag{3.25}$$

Gęstości prawdopodobieństwa $g_1(\tau)$, $g_{N-1}(\tau)$ rozpoczęcia procesów $X_p(t)$ w punktach $x = 1$ i $x = N-1$ w chwili τ, czyli strumienie masy prawdopodobieństwa wypływające z ekranów i pojawiające się na osi w punktach $x = 1$ i $x = N-1$, są określone przez równania

$$g_1(\tau) = \int_0^\tau \vartheta_0(t)l_0(\tau-t)dt,$$
$$g_{N-1}(\tau) = \int_0^\tau \vartheta_N(t)l_N(\tau-t)dt. \tag{3.26}$$

$\vartheta_0(t)$, $\vartheta_N(t)$ - całkowite strumienie mas prawdopodobieństwa wpływające w chwili t do ekranów - można wyrazić jako

$$\vartheta_0(t) = p_0(0)\delta(t) + \vartheta_{\Phi,0}(t) + \int_0^t g_1(\tau)\vartheta_{1,0}(t-\tau)d\tau + \int_0^t g_{N-1}(\tau)\vartheta_{N-1,0}(t-\tau)d\tau,$$
$$\vartheta_N(t) = p_N(0)\delta(t) + \vartheta_{\Phi,N}(t) + \int_0^t g_1(\tau)\vartheta_{1,N}(t-\tau)d\tau + \int_0^t g_{N-1}(\tau)\vartheta_{N-1,N}(t-\tau)d\tau. \tag{3.27}$$

Postać równań (3.26), (3.27) upraszcza się nieco, gdy poddamy je transformacji Laplace'a

$$\bar{g}_1(s) = \bar{\vartheta}_0(s)\bar{l}_0(s), \quad \bar{g}_{N-1}(s) = \bar{\vartheta}_N(s)\bar{l}_N(s), \tag{3.28}$$

$$\bar{\vartheta}_0(s) = p_0(0) + \bar{\vartheta}_{\Phi,0}(s) + \bar{g}_1(s)\bar{\vartheta}_{1,0}(s) + \bar{g}_{N-1}(s)\bar{\vartheta}_{N-1,0}(s),$$
$$\bar{\vartheta}_N(s) = p_N(0) + \bar{\vartheta}_{\Phi,N}(s) + \bar{g}_1(s)\bar{\vartheta}_{1,N}(s) + \bar{g}_{N-1}(s)\bar{\vartheta}_{N-1,N}(s), \tag{3.29}$$

gdzie

$$\bar{f}(s) = \int_0^\infty e^{-st}f(t)dt.$$

Powyższe zależności można zobrazować konstruując graf przedstawiony na rys. 3.2 i porównując rozchodzenie się strumieni mas prawdopodobieństwa w układzie złożonym z przedziału $(0,N)$ oraz dwóch barier do rozchodzenia się sygnału w przedstawionym grafie.

Określmy funkcje przejścia gałęzi tego grafu.

Rozkłady $l_0(t)$, $l_N(t)$ reprezentujące czas trwania okresów bezczynności i okresów, w których kolejka jest zapełniona, nie są znane w modelu analitycznym stanowiska G/G/1/N. Przyjmiemy, że są one równe odpowiednio rozkładowi czasów między zgłoszeniami i rozkładowi czasów obsługi, pamiętając że jest to przybliżenie. Jest więc $\bar{l}_0(s) = \bar{a}(s)$, $\bar{l}_N(s) = \bar{b}(s)$. Podkreślmy, że $l_0(t)$, $l_N(t)$ są rozkładami dowolnymi, a nie wykładniczymi.

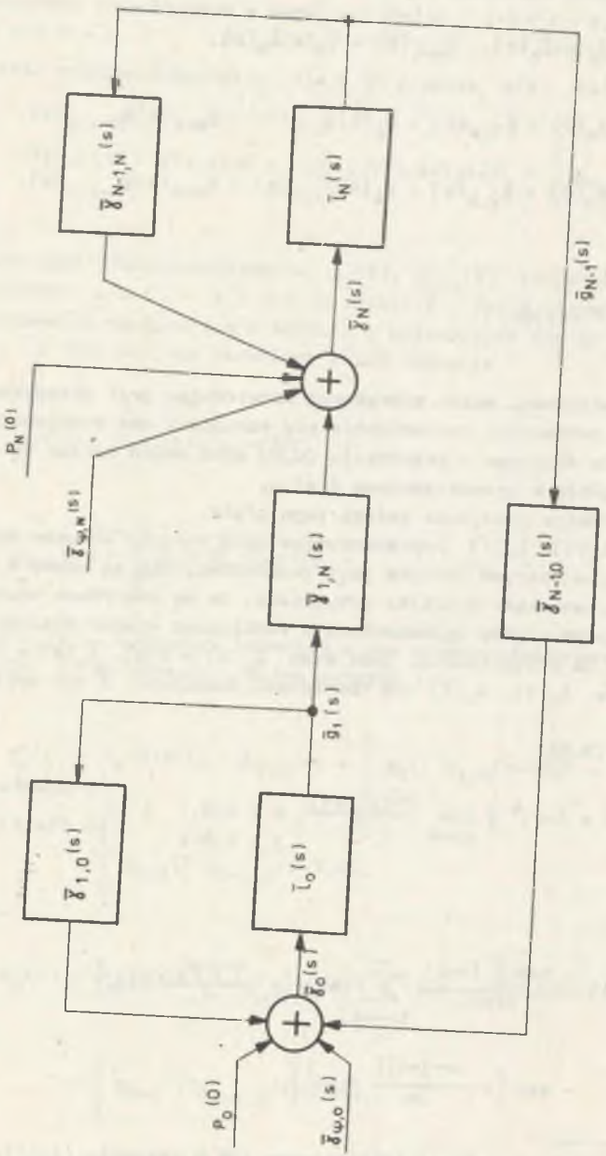
Zgodnie z (3.23)

$$\bar{\vartheta}_{j,k}(s) = (-1)^i \frac{\alpha}{2} \lim_{x\to k} \frac{\partial\bar{\varphi}(x,s;j)}{\partial x}, \quad \begin{matrix} k = 0,N, \\ j = 1,N-1 \end{matrix} \quad i = \begin{cases} 1 \text{ dla } k = 0 \\ 0 \text{ dla } k = N \end{cases}$$

gdzie

$$\bar{\varphi}(x,s;j) = \frac{\exp\frac{\beta}{\alpha}(x-1)}{A(s)} \sum_{i=-\infty}^{\infty} \left\{ \exp\left[-\frac{x-j-x_i'}{\alpha}A(s)\right] + \right.$$
$$\left. - \exp\left[-\frac{|x-j-x_i''|}{\alpha}A(s)\right] \right\}, \tag{3.30}$$

$A(s) = \sqrt{\beta^2 + 2\alpha s}$; x_i', x_i'' są zdefiniowane jak w równaniu (3.21).



Rys. 3.2. Graficzna interpretacja równań obiegu masy prawdopodobieństwa
Fig. 3.2. A graphic interpretation of the probability mass circulation

Można więc obliczyć $\bar{g}_1(s)$ i $\bar{g}_{N-1}(s)$:

$$\bar{g}_1(s) = \left\{ p_0(0) + \bar{\gamma}_{\phi,0}(s) + \left[p_N(0) + \bar{\gamma}_{\phi,N}(s) \right] \frac{\bar{l}_N(s)\bar{\gamma}_{N-1,0}(s)}{1 - \bar{l}_N(s)\bar{\gamma}_{N-1,N}(s)} \right\} \cdot$$

$$\cdot \frac{\bar{l}_0(s)}{1-\bar{l}_0(s)\bar{\gamma}_{1,0}(s)} \left[1 - \frac{\bar{l}_0(s)\bar{\gamma}_{1,N}(s)}{1-\bar{l}_0(s)\bar{\gamma}_{1,0}(s)} \frac{\bar{l}_N(s)\bar{\gamma}_{N-1,0}(s)}{1-\bar{l}_N(s)\bar{\gamma}_{N-1,N}(s)} \right]^{-1}, \qquad (3.31)$$

$$\bar{g}_{N-1}(s) = \bar{g}_1(s) \frac{\bar{\gamma}_{1,N}(s)\bar{l}_N(s)}{1-\bar{l}_N(s)\bar{\gamma}_{N-1,N}(s)}$$

i transformatę gęstości prawdopodobieństwa procesu X(t)

$$\bar{f}(x,s) = \bar{\varphi}(x,s;\psi) + \bar{g}_1(s)\bar{\varphi}(x,s;1) + \bar{g}_{N-1}(s)\bar{\varphi}(x,s;N-1). \qquad (3.32)$$

Prawdopodobieństwo, że proces w chwili t ma wartość x = 0, czyli przybliżone prawdopodobieństwo pustego systemu obsługi w chwili t, to

$$p_0(t) = \int_0^t \gamma_0(t)dt$$

lub

$$\bar{p}_0(s) = \frac{1}{s}\bar{\gamma}_0(s). \qquad (3.33)$$

Stanowisko G/G/1

N(t) jest aproksymowany procesem dyfuzji z jedną barierą, umieszczoną w x = 0; w punkcie x = 0 proces zatrzymuje się przez czas zdefiniowany rozkładem $l_0(t)$ i przeskakuje następnie do x = 1. Proces $X_p(t)$ rozpoczyna się w dowolnym $x_0 > 0$ i kończy po osiągnięciu zera. Gęstość tego procesu jest rozwiązaniem równania dyfuzji z jedną barierą pochłaniającą w x = 0 i ma postać (3.6) przy warunku początkowym $f(x,0;\psi(x)) = \phi(x)$, $x > 0$ lub (3.7) przy warunku początkowym $f(x,0;x_0) = \delta(x-1)$. W tym ostatnim przypadku strumień masy prawdopodobieństwa wpadający do ekranu w x = 0 jest określony przez (3.8)

$$\gamma_{1,0}(t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{(\beta t+1)^2}{2\alpha t}}. \qquad (3.34)$$

Transformaty Laplace'a funkcji $f(x,t;1)$, $\gamma_{1,0}(t)$ mają znaną postać

$$\bar{\varphi}(x,s;1) = \frac{\exp\left[\frac{\beta}{\alpha}(x-1)\right]}{A(s)}\left\{\exp\left[-\frac{|x-1|}{\alpha}A(s)\right]\right.$$

$$\left. - \exp\left[-\frac{|x+1|}{\alpha}A(s)\right]\right\}, \qquad (3.35)$$

$$\bar{\gamma}_{1,0}(s) = \exp\left\{-\left[\beta + A(s)\right]/\alpha\right\}. \qquad (3.36)$$

Równania (3.25)-(3.27) przybierają w przypadku stanowiska G/G/1 postać

$$f(x,t;\Psi) = \varphi(x,t;\Psi) + \int_0^t g_1(\tilde{\tau})f(x,t-\tilde{\tau};1)d\tilde{\tau}, \qquad (3.37)$$

$$g_1(\tilde{\tau}) = \int_0^t \gamma_0(t)l_0(\tilde{\tau}-t)dt, \qquad (3.38)$$

$$\gamma_0(t) = p_0(0)\delta(t) + \gamma_{\Psi,0}(t) + \int_0^t g_1(\tilde{\tau})\gamma_{1,0}(t-\tilde{\tau})d\tilde{\tau}, \qquad (3.39)$$

a równania (3.32), (3.28), (3.29) przechodzą w

$$\bar{f}(x,s) = \bar{\varphi}(x,s;\Psi) + \bar{g}_1(s)\bar{\varphi}(x,s;1), \qquad (3.40)$$

$$\bar{g}_1(s) = \bar{\gamma}_0(s)\bar{l}_0(s), \qquad (3.41)$$

$$\bar{\gamma}_0(s) = p_0(0) + \bar{\gamma}_{\Psi,0}(s) + \bar{g}_1(s)\bar{\gamma}_{1,0}(s). \qquad (3.42)$$

Przyjmujemy jak poprzednio $\bar{l}_0(s) = \bar{a}(s)$ i obliczamy

$$\bar{g}_1(s) = \left[p_0(0) + \bar{\gamma}_{\Psi,0}(s)\right]\frac{\bar{l}_0(s)}{1 - \bar{l}_0(s)\bar{\gamma}_{1,0}(s)}, \qquad (3.43)$$

co uwzględnione w (3.40) daje rozwiązanie - funkcję gęstości prawdopodobieństwa procesu $X(t)$ w formie transformaty Laplace'a. W przypadku wykładniczego rozkładu $l_0(x)$ rozwiązanie $\bar{f}(x,s;x_0)$, również w formie transformaty Laplace'a, można otrzymać bez uciekania się do powyższego rachunku mas prawdopodobieństwa [Duda 84]. Obecność w jawnej postaci rozkładu $l_0(x)$ we wzorze (3.43) stwarza możliwość badania wpływu rozkładu

czasu pobytu w barierze, a więc rozkładu trwania okresu bezczynnego stanowiska, na dynamikę zmian długości kolejki.

Aproksymację dyfuzyjną stanów nieustalonych procesu $N(t)$ można rozpatrywać w sposób prostszy za pomocą równania dyfuzji z warunkiem brzegowym w postaci bariery odbijającej w punkcie $x = 0$ [Koba 74b]. Gęstość procesu dyfuzji ma w tym przypadku postać

$$f(x,t;x_0) = \frac{\partial}{\partial x}\left\{\Phi\left((x-x_0-\beta t)/\sqrt{\alpha t}\right) - e^{\frac{2\beta x}{\alpha}}\Phi\left(-(x+x_0+\beta t)/\sqrt{\alpha t}\right)\right\},$$

gdzie $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz$.

Dokładność tej aproksymacji można zwiększyć przesuwając barierę z $x = 0$ do $x = -C_b$ (dla $C_b \leq 1$) [Fili 83, Fili -].

Poniżej rozpatrzymy dokładniej, wraz z wynikami numerycznymi, dwa szczególne przypadki rozwiązania (3.40) - przypadek ładowania i rozładowania kolejki.

Ł a d o w a n i e s t a n o w i s k a G/G/1

Do pustego stanowiska obsługi zaczynają nadchodzić klienci. Rozkład długości kolejki zmienia się w funkcji czasu, dążąc przy $t\to\infty$ do stanu ustalonego, o ile ten istnieje. W chwili $t = 0$ stanowisko jest puste, jest więc $p_0(0) = 1$ oraz $\Psi(x) = 0$ dla $x > 0$,

$$f(x,s;0) = \frac{\bar{a}(s)}{\left[1 - \bar{\gamma}_{1,0}(s)\bar{a}(s)\right]A(s)}\exp\left[\frac{\beta}{\alpha}(x-1)\right]\left\{\exp\left[-\frac{|x-1|}{\alpha}A(s)\right] + \right.$$

$$\left. - \exp\left[-\frac{|x+1|}{\alpha}A(s)\right]\right\} \quad \text{dla } x > 0 \qquad (3.44)$$

oraz

$$\bar{p}_0(s) = \frac{1}{2}\left[1 + \bar{\gamma}_{1,0}(s) - \bar{g}_1(s)\right]. \qquad (3.45)$$

Wartość średnia procesu $X(t)$ w chwili t, czyli przybliżona liczba klientów w systemie w chwili t, to

$$E\left[X(t)\right] = \int_0^\infty xf(x,t;0)dx$$

lub po transformacji

$$E[\bar{X}(s)] = \int_0^\infty x\bar{f}(x,s;0)dx = \lambda\bar{p}_0(s)\frac{s+\beta[1-\bar{\gamma}_{1,0}(s)]}{s^2}. \qquad (3.46)$$

Gdy czas dąży do nieskończoności, rozwiązania (3.44)-(3.46) dążą do znanych rozwiązań stanu ustalonego [Gela 75]:

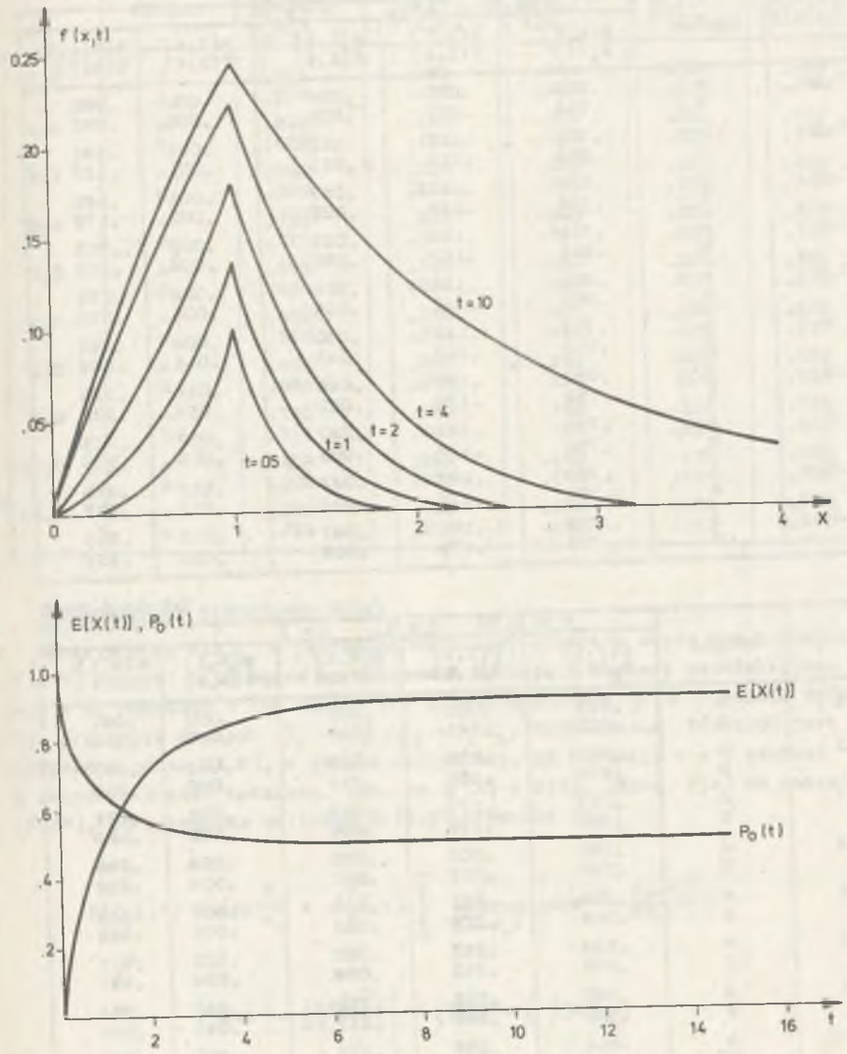
$$f(x) = \lim_{t\to\infty} f(x,t) = \lim_{s\to 0} s\bar{f}(x,s) = \begin{cases} \frac{p_0\lambda}{-\beta}(1 - e^{zx}) \text{ dla } 0 < x \leq 1 \\ \frac{p_0\lambda}{-\beta}(e^{-z} - 1)e^{zx}, \text{ dla } x \geq 1 \end{cases} \qquad (3.47)$$

$$z = 2\beta/\alpha$$

$$p_0 = \lim_{t\to\infty} p_0(t) = \lim_{s\to 0} s\bar{p}_0(s) = \frac{E[l_0]}{E[l_0] + E[\gamma_{1,0}]} = 1 - \rho \qquad (3.48)$$

$$E[X] = \lim_{t\to\infty} E[X(t)] = \lim_{s\to 0} s\,E[\bar{X}(s)] = \frac{1}{2}\rho(1 + \frac{\rho C_a + C_b}{1-\rho}). \qquad (3.49)$$

Niestety, analityczne znalezienia oryginałów transformat $\bar{p}_0(s)$, $\bar{f}(x,s;0)$, $E[\bar{X}(s)]$ wydaje się niemożliwe. W zamieszczonych poniżej przykładach poszukiwano ich numerycznie posługując się algorytmem [Steh 79]. Rys. 3.3 przedstawia przykładowe wyniki tej numerycznej inwersji, a w tabeli 3.III zestawiono wartości $p_0(t)$, $f(i,t;0)$ dla $i = 1,2,3$ oraz $E[X(t)]$ ze znanymi w tym przypadku dokładnymi wartościami $p(i,t)$ i $E[N(t)]$ - przykład dotyczy bowiem stanowiska M/M/1.



Rys. 3.3. Przykładowe rozwiązanie równania dyfuzji podczas ładowania stanowiska; $\lambda = 1$, $\mu = 2$, $C_a = C_b = 1$

Fig. 3.3. An exemplary solution of the diffusion equation - transient states during charging a previously empty service station; $\lambda = 1$, $\mu = 2$, $C_a = C_b = 1$

Tabela 3.IIIa

Porównanie wyników dokładnych (A) i przybliżonych (B) dla stanów nieustalonych w czasie ładowania stanowiska M/M/1

t	Metoda	ρ=0.25 λ=1. μ=4				
		$p(o,t)$ $p_o(t)$	$p(1,t)$ $f(1,t)$	$p(2,t)$ $f(2,t)$	$p(3,t)$ $f(3,t)$	$E[N(t)]$ $E[X(t)]$
0.1	A	.921	.076	.004	.001	.082
	B	.914	.097	.005	.000	.082
0.2	A	.871	.118	.010	.001	.141
	B	.864	.122	.015	.001	.137
0.3	A	.838	.143	.017	.001	.182
	B	.834	.148	.022	.002	.178
0.4	A	.816	.158	.023	.003	.213
	B	.823	.157	.030	.004	.208
0.5	A	.801	.168	.028	.004	.236
	B	.800	.162	.035	.006	.229
1.0	A	.767	.183	.040	.008	.295
	B	.768	.175	.045	.011	.289
2.0	A	.753	.187	.046	.011	.325
	B	.754	.179	.052	.014	.322
5.0	A	.750	.187	.047	.012	.333
	B	.750	.175	.053	.016	.333
10.0	A	.750	.187	.047	.014	.333
	B	.750	.175	.053	.016	.333
15.0	A	.750	.187	.047	.016	.333
	B	.750	.175	.053	.016	.333

Tabela 3.IIIb

t	Metoda	ρ = 0.50 λ = 1, μ = 2				
		$p(o,t)$ $p_o(t)$	$p(1,t)$ $f(1,t)$	$p(2,t)$ $f(2,t)$	$p(3,t)$ $f(3,t)$	$E[N(t)]$ $E[X(t)]$
0.1	A	.913	.083	.004	.000	.091
	B	.907	.121	.004	.000	.090
0.2	A	.848	.138	.013	.001	.167
	B	.838	.159	.016	.000	.165
0.3	A	.797	.177	.024	.002	.231
	B	.789	.196	.030	.002	.228
0.4	A	.756	.203	.035	.004	.286
	B	.749	.219	.041	.004	.282
0.5	A	.726	.221	.045	.007	.335
	B	.719	.228	.052	.008	.329
1.0	A	.634	.257	.082	.021	.501
	B	.632	.255	.088	.024	.497
2.0	A	.565	.263	.111	.042	.694
	B	.566	.256	.113	.045	.685
5.0	A	.516	.255	.124	.058	.896
	B	.517	.248	.124	.060	.889
10.0	A	.503	.251	.125	.062	.975
	B	.504	.249	.125	.064	.972
15.0	A	.501	.250	.125	.062	.993
	B	.502	.246	.125	.064	.992

Tabela 3.IIIc

t	Metoda	ρ = 0.75 λ = 1, μ = 1.333				
		$p(0,t)$ $p_o(t)$	$p(1,t)$ $f(1,t)$	$p(2,t)$ $f(2,t)$	$p(3,t)$ $f(3,t)$	$E[N(t)]$ $E[N(t)]$
0.1	A	.911	.085	.004	.000	.094
	B	.906	.135	.002	.000	.094
0.2	A	.839	.146	.014	.001	.177
	B	.829	.185	.015	.000	.176
0.3	A	.781	.190	.026	.003	.251
	B	.769	.217	.932	.001	.249
0.4	A	.733	.222	.040	.005	.319
	B	.721	.247	.047	.004	.315
0.5	A	.693	.245	.053	.008	.381
	B	.686	.261	.064	.008	.376
1.0	A	.565	.290	.107	.030	.628
	B	.560	.294	.117	.033	.619
2.0	A	.454	.288	.153	.068	.962
	B	.454	.285	.156	.072	.954
5.0	A	.351	.250	.168	.105	1.525
	B	.351	.248	.167	.106	1.516
10.0	A	.302	.222	.160	.112	2.003
	B	.302	.221	.159	.112	1.998
15.0	A	.282	.210	.154	.111	2.275
	B	.283	.208	.153	.111	2.268

Rozładowanie stanowiska G/G/1

Stacjonarny dla $t<0$ strumień wejściowy klientów urywa się w chwili $t = 0$. Proces $N(t)$ można aproksymować dyfuzją z barierą pochłaniającą w $x = 0$. Ponieważ w tym czasie $\lambda = 0$, współczynniki β, α równania dyfuzji przyjmują wartość $\beta_r = -\mu$, $\alpha_r = \mu c_b$. Rozwiązanie $f(x,t;\psi)$ jest określone przez (3.6), a jeżeli przyjmiemy, że w chwili $t = 0$ panował w stanowisku stan ustalony, tzn. że $\psi(x) = f(x)$, gdzie $f(x)$ ma postać (3.47), to wykonanie obliczeń w (3.6) prowadzi do

$$f(x,t;f) = \exp\left(\frac{\beta_r}{\alpha_r}x - \frac{\beta_r^2}{\alpha_r}t\right)\int_0^\infty \frac{1}{\sqrt{2\pi\alpha_r t}}\left\{\exp\left[-\frac{(x-\xi)^2}{2\alpha_r t}\right]\right.$$

$$\left. - \exp\left[-\frac{(x+\xi)^2}{2\alpha_r t}\right]\right\} f(\xi)\exp\left(-\frac{\beta}{\alpha}\xi\right)d\xi =$$

$$= \frac{(1-\rho)}{-\beta}\exp\left(\frac{\beta_r}{\alpha_r}x - \frac{\beta_r^2}{\alpha_r}t\right)\left(\exp\left(-\frac{\beta}{\alpha}x + \alpha_r\frac{\beta^2}{2\alpha}t\right)\right)\left\{\phi\left(\frac{1-x+z_2 t}{\sqrt{\alpha_r t}}\right)\right.$$

$$- \Phi\left(\frac{-x-z_2 t}{\sqrt{\alpha_r t}}\right) + \Phi\left(\frac{1+x-z_2 t}{\sqrt{\alpha_r t}}\right) - \Phi\left(\frac{x-z_2 t}{\sqrt{\alpha_r t}}\right) - \left[e^{-z} - 1\right]$$

$$\cdot \left[1 - \Phi\left(\frac{1+x-z_2 t}{\sqrt{\alpha_r t}}\right)\right] \Bigg\} + \exp\left(\frac{\beta}{\alpha} x + \alpha_r \frac{\beta^2}{2\alpha} t\right) \left\{\Phi\left(\frac{-x-z_2 t}{\sqrt{\alpha_r t}}\right) + \right.$$

$$- \Phi\left(\frac{1-x-z_2 t}{\sqrt{\alpha_r t}}\right) - \Phi\left(\frac{1+x+z_2 t}{\sqrt{\alpha_r t}}\right) - \Phi\left(\frac{x+z_2 t}{\sqrt{\alpha_r t}}\right) +$$

$$\left. + \left[e^{-z} - 1\right] \left[1 - \Phi\left(\frac{1-x-z_2 t}{\sqrt{\alpha_r t}}\right)\right] \right\} \Bigg), \qquad (3.50)$$

gdzie

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad z = \frac{2\beta}{\alpha}, \quad z_2 = \frac{\alpha_r \beta}{\alpha}.$$

β, α są współczynnikami występującymi w funkcji (3.47) określającej stan początkowy.

W praktyce łatwiej jednak obliczyć całkę

$$\bar{f}_r(x,s;f) = \int_0^{\infty} \frac{e^{\beta_r (x-s)/\alpha_r}}{A_r(s)} \left[e^{-\frac{|x-\xi|}{\alpha} A_r(s)} - e^{-\frac{|x+\xi|}{\alpha_r} A_r(s)} \right] f(\xi) d\xi, \qquad (3.51)$$

gdzie $A_r(s) = \sqrt{\beta_r^2 + 2\alpha_r s}$,

a następnie poszukiwać numerycznie oryginałów funkcji

$$\bar{f}_r(x,s;f), \quad E\left[\bar{x}(s)\right] = \int_0^{\infty} x \bar{f}_r(x,s;f) dx$$

i $\quad \bar{p}_0(s) = \frac{1}{s}\left[p_0(0) + \bar{\gamma}\,\phi_{,0}(s)\right]$.

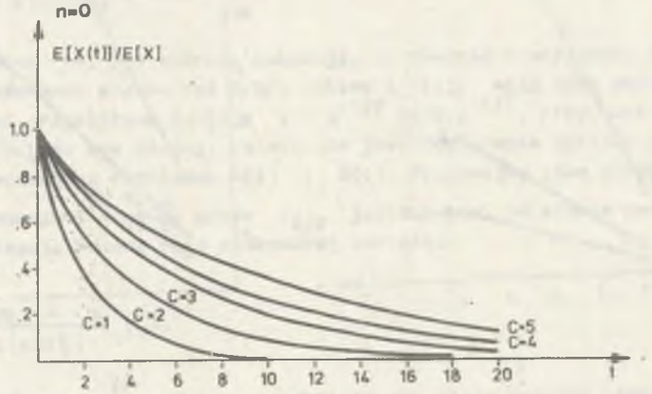
D y n a m i k a ł a d o w a n i a i r o z ł a d o w a n i a s t a n o w i s k a G/G/1

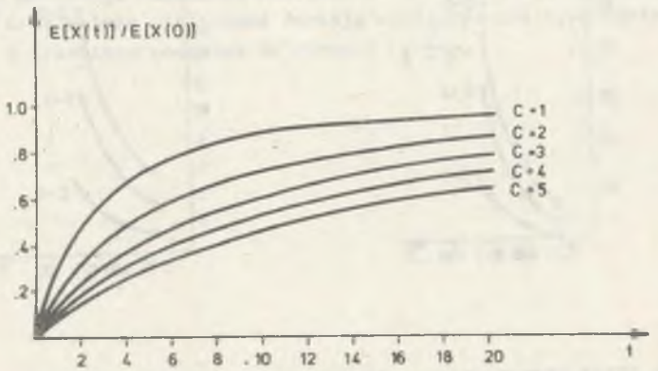
W przypadku stanowiska M/M/1 można prosto przeprowadzić dokładną analizę procesu rozładowania: w chwili początkowej rozkład liczby klientów w stanowisku w stanie równowagi ma postać [Klei 75]

$$p(n,0) = (1-\rho)\rho^n, \qquad n = 0,1,\ldots$$

a średnia liczba klientów w stanowisku, to

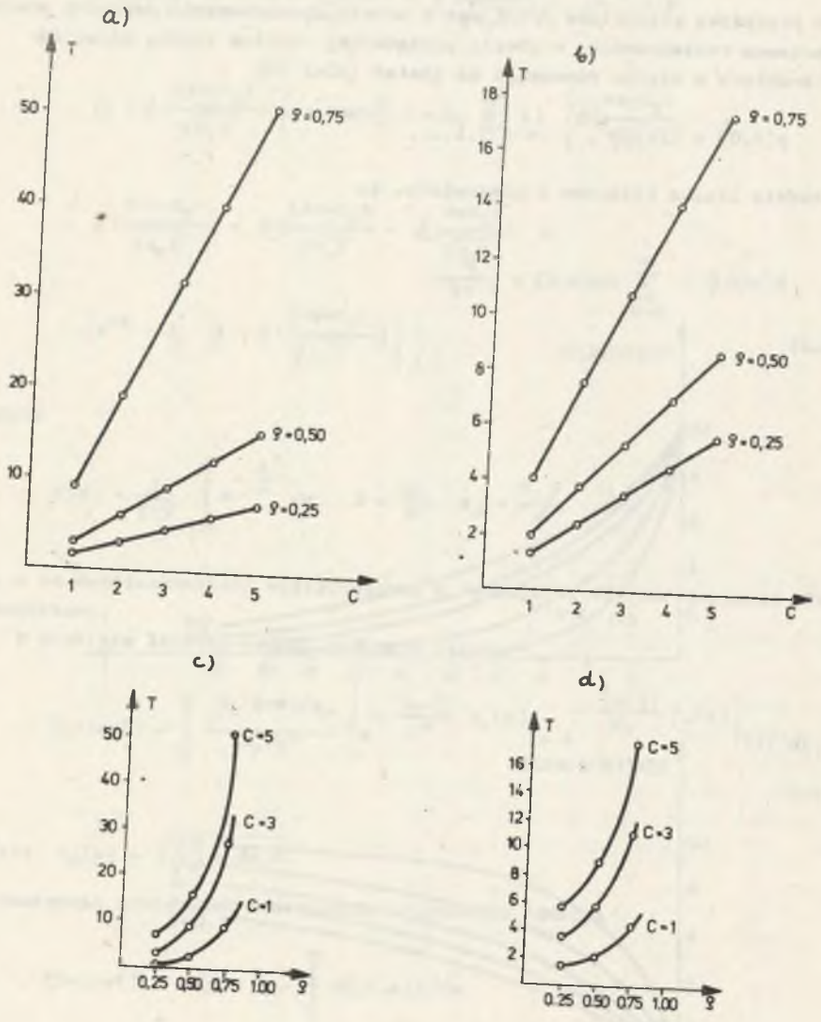
$$E[N(0)] = \sum_{n=0}^{\infty} np(n,0) = \frac{\rho}{1-\rho}.$$





Rys. 3.4. Przybliżona średnia liczba klientów w czasie ładowania (a) i rozładowania (b) stanowiska, wartości znormalizowane; $\lambda = 0,5$, $\mu = 1$, $C_a = C_b = C$

Fig. 3.4. The approximate mean number of customers during the charging (a) and discharging (b) of a service station; the values are normalized; $\lambda = 0,5$, $\mu = 1$, $C_a = C_b = C$



Rys. 3.5. Stała czasowa zmian $E[N(t)]$ przy ładowaniu (a,c) i rozładowaniu (b,d) stanowiska w funkcji $C_a = C_b = C$ i ρ

Fig. 3.5. Time-constant of the changes of $E[N(t)]$ during charging (a,c) and discharging (b,d) of a service station as a function of $C_a = C_b = C$ and ρ

Dla $t > 0$

$$p(n,t) = \sum_{k=0}^{\infty} p(n+k,0)\,\frac{(\mu t)^k}{k!}\,e^{-\mu t} = (1-\rho)\rho^n \sum_{k=0}^{\infty} \rho^k\,\frac{(\mu t)^k}{k!} =$$

$$= (1-\rho)\rho^n e^{-(\mu-\lambda)t} = p(n,0)e^{-(\mu-\lambda)t},$$

czyli

$$E[N(t)] = E[N(0)]\,e^{-(\mu-\lambda)t}.$$

Zamieszczone na rys. 3.4 wykresy wskazują, że również w przypadku ładowania i rozładowania stanowiska G/G/1 zmiany $E[N(t)]$ mają taki charakter, że mogę być przybliżone funkcją $1 - e^{-t/T}$ bądź $e^{-t/T}$, przy czym stan nieustalony trwa tym dłużej, im większa jest obciążenie systemu i im większe są wariancje rozkładów A(t) i B(t). Przyjmijmy jako stałą czasową tych przebiegów $T = \frac{t_{1/2}}{\ln 2}$, gdzie $t_{1/2}$ jest czasem, po którym zmiany $E[X(t)]$ osiągają połowę swej całkowitej wartości:

$$\frac{E[X(t_{1/2})]}{|E[X] - E[X(0)]|} = 0,5.$$

Jak przedstawiono na rys. 3.5, tak zdefiniowana stała czasowa jest liniową funkcją wariancji rozkładów odstępów czasu między zgłoszeniami i czasem obsługi oraz silnie nieliniową funkcją współczynnika wykorzystania stanowiska ρ, o przebiegu podobnym do funkcji $\frac{\rho}{1-\rho}$.

Rozdział IV

APROKSYMACJA DYFUZYJNA WIRTUALNEGO CZASU CZEKANIA W STANOWISKU G/G/1

Twierdzenia graniczne przedstawione w rozdz. II pokazują, że proces dyfuzji nadaje się do przybliżonego opisu zarówno procesu N(t) jak procesu U(t) - wirtualnego czasu czekania w kolejce przed stanowiskiem.

Proces $\{U(t),\ t \geq 0\}$ został opisany w rozdz. II, a rys. 2.1b przedstawia jego przykładową realizację. Przypomnijmy tylko, że U(t) wyraża zapas pracy - czas potrzebny do zakończenia obsługi wszystkich klientów obecnych w stanowisku w chwili t.

Wartość procesu wzrasta skokowo w momencie nadejścia nowego klienta o wartość czasu jego obsługi, a później maleje w miarę postępu obsługi. W momencie zakończenia obsługi ostatniego w okresie czynnym klienta, wartość U(t) dochodzi do zera i zachowuje ją do momentu nadejścia następnego klienta i rozpoczęcia nowego okresu czynnego. O ile wykorzystaniu aproksymacji dyfuzyjnej procesu N(t) poświęcono już co najmniej kilkadziesiąt prac, to znacznie mniej miejsca poświęcono procesowi U(t). Praca Heymana [Heym 74] wykorzystuje dyfuzyjny opis U(t) do analizy okresu czynnego w stanowisku obsługi - w punkcie 3.2 zestawiono jego rezultaty z analogicznymi wynikami uzyskanymi za pomocą aproksymacji procesu N(t).

Poniżej będziemy się odwoływać do pracy Gavera [Gave 68], który proces V(t) w stanowisku M/G/1 aproksymował procesem dyfuzji z barierą odbijającą. Pokażemy, jak wprowadzenie powrotów elementarnych zwiększa dokładność tej aproksymacji.

Ponieważ model Gavera odnosi się do stanowiska M/G/1, a także dlatego, że tylko w przypadku poissonowskiego strumienia zgłoszeń istnieją analityczne wyniki dotyczące U(t), z którymi można porównać wyniki aproksymacji, będziemy dalej mówić o stanowisku M/G/1. Przedstawiony model jest jednak ogólny i odnosi się do stanowiska G/G/1. Jest to tylko kwestia zamiany współczynników β, α w równaniu dyfuzji na współczynniki zgodne z Tw. 1,2 w rozdziale II.

Ponieważ jednym z ograniczeń dokładności aproksymacji dyfuzyjnej jest fakt, iż uwzględnia ona tylko dwa pierwsze momenty rozkładów A(x), B(x), w punkcie 4.4 przedstawiono aproksymację procesu U(t) przy pomocy procesu Markowa, który bierze pod uwagę także trzeci moment rozkładu czasu obsługi, i zbadano dokładność takiego rozwiązania.

4.1. Rezultaty dokładne dotyczące procesu U(t)

Mieszane, tj. uwzględniające zarówno zmiany ciągłe jak i skokowe, procesy Markowa z ciągłym czasem były opisywane przez Kołmogorowa [Kolm 31] i Fellera [Fell 36].

Poniżej zamieszczono dwa najważniejsze rezultaty (i), (ii) dotyczące procesu U(t) w stanowisku M/G/1.

A(x), a(x) oznaczają jak zawsze dystrybuantę i gęstość rozkładu odstępów czasu $a_i = \tau_{i+1} - \tau_i$ między nadejściem kolejnych klientów, a B(x) i b(x) dystrybuantę i gęstość rozkładu czasów obsługi b_i; $a(x) = 1-e^{-\lambda x}$, B(x) jest dowolne.

Oznaczmy przez $F_u(x,t;u_o)$ dystrybuantę

$$F_u(x,t;u_o) = Pr[U(t) \leq x \mid U(0) = u_o]$$

oraz przez $\hat{F}_w(t,t;u_o)$, jej transformatę Laplace'a względem zmiennej x:

$$\hat{F}_w(r,t;u_o) = \int_0^\infty e^{-rx} d_x F_w(x,t;u_o)$$

(będziemy poniżej oznaczać przez $\hat{f}(r,t;u_o)$ i $\bar{f}(x,s;u_o)$ transformaty Laplace'a funkcji $f(x,t;u_o)$ odpowiednio względem zmiennej x i zmiennej t).

(i) $F_u(r,t;u_o)$ jest określona równaniem (4.1)

$$\hat{F}_u(r,t;u_o) = e^{-rt-[1-\hat{b}(r)]\lambda t}\Big\{\hat{F}_u(r,0;u_o) - r\int_0^t e^{-ru+[1-\hat{b}(r)]\lambda u} p_o(u)du\Big\} \tag{4.1}$$

$\bar{p}_o(t) = Pr[U(t) = 0]$ jest określone transformatę

$$\bar{p}_o(s) = \int_0^\infty e^{-st} p_o(t)dt = \frac{\hat{F}_u\{s+\lambda[1-\bar{g}(s)],0,u_o\}}{s+\lambda-\lambda z(s)},$$

gdzie $z = z(s)$ jest pierwiastkiem równania

$$z = \bar{b}(s+\lambda-z)$$

posiadającym najmniejszą wartość bezwzględną wewnątrz jednostkowego koła $|z| < 1$.

Wyprowadzenie (4.1) zawiera [Taká 62b].

(ii) $F_u(x,t;u_o)$ spełnia równania różniczkowo-całkowe

$$\frac{F_u(x,t;u_o)}{\partial t_x} = \frac{F_u(x,t;u_o)}{\partial x} - \lambda\left[F_u(x,t;u_o) - \int_0^x B(x-\xi)\,d_\xi F_u(\xi,t;u_o)\right], \tag{4.2}$$

wyprowadzone w [Taká 62a] i rozwiązane w [Taká 62b] przez Takácsa (por. też dyskusję w [Haso 63]).

Podwójna transformata Laplace'a gęstości $f_u(x,t;u_o) = \frac{dF_u(x,t;u_o)}{dx}$ ma postać

$$\hat{\tilde{f}}_u(r,s;u_o) = \frac{\left(\frac{r}{\eta}\right)^{-\eta u_o} - e^{-r\,u_o}}{\lambda\hat{b}(r) - \lambda + r - s}, \tag{4.3}$$

gdzie $\eta = \eta(s)$ jest jedynym, por. [Bene 56], w obszarze $Re(s) > 0$, $Re(r) > 0$ pierwiastkiem mianownika.

Jeżeli istnieje stan ustalony dla $t \to \infty$, to $\hat{f}(r,s;u_o)$ ma w nim postać

$$\hat{f}_u(r) = \lim_{s\to 0} s\hat{\tilde{f}}_u(r,s;u_o) = \frac{r(1-\rho)}{r - \lambda + \lambda\hat{b}(r)}, \tag{4.4}$$

gdzie $\rho = \lambda/\mu$.

Wzór (4.4) jest znanym od dawna [Poll 30, Khin 32] wzorem Pollaczka-Chinczyna.

4.2. Aproksymacja Gavera

Gaver [Gave 68] zaproponował aproksymację procesu $U(t)$ procesem dyfuzji $X(t)$ o parametrach

$$\beta = \lim_{\Delta t\to 0} \frac{E[X(t+\Delta t) - X(t)|X(t)]}{\Delta t} = \lambda E[b] - 1 + O(\Delta t)$$

$$\alpha = \lim_{\Delta t\to 0} \frac{\mathrm{var}[X(t+\Delta t) - X(t)|X(t)]}{\Delta t} = \lambda E[b^2] + O(\Delta t)$$

z barierą odbijającą umieszczoną w $x = 0$. Ten sam model był rozważany później przez Kleinrocka [Klei 76]. Niech $f_G(x,t;x_o)$ i $F_G(x,t;x_o)$ oznaczają gęstość i dystrybuantę tego procesu.

Rozwiązanie równania dyfuzji

$$\frac{\partial f_G(x,t;x_o)}{\partial t} = -\beta\frac{f_G(x,t;x_o)}{\partial x} + \frac{\alpha}{2}\frac{\partial^2 f_G(x,t;x_o)}{\partial x^2} \tag{4.5}$$

wraz z warunkami brzegowymi w postaci ekranu odbijającego w $x = 0$ oraz

$$\lim_{x\to\infty} f_G(x,t;x_o) = 0$$

i warunkiem początkowym

$$f_G(x,0;x_o) = \delta(x - x_o)$$

jest znane w formie transformaty (por. np. [Chan 43, Prab 65]):

$$\tilde{f}_G(x,s;x_o) = \int_0^\infty e^{-st} f_G(x,t)\,dt = \frac{2\,e^{-\vartheta x_o}}{2\vartheta(\xi_1 - \xi_2)}\left[\xi_1 e^{-\xi_1 x} - \xi_2 e^{-\xi_2 x}\right] + \begin{cases} 0 & \text{jeżeli } x < x_o \\ \frac{2\,e^{\xi_1(x-x_o)}}{(\xi_1 - \xi_2)} + \frac{2\,e^{\xi_2(x-x_o)}}{(\xi_1 - \xi_2)\alpha} & \text{jeżeli } x > x_o \end{cases} \tag{4.6}$$

gdzie:

$$\xi_1 = \xi_1(s) = \frac{\beta}{\alpha}\left[1 + \sqrt{1 + 2\frac{\alpha s}{\beta^2}}\right] = \frac{1}{\alpha}\left[\beta + A(s)\right],$$

$$\xi_2 = \xi_2(s) = \frac{\beta}{\alpha}\left[1 - \sqrt{1 + 2\frac{\alpha s}{\beta^2}}\right] = \frac{1}{\alpha}\left[\beta - A(s)\right],$$

$$A(s) = \sqrt{\beta^2 + 2\alpha s}.$$

Pierwiastki ξ_1, ξ_2 mają przeciwne znaki; ϑ oznacza pierwiastek dodatni:

jeżeli $\rho > 1$ (tj. $\beta > 0$), to $\vartheta \equiv \xi_1$,

jeżeli $\rho < 1$ (tj. $\beta < 0$), to $\vartheta \equiv \xi_2$.

Dla $x_o = 0$, (4.6) przybiera postać

$$\tilde{f}_G(x,s;0) = \begin{cases} -\frac{1}{s}\xi_1 e^{\xi_1 x} & \text{jeżeli } \quad < 1 \\ -\frac{1}{s}\xi_2 e^{\xi_2 x} & \text{jeżeli } \quad > 1, \end{cases} \tag{4.7}$$

a w stanie ustalonym, gdy $t\to\infty$, $f_G(x,t;x_0) \longrightarrow f_G(x)$,

$$f_G(x) = \lim_{s\to 0} s\,\bar{f}_G(x,s) = \begin{cases} \frac{2(1-\rho)}{\rho^2}\exp\left[-\frac{2(1-\rho)}{\rho^2}x\right] & \text{jeżeli } \rho < 1 \\ 0 & \text{jeżeli } \rho > 1\,. \end{cases} \tag{4.8}$$

Wartość średnia

$$E[X] = \int_0^\infty x\, f_G(x)dx$$

jest równa dokładnej wartości średniej czasu czekania w kolejce M/G/1:

$$E[X] = E[U]_{M/G/1} = \frac{E[b^2]}{1-\rho} = E[w],$$

a wariancja procesu dyfuzji jest mniejsza od wariancji czasu czekania,

$$\mathrm{var}[X] < \mathrm{var}[U]_{M/G/1}.$$

4.3. Aproksymacja procesem dyfuzji z powrotami elementarnymi - stan ustalony

Wprowadźmy do procesu dyfuzji powroty elementarne: proces po osiągnięciu wartości zerowej zachowuje ją przez czas o rozkładzie wykładniczym $L_0(t)$ (przyjmiemy, jak poprzednio, że jest to rozkład identyczny z rozkładem $A(t)$), by następnie rozpocząć się w punkcie x określonym przez rozkład $B(x)$ - podobnie jak wirtualny czas czekania, który w okresie bezczynności stanowiska obsługi jest zerowy, a wraz z nadejściem pierwszego klienta w okresie czynnym wzrasta skokowo o wartość określoną rozkładem $B(x)$.

Równania dyfuzji mają postać:

$$\frac{\partial f(x,t;x_0)}{\partial t} = \frac{\alpha}{2}\frac{\partial^2 f(x,t;x_0)}{\partial x^2} - \beta\frac{\partial f(x,t;x_0)}{\partial x} + \lambda p_0(t)b(x)$$

$$\frac{dp_0}{dt} = \lim_{x\to 0}\left[\frac{\alpha}{2}\frac{\partial f(x,t;x_0)}{\partial x} - \beta f(x,t;x_0)\right] - \lambda p_0(t) \tag{4.9}$$

$$\lim_{x\to 0} f(x,t;x_0) = \lim_{x\to\infty} f(x,t;x_0) = 0,$$

gdzie $p_0(t) = P[x(t) = 0]$.

Zachowamy te same, co w aproksymacji Gavera, współczynniki β i α.

Dla stanu ustalonego, w którym $\lim_{t\to\infty} f(x,t;x_0) = f(x)$, $\lim_{t\to\infty} p_0(t) = p_0$, równania (4.9) przybierają postać

$$0 = \frac{\alpha}{2}\frac{d^2f(x)}{dx^2} - \frac{df(x)}{dx} + \lambda p_0 b(x),$$

$$\lim_{x\to 0}\left[\frac{\alpha}{2}\frac{df(x)}{dx} - \beta f(x)\right] = \lambda p_0,$$

$$\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x) = 0, \tag{4.10}$$

skąd po rozwiązaniu otrzymujemy

$$f(x) = \frac{2\lambda p_0}{\alpha} e^{zx}\int [1 - B(x)]\, e^{-zx}dx, \tag{4.11}$$

gdzie $z = 2\beta/\alpha$.

Dla dalszych obliczeń trzeba przyjąć konkretną postać rozkładu $B(x)$. Załóżmy najpierw wykładniczy rozkład czasu obsługi $B(x) = 1 - e^{-\mu x}$. Wówczas, dla $\mu + z \neq 0$,

$$f(x) = \frac{2\lambda p_0}{\alpha}\frac{1}{\mu+z}\left[e^{+zx} - \frac{-\mu x}{\mu}\right] + p_0\delta(x)$$

oraz

$$F(x) = \frac{2\lambda p_0}{\alpha}\frac{1}{\mu+z}\left[1 - e^{zx}) - \frac{1}{\mu}(1 - e^{-\mu x})\right] + p_0, \tag{4.12}$$

gdzie $\delta(x)$ jest funkcją delta Diraca, lub, dla $\mu + z = 0$,

$$f(x) = \frac{2\lambda p_0}{\alpha} x e^{zx} + p_0\delta(x),$$

$$F(x) = \frac{\lambda p_0}{\alpha}\left[e^{zx}(x - \tfrac{1}{z})\right] + \frac{1}{z} + p_0. \tag{4.12a}$$

Po uwzględnieniu warunku normalizacyjnego $\lim_{x\to\infty} F(x) = 1$ otrzymujemy $p_0 = 1 - \rho$ oraz

$$E[X] = \int_0^\infty x dFx = \frac{\rho/\mu}{1-\rho}.$$

a więc dokładne wartości prawdopodobieństwa pustego stanowiska i średniego czasu czekania w stanowisku M/M/1.

Drugi moment rozkładu F(x) ma postać

$$E[x^2] = \frac{4\lambda\rho_0}{\alpha}\frac{1}{\mu+z}\left[\frac{1}{z^3} - \frac{1}{\mu^3}\right] \quad \text{dla} \quad \mu + z \neq 0$$

oraz

$$E[x^2] = \frac{3\lambda\rho_0\alpha^3}{4\beta^4} \quad \text{dla} \quad \mu + z = 0$$

i może przyjmować zarówno wartości większe jak i mniejsze od dokładnej wartości $E[u^2]$ w M/M/1.

Dla rozkładu hiperwykładniczego, reprezentującego rozkłady o dużej wariancji:

$$B(x) = 1 - \sum_{i=1}^{K} b_i e^{-\mu_i x}, \quad \sum_{i=1}^{K} b_i = 1 \tag{4.13}$$

f(x) jest liniową kombinacją rozwiązań (4.12) lub (4.12a).

Dla rozkładu Coxa drugiego rzędu, używanego często do aproksymacji innych rozkładów,

$$b(t) = (1 - a)\mu_1 e^{-\mu_1 t} + a\mu_1 e^{-\mu_1 t} * \mu_2 e^{-\mu_2 t}$$

gdzie * jest operacją splotu,
otrzymujemy po przekształceniu

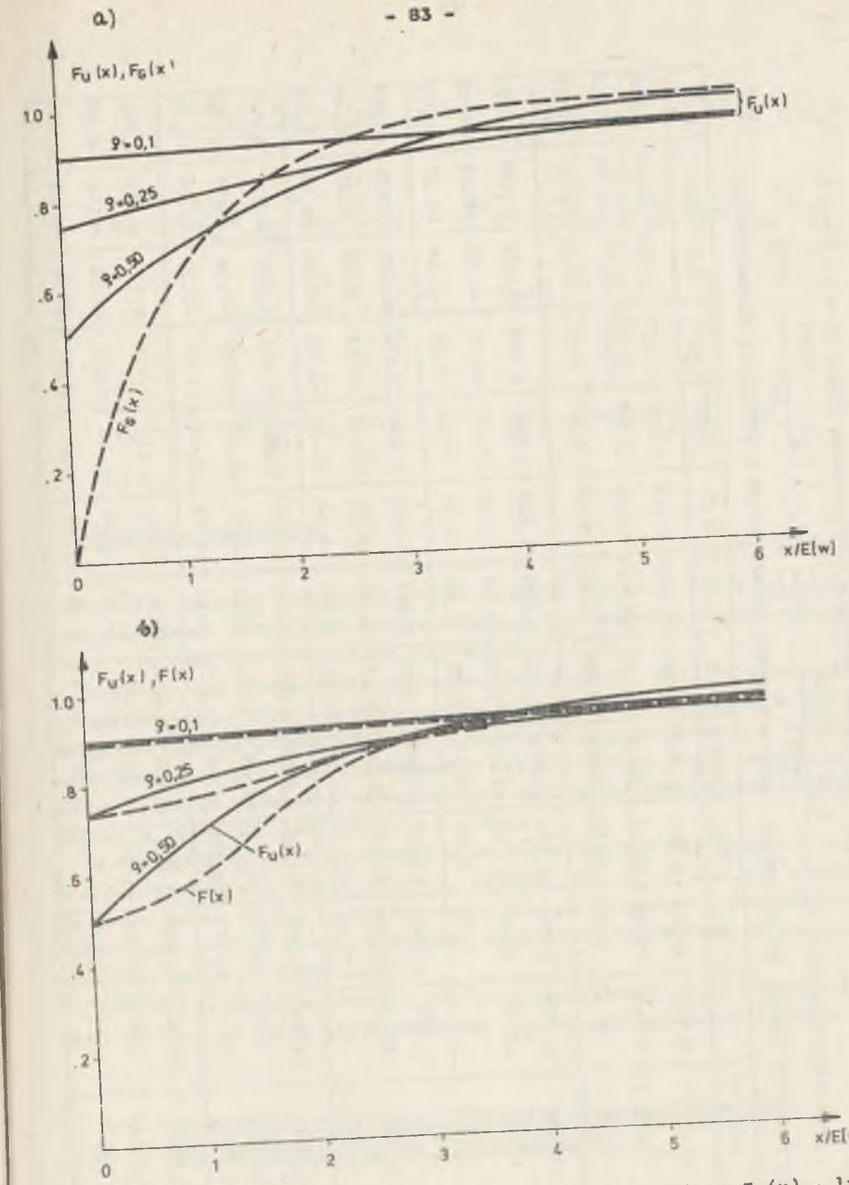
$$b(t) = b_1\mu_1 e^{-\mu_1 t} + b_2\mu_2 e^{-\mu_2 t},$$

gdzie

$$b_1 = (1 - a) - a\frac{\mu_2}{\mu_1 - \mu_2},$$

$$b_2 = a\frac{\mu_1}{\mu_1 - \mu_2} \tag{4.14}$$

i rozwiązanie (4.13) w postaci



Rys. 4.1. Stanowisko M/M/1, ρ = 0,1, 0,25, 0,50; przebieg $F_u(x)$ - linia ciągła - oraz $F_G(x)$, F(x) - linia przerywana

Fig. 4.1. Sation M/M/1, ρ = 0,1, 0,25, 0,50; probability distribution functions of virtual waiting time: $F_u(x)$ - solid curve - exact function; $F_d(x)$, F(x) - broken curves - approximate functions

Tabela 4.I

Wartości dokładne dystrybuanty $F_u(x)$ czasu czekania w stanowisku M/M/1 ($\mu = 1$, $\lambda = \rho$)

oraz błąd względny $\varepsilon_G(x) = \frac{F_u(x) - F_G(x)}{F_u(x)} \cdot 100\%$, $\varepsilon(x) = \frac{F_u(x) - F(x)}{F_u(x)} \cdot 100\%$

wprowadzony przez aproksymację Gavera i przez aproksymację z powrotami elementarnymi

	x/E [w]	0	0,2	0,4	0,6	0,8	1,0	2,0	3,0	4,0	5,0	6,0	7,0
$\rho = 0,10$	$F_u(x)$	0,900	0,902	0,904	0,906	0,908	0,910	0,918	0,926	0,933	0,939	0,945	0,950
	ε_G[%]	100,0	79,90	63,53	50,19	39,33	30,50	5,82	-2,62	-5,22	-5,74	-5,54	-5,13
	ε [%]	0	0,20	0,35	0,46	0,55	0,61	0,71	0,64	0,52	0,41	0,30	0,21
$\rho = 0,25$	$F_u(x)$	0,750	0,762	0,774	0,785	0,795	0,805	0,848	0,882	0,908	0,928	0,944	0,957
	ε_G[%]	100,0	76,20	57,39	42,51	30,76	21,50	-1,92	-7,75	-8,11	-6,99	-5,65	-4,45
	ε [%]	0	1,40	2,35	2,96	3,32	3,48	2,83	1,55	0,51	-0,18	-0,57	-0,75
$\rho = 0,50$	$F_u(x)$	0,500	0,548	0,591	0,630	0,665	0,697	0,816	0,888	0,932	0,959	0,975	0,985
	ε_G[%]	100,0	66,90	44,18	28,34	17,17	9,27	-5,96	-6,95	-5,29	-3,58	-2,30	-1,44
	ε [%]	0	7,09	10,13	10,90	10,41	9,27	2,34	-1,35	-2,35	-2,17	-1,66	-1,16
$\rho = 0,75$	$F_u(x)$	0,250	0,355	0,444	0,522	0,588	0,646	0,833	0,921	0,963	0,982	0,992	0,996
	ε_G[%]	100	48,86	25,81	13,53	6,41	2,11	-3,84	-3,18	-1,98	-1,11	-0,59	-0,30
	ε [%]	0	19,67	19,25	14,80	10,17	6,34	-1,92	-2,51	-1,74	-1,02	-0,56	-0,29
$\rho = 0,90$	$F_u(x)$	0,100	0,248	0,372	0,476	0,562	0,634	0,851	0,934	0,975	0,990	0,996	0,998
	ε_G[%]	100	26,98	11,40	5,12	2,00	0,31	-1,58	-1,14	-0,64	-0,33	-0,16	-0,07
	ε [%]	0	23,61	12,82	6,45	2,99	1,03	-1,38	-1,07	-0,62	-0,32	-0,16	-0,07

$$f(x) = \frac{2\lambda p_0}{\alpha}\left[\frac{b_1}{\mu_1 + z}\left(e^{zx} - e^{-\mu_1 x}\right) + \frac{b_2}{\mu_2 + z}\left(e^{zx} - e^{-\mu_2 x}\right)\right] + p_0\delta(x),$$

$$F(x) = \frac{2\lambda p_0}{\alpha}\left\{\frac{b_1}{\mu_1 + z}\left[-\frac{1}{z}(1 - e^{zx}) - \frac{1}{\mu_1}(1 - e^{-\mu_1 x})\right] + \frac{b_2}{\mu_2 + z}\right.$$

$$\left. \cdot \left[-\frac{1}{z}(1 - e^{zx}) - \frac{1}{\mu_2}(1 - e^{-\mu_2 x})\right]\right\} + p_0,$$

$$p_0 = \left\{1 + \frac{2\lambda}{\alpha}\left[\frac{b_1}{\mu_1 + z}\left(-\frac{1}{z} - \frac{1}{\mu_1}\right) + \frac{b_2}{\mu_2 + z}\left(-\frac{1}{z} - \frac{1}{\mu_2}\right)\right]\right\}^{-1}. \qquad (4.15)$$

Przykłady numeryczne

Rysunek 4.1a przedstawia dystrybuanty $F_u(x) = 1 - \rho e^{-\mu(1-\rho)x}$ czasu czekania w kolejce stanowiska M/M/1 w skali czasu $x = t/E$ [w] przy różnym obciążeniu stanowiska i odpowiadającą im, przy zastosowanej normalizacji tę samą dla wszystkich, aproksymację $F_G(x)$.

Rysunek 4.1b przedstawia ponownie dystrybuanty $F_u(x)$ i odpowiadające im aproksymacje $F(x)$. Wykresy uwidaczniają zalety rozwiązania (4.12), uwzględniającego proces powrotów elementarnych, w porównaniu z rozwiązaniem Gavera, które jako aproksymacja spełnia dobrze swoje zadanie przy dużym (bliskim jedności) współczynniku obciążenia stanowiska (przykłady podane w [Gave 68] zakładają $\rho = 0,95$).

W przypadku słabo obciążonego stanowiska bariera odbijająca w punkcie $x = 0$, sprawiająca, iż $F_G(0) = 0$, podczas gdy $F_u(0) = 1 - \rho$, wprowadza do rozwiązania $F_G(x)$ istotne błędy. Dla aproksymacji z powrotami elementarnymi $F(0) = 1 - \rho$ i stopień obciążenia stanowiska ma znacznie mniejszy wpływ na błędy metody.

Tabela 4.I zawiera wartości dokładne dystrybuant $F_u(x)$ przedstawionych na rys. 4.1 oraz błędy względne ε_G, ε aproksymacji $F_G(x)$ i $F(x)$.

4.4. Aproksymacja U(t) z uwzględnieniem trzeciego momentu rozkładu czasu obsługi - stan ustalony

Trzeci moment rozkładu $B(x)$ ma niekiedy decydujący wpływ na rozkład liczby klientów w stanowisku i rozkład czasu czekania. Szereg przykładów numerycznych dla udokumentowania tego faktu zawiera [Bere 81]. Ograniczenie charakterystyki rozkładu czasu obsługi do pierwszych dwóch momentów jest jednym z zarzutów częściej stawianych aproksymacji dyfuzyjnej.

Jak pokazuje [Klei 76], z równania Chapmana-Kołmogorowa

$$f(x,t;y,\tau) = \int_{-\infty}^{\infty} f(w,u;y,\tau)f(x,t;w,u)dw$$

wynika zależność

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[A_n(x,t)\right], \tag{4.16}$$

gdzie

$$A_n(x,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (y-x)^n f(x,t;y,t+\Delta t)dy. \tag{4.17}$$

Tak więc aproksymacja dyfuzyjna jest szczególnym przypadkiem (4.16), biorącym pod uwagę tylko dwa pierwsze składniki sumy, a dla wybranej aproksymacji U(t)

$$A_1(x,t) = \beta = \rho - 1, \quad A_2(x,t) = \alpha = \lambda E[b^2] + O(\Delta t).$$

Dołączmy trzeci wyraz sumy w (4.16), biorąc

$$A_3(x,t) = \gamma(x,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (y-x)^3 f(x,t;y,t+\Delta t)dy,$$

to jest

$$\gamma \Delta t = E\left(\left\{\left[X_T(t+\Delta t) - X_T(t)\right] - \beta \Delta t\right\}^3 \middle| X(t)\right) = \lambda \Delta t E[b^3] + O(\Delta t).$$

Przyjmujemy $\gamma = \lambda E[b^3]$, uwzględniając w ten sposób trzeci moment rozkładu czasu obsługi i aproksymując U(t) procesem Markowa $X_T(t)$ opisanym równaniem różniczkowym trzeciego rzędu. W stanie ustalonym, przy analogicznych do poprzednich warunkach brzegowych, równania opisujące proces mają postać

$$0 = -\frac{\gamma}{6} \frac{d^3 f_T(x)}{dx^3} + \frac{\alpha}{2} \frac{d^2 f_T(x)}{dx^2} - \beta \frac{df_T(x)}{dx} + p_0 b(x)$$

$$\lim_{x \to 0} \left[-\frac{\gamma}{6} \frac{d^2 f_T(x)}{dx^2} + \frac{\alpha}{2} \frac{df_T(x)}{dx} - \beta f_T(x)\right] = \lambda p_0, \tag{4.18}$$

$$\lim_{x \to 0} f_T(x) = \lim_{x \to \infty} f_T(x) = 0.$$

Przyjmijmy jako rozkład B(x) ten sam, co poprzednio, określony w (4.14), rozkład Coxa drugiego rzędu i obliczmy pierwsze trzy momenty tego rozkładu

$$\begin{aligned} E[b] &= s_1 + as_2 \\ E[b^2] &= 2s_1^2 + 2as_1 s_2 + 2as_2^2 \\ E[b^3] &= 6s_1^3 + 6a(s_2^3 + s_1 s_2^2 + s_1^2 s_2), \end{aligned} \tag{4.19}$$

gdzie $s_1 = 1/\mu_1$, $s_2 = 1/\mu_2$

Rozwiązanie równań (4.18) przy tak określonym b(x) to

$$f_T(x) = -p_0\lambda(F_1+F_2)e^{l_1 x} + p_0\lambda F_1 e^{-\mu_1 x} + p_0\lambda F_2 e^{-\mu_2 x}, \tag{4.20}$$

skąd

$$F_T(x) = p_0\lambda \frac{F_1 + F_2}{l_1}(1 - e^{l_1 x}) + p_0 \frac{\lambda F_1}{\mu_1}(1 - e^{-\mu_1 x}) + p_0 \frac{F_2}{\mu_2}(1 - e^{-\mu_2 x}) + p_0. \tag{4.21}$$

gdzie

$$l_1 = \frac{-\frac{\alpha}{4} + \frac{1}{2}\sqrt{\frac{\alpha^2}{4} - \frac{2}{3}\lambda\beta}}{-\frac{\gamma}{6}}$$

jest jedynym ujemnym pierwiastkiem równania charakterystycznego

$$-\frac{\gamma}{6} l^2 + \frac{\alpha}{2} l - \beta = 0,$$

$$F_1 = \frac{-b_1/\mu_1}{\frac{\gamma}{6}\mu_1^2 + \frac{\alpha}{2}\mu_1 + \beta}, \qquad F_2 = \frac{-b_2/\mu_2}{\frac{\gamma}{6}\mu_2^2 + \frac{\alpha}{2}\mu_2 + \beta},$$

a z warunku normalizacyjnego otrzymujemy

$$p_0 = \left[1 + F_1/\mu_1 + F_2/\mu_2 + \frac{F_1 + F_2}{1_1}\right]^{-1}.$$

Przykłady numeryczne

Dla lepszego poznania dokładności aproksymacji dyfuzyjnej z procesem powrotów elementarnych oraz sprawdzenia celowości uwzględnienie w aproksymacji trzeciego momentu rozkładu czasu obsługi przeprowadzono szereg obliczeń. Przyjęto stanowisko M/G/1 z rozkładem B(x) określonym w (4.14), którego parametry a, μ_1, μ_2 są obliczane na podstawie założonych i uważanych za znane wartości średniej i wariancji:

$$\mu_1 = 2\Theta\mu, \quad \mu_2 = 2(1-\Theta)\mu, \quad a = 1 - \Theta - \frac{(1-\Theta)^2}{\Theta},$$

gdzie $\Theta = \frac{1}{2}\left(1 - \sqrt{\frac{c_b - 1}{c_b + 1}}\right)$.

Jest to rozkład zaproponowany w [Saue 75] i traktowany od tego czasu jako standardowy rozkład w obliczeniach testowych. Dla rozkładu (4.14) wzór Pollaczka-Chinczyna (4.4) przyjmuje postać

$$\hat{f}_u(r) = c_1 + \frac{c_2}{r - r_1} + \frac{c_3}{r - r_2},$$

gdzie

$$r_{1,2} = \frac{\lambda - \mu_1 - \mu_2 \pm \sqrt{\Delta}}{2},$$

$$\Delta = (\mu_1 - \mu_2 - \lambda)^2 + 4a\lambda\mu_1,$$

$$c_1 = 1 - \rho,$$

$$c_2 = \frac{1-\rho}{r_1 - r_2}\left[\mu_1\mu_2 + r_1(\mu_1 + \mu_2 + r_1)\right],$$

$$c_3 = \frac{1-\rho}{r_2 - r_1}\left[\mu_1\mu_2 + r_2(\mu_1 + \mu_2 + r_2)\right],$$

czyli po transformacji odwrotnej

$$f_u(x) = c_1\delta(x) + c_2 e^{r_1 x} + c_3 e^{r_2 x} \qquad (4.22)$$

oraz

$$F_u(x) = c_1 + \frac{c_2}{r_1}(e^{r_1 x} - 1) + \frac{c_3}{r_2} e^{r_2 x} - 1 \qquad (4.23)$$

Wyniki aproksymacji dyfuzyjnej F(x) (4.15), aproksymacji $F_T(x)$ (4.21) można więc porównać ze znanymi w tym przypadku wynikami dokładnymi (4.23).

Obliczenia przeprowadzono dla różnego obciążenia systemu i dla różnych wariancji rozkładu B(x), przyjmując zawsze $E[b] = 1$.

Rysunki 4.2a,b,c przedstawiają przebieg błędów względnych

$$\varepsilon(x) = \frac{F_u(x) - F(x)}{F_u(x)}, \qquad \varepsilon_T(x) = \frac{F_u(x) - F_T(x)}{F_u(x)}$$

dla wybranych wartości ρ i wartości $C_b = 1$, $C_b = 10$.

Widać, że wprowadzenie do modelu trzeciego momentu rozkładu B(x) zmniejsza na ogół błąd aproksymacji, lecz nie powoduje radykalnego zwiększenia dokładności.

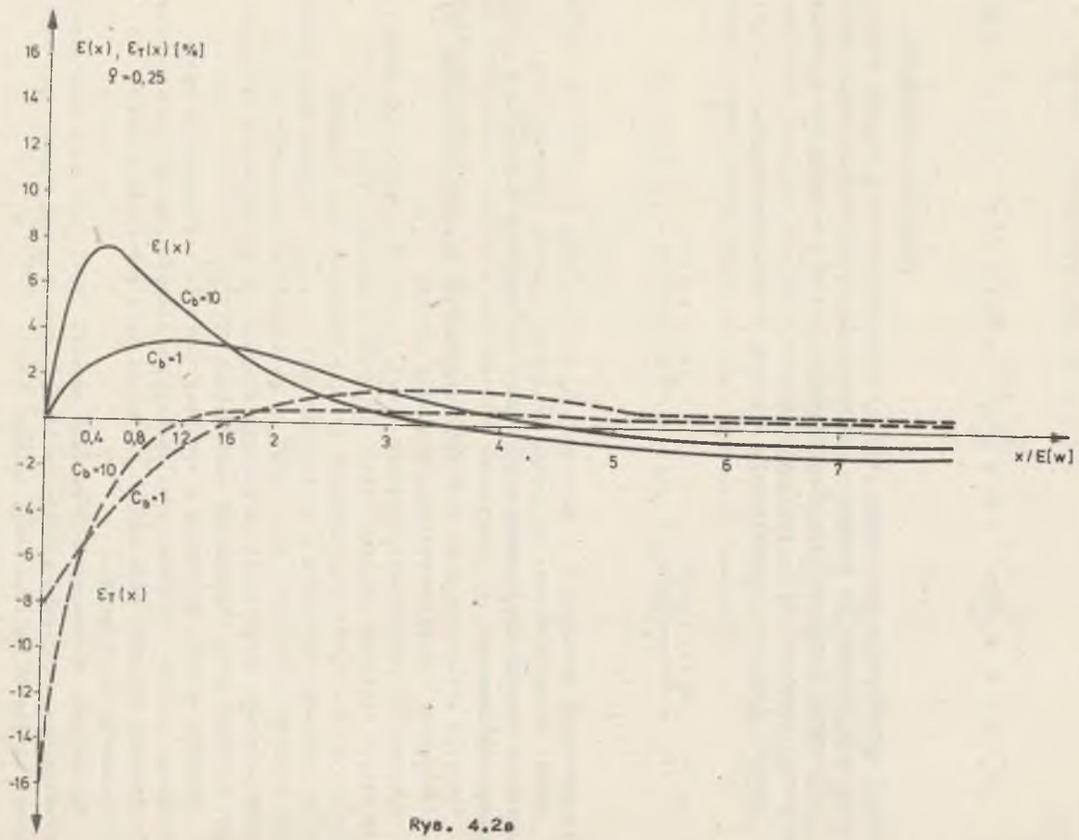
Powstaje stosunkowo duży błąd przy wyznaczaniu prawdopodobieństwa p_0 (aproksymacja dyfuzyjna wyznacza dokładnę wartość p_0).

Rysunek 4.3 przedstawia przebieg $\varepsilon(x)$, $\varepsilon_T(x)$ dla ustalonych dwóch pierwszych momentów rozkładu B(x) (wybrano przypadek $E[b] = 1$, $E[b^2] = 5$, $\rho = 0{,}50$) przy zmiennym trzecim momencie tego rozkładu. W tym przypadku obliczano z (4.19) wartości μ_1, μ_2 rozkładu Coxa (4.14) jako funkcja ustalonych $E[b]$, $E[b^2]$ i zmiennego jako parametr a. Lekka przewaga aproksymacji trzeciego stopnia nad aproksymacją dyfuzyjną jest widoczna przy niewielkich wartościach $E[b^3]$.

Uzyskane wyniki, porównane z wynikami [Bere 81], wskazują, że wpływ trzeciego momentu rozkładu czasu obsługi na długość kolejki jest większy w przypadku systemów zamkniętych, z ograniczoną populacją klientów niż w rozważanym tu przypadku.

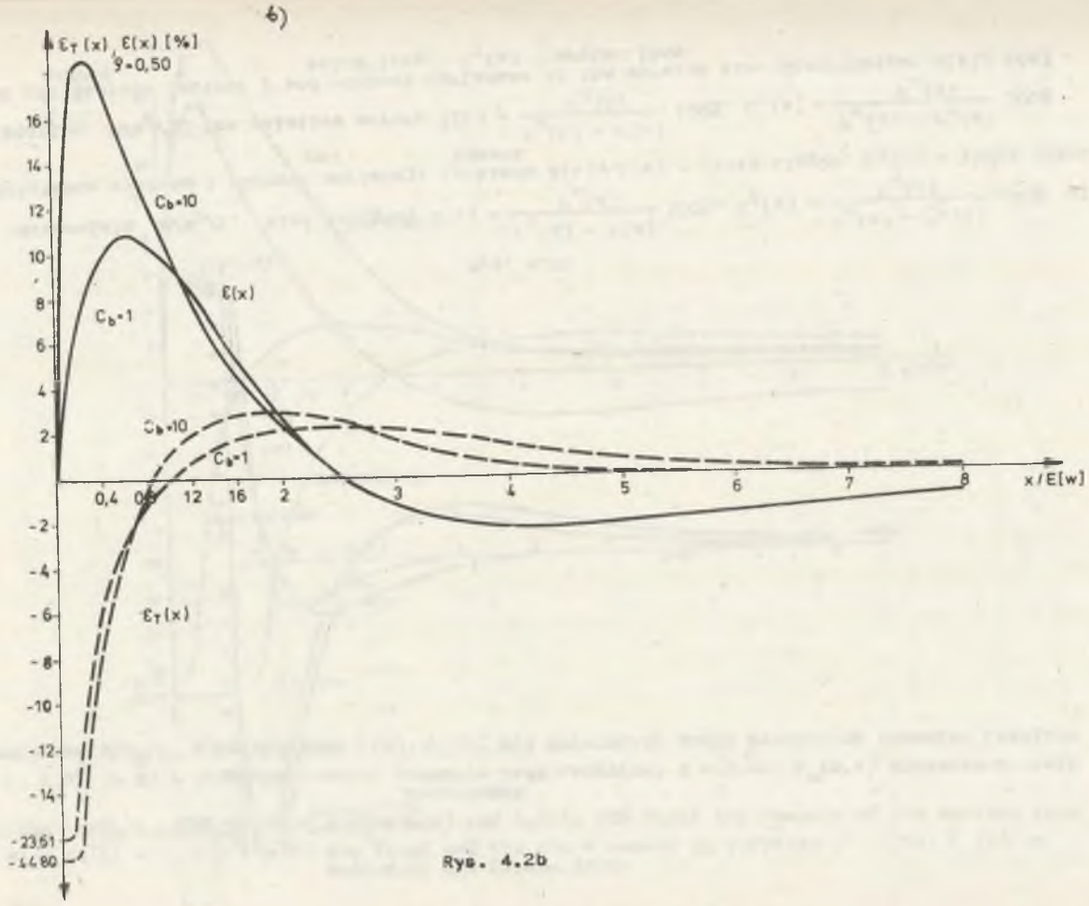
Reasumując, wprowadzenie trzeciego momentu B(x) poprawia nieznacznie aproksymację, lecz wprowadzenie następnych momentów nie wydaje się, zważywszy trudności obliczeniowe, celowe.

a)

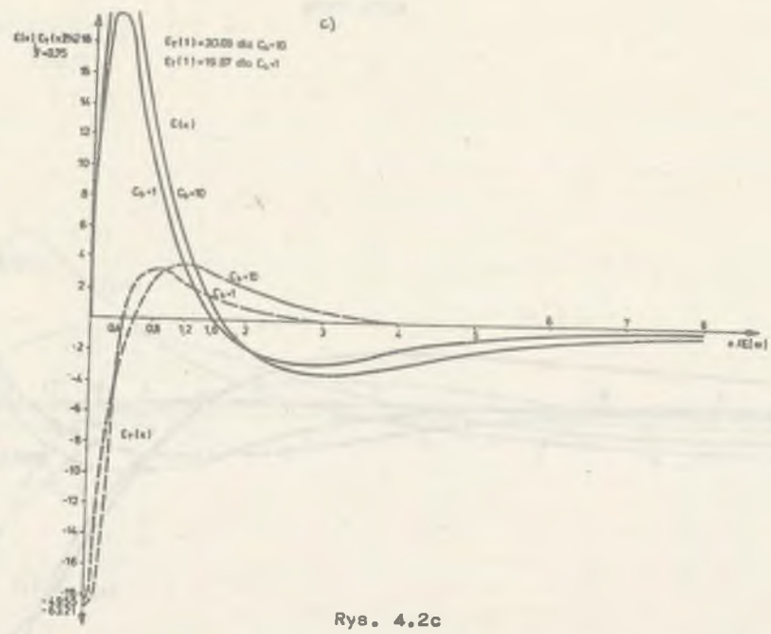


Rys. 4.2a

b)



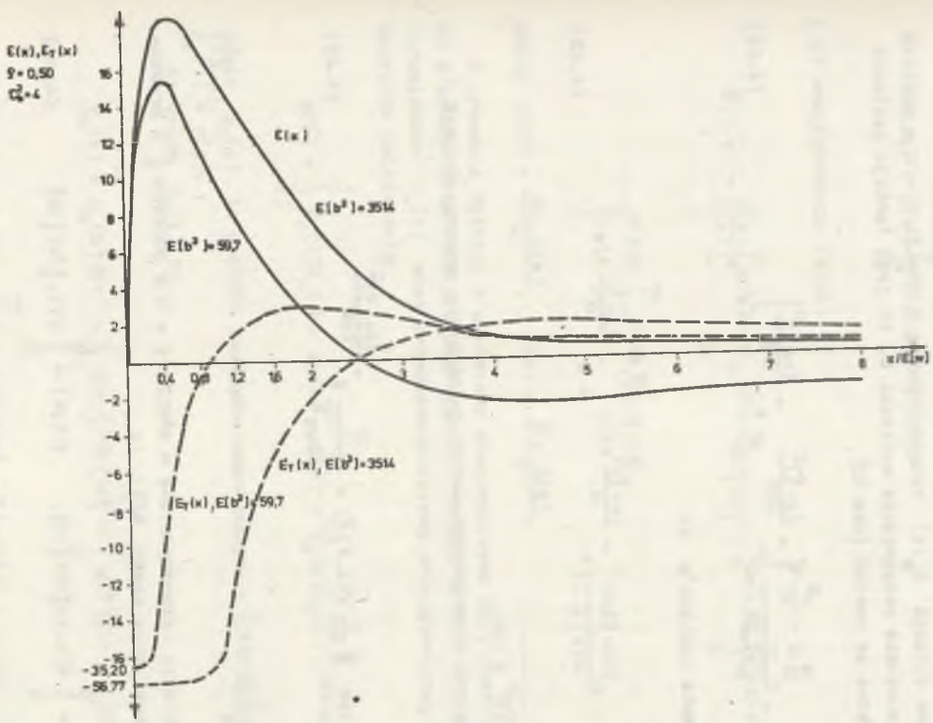
Rys. 4.2b



Rys. 4.2c

Rys. 4.2. Stanowisko $M/K_2/1$, błąd względny $\varepsilon(x) = \frac{F_u(x) - F(x)}{F_u(x)}$ 100%, $\varepsilon_T(x) = \frac{F_u(x) - F_T(x)}{F_u(x)}$ 100% dla różnego obciążenia systemu i różnych wariancji rozkładu B(x); $\varepsilon(x)$ - linia ciągła, $\varepsilon_T(x)$ - linia przerywana

Fig.4.2. Station $M/K_2/1$, the relative errors $\varepsilon(x) = \frac{F_u(x) - F(x)}{F_u(x)}$ 100%, $\varepsilon_T(x) = \frac{F_u(x) - F_T(x)}{F_u(x)}$ 100% for various utilization factors ρ and various variances of the service time distribution B(x); $\varepsilon(x)$ - solid line, $\varepsilon_T(x)$ - broken line



Rys. 4.3. Stanowisko $M/K_2/1$, błąd względny $\varepsilon(x)$, $\varepsilon_T(x)$ dla ustalonych dwóch pierwszych momentów rozkładu B(x) ($E[b] = 1$, $E[b^2] = 5$) i zmiennym trzecim momencie tego rozkładu; $\rho = 0,50$, $F_u(x,t)$ zaznaczono linią przerywaną

Fig. 4.3. Station $M/K_2/1$, the relative errors $\varepsilon(x)$ and $\varepsilon_T(x)$; the first two moments of the service time distribution B(x) ($E[b] = 1$, $E[b^2] = 5$) are fixed and the third moment is varying; $\rho = 0,50$, $F_u(x)$ is marked by the broken line

4.5. Aproksymacja stanów nieustalonych i dynamiki zmian średniego czasu czekania

Dla rozwiązania układu równań (4.9) dla dyfuzji z powrotami elementarnymi powrócimy do techniki omówionej i wykorzystanej w rozdziale trzecim, polegającą na przedstawieniu rozwiązania (4.9) w funkcji rozwiązania równania dyfuzji z założonym w $x = 0$ ekranie pochłaniającym.

Jeżeli proces dyfuzji $X_p(t)$ rozpoczyna się w chwili $t = 0$ w punkcie ξ, a kończy w momencie osiągnięcia wartości 0, to jego funkcja gęstości prawdopodobieństwa ma postać [Cox 65]

$$\varphi(x,t;\xi) = \frac{e^{\frac{\beta}{\alpha}x - \frac{\beta^2}{\alpha}t}}{\sqrt{2\pi\alpha t}}\left[e^{-\frac{(x-\xi)^2}{2\alpha t}} - e^{-\frac{(x+\xi)^2}{2\alpha t}}\right], \tag{4.24}$$

a jej transformata Laplace'a to

$$\bar{\varphi}(x,s;\xi) = \frac{e^{\beta(x-\xi)/\alpha}}{A(s)}\left[e^{-\frac{|x-\xi|}{\alpha}A(s)} - e^{-\frac{|x+\xi|}{\alpha}A(s)}\right], \tag{4.25}$$

gdzie $A(s) = \sqrt{\beta^2 + 2\alpha s}$.

Obliczamy gęstość prawdopodobieństwa zakończenia procesu dyfuzji w chwili t:

$$\gamma(t,\xi) = \lim_{x\to 0}\frac{\alpha}{2}\frac{\partial}{\partial x}\varphi(x,t;\xi) = \frac{\xi}{\sqrt{2\pi\alpha t^3}}e^{-\frac{(\beta t+\xi)^2}{2\alpha t}}, \tag{4.26}$$

$$\bar{\gamma}(s,\xi) = e^{\frac{\xi}{\alpha}[\beta+A(s)]}. \tag{4.27}$$

Jeżeli proces dyfuzji rozpoczyna się w chwili $t = 0$ w punkcie ξ z prawdopodobieństwem określonym przez $b(\xi)$, to

$$f(x,t;b) = \int_0^\infty \varphi(x,t;\xi)b(\xi)d\xi, \qquad \gamma(t;b) = \int_0^\infty \gamma(t,\xi)b(\xi)d\xi \tag{4.28}$$

oraz

$$\bar{\gamma}(s;b) = \int_0^\infty \bar{\gamma}(s,\xi)b(\xi)d\xi = \int_0^\infty e^{\frac{\xi}{\alpha}[\beta+A(s)]}b(\xi)d\xi = \bar{b}\left\{\frac{1}{\alpha}[\beta+A(s)]\right\} \tag{4.29}$$

Załóżmy, że chwilę $\tau \in (0,t)$ rozpoczęcia się procesu określa rozkład o gęstości $g(\tau)$, wówczas gęstość $\varkappa(x,t)$ tego procesu ma postać

$$\varkappa(x,t) = \int_0^t g(\tau)\int_0^\infty \varphi(x,t-\tau;\xi)b(\xi)d\xi d\tau; \tag{4.30}$$

$$\bar{\varkappa}(x,s) = \bar{g}(s)\int_0^\infty \bar{\varphi}(x,s;\xi)b(\xi)d\xi, \tag{4.31}$$

a po uwzględnieniu (4.25)

$$\bar{\varkappa}(x,s) = \frac{\bar{g}(s)}{A(s)}\left[e^{D(s)x}\int_0^x e^{-D(s)\xi}b(\xi)d\xi + e^{D(s)x}\int_x^\infty e^{-E(s)}b(\xi)d\xi\right.$$

$$\left. - e^{D(s)x}\int_0^\infty e^{-E(s)\xi}b(\xi)d\xi\right]. \tag{4.32}$$

gdzie $D(s) = \frac{\beta - A(s)}{\alpha}$, $E(s) = \frac{\beta + A(s)}{\alpha}$.

W procesie dyfuzji z powrotami elementarnymi $g(\tau)$ jest strumieniem masy prawdopodobieństwa przepływającym z ekranu na oś $x > 0$ i związanym ze strumieniem $\gamma_o(t)$ masy prawdopodobieństwa przepływającej w odwrotnym kierunku zależnością

$$g(\tau) = \int_0^\tau \gamma_o(t)l_o(\tau - t)dt, \qquad \bar{g}(s) = \bar{\gamma}_o(s)\bar{l}_o(s), \tag{4.33}$$

gdzie $l_o(t)$ to gęstość rozkładu czasu pobytu w ekranie; zakładamy $l_o(t) = a(t)$.

Strumień $\gamma_o(t)$ to

$$\gamma_o(t) = p_o(0)\delta(t) + \int_0^\infty \gamma(t,\xi)\Phi(\xi)d\xi + \int_0^t g(\tau)\gamma(t-\tau)d\tau,$$

$$\bar{\gamma}_o(s) = p_o(0) + \int_0^\infty \bar{\gamma}(s,\xi)\Phi(\xi)d\xi + \bar{\gamma}(s)\bar{g}(s) = p_o(0) + \bar{\Phi}\left\{\frac{1}{\alpha}[\beta + A(s)]\right\} + \tag{4.34}$$

$$+ \bar{g}(s)\bar{\gamma}(s).$$

gdzie $p_0(0)$ i $\Psi(\xi)$ to odpowiednio masa prawdopodobieństwa w ekranie i jej rozkład na półosi $x > 0$ w chwili początkowej, a gęstość prawdopodobieństwa procesu ma postać (4.30), (4.31) uzupełnioną o człon opisujący rozładowanie warunków początkowych:

$$f(x,t) = \int_0^t g(\tau) \int_0^\infty \varphi(x,t-\tau;\xi)b(\xi)d\xi d\tau + \int_0^\infty \varphi(x,t;\xi)\Psi(\xi)d\xi,$$

$$\bar{f}(x,s) = \bar{g}(s) \int_0^\infty \bar{\varphi}(x,s;\xi)b(\xi)d\xi + \int_0^\infty \bar{\varphi}(x,s;\xi)\Psi(\xi)d\xi, \quad x > 0 \qquad (4.35)$$

Prawdopodobieństwo $p_0(t)$ pobytu procesu w punkcie $x = 0$ obliczamy z bilansu

$$p_0(t) = \int_0^t \left[\gamma_0(t) - g(\tau)\right] d\tau,$$

$$\bar{p}_0(s) = \frac{1}{s}\left[\bar{\gamma}_0(s) - \bar{g}(s)\right] = \frac{\bar{g}(s)}{s} \frac{1 - \bar{l}_0(s)}{\bar{l}_0(s)}, \qquad (4.36)$$

a funkcje $\gamma_0(t)$, $g(t)$ czy też w praktyce $\bar{\gamma}_0(s)$, $\bar{g}(s)$, wyznaczamy z (4.33), (4.34).

Niech dla $t < t_i$ zachodzi $A(x) = A_{(i)}(x)$, $B(x) = B_{(i)}(x)$ i proces dyfuzji jest określony gęstością $f_{(i)}(x,t)$, a w momencie $t = t_i$ następuje zmiana warunków procesu: $A(x) = A_{(i+1)}(x)$, $B(x) = B_{(i+1)}(x)$, dla $t \geq t_i$.

Nową wartość przyjmuję współczynniki β, α : $\beta = \beta_{(i+1)}$, $\alpha = \alpha_{(i+1)}$. Mamy wówczas warunek początkowy w postaci $p_{0(i)}(t_i)$ i $f_{(i)}(x,t_i)$, a z (4.33)-(4.35) obliczamy

$$\bar{f}_{(i+1)}(x,s) = \bar{g}_{(i+1)}(x) \int_0^\infty \varphi_{(i+1)}(x,s;\xi)b_{(i+1)}(\xi)d\xi +$$

$$+ \int_0^\infty \bar{\varphi}_{(i+1)}(x,s;\xi)f_{(i)}(\xi,t_i)d\xi, \qquad (4.37)$$

$$\bar{g}_{(i+1)}(s) = \frac{p_{0(i)}(t_i) + \int_0^\infty \bar{\gamma}_{(i+1)}(s,\xi)f_{(i)}(\xi,t_i)d\xi}{1 - \bar{l}_{(i+1)}(s)\bar{\gamma}_{(i+1)}(s;b)} =$$

$$= \frac{p_{0(i)}(t_i) + \hat{f}_{(i)}\left\{\frac{1}{\alpha_{(i+1)}}\left[\beta_{(i+1)} + A_{(i+1)}(s)\right], t_i\right\}}{1 - l_{(i+1)}(s)\bar{\gamma}_{(i+1)}(s;b)} \qquad (4.38)$$

gdzie $\bar{\varphi}_{(i+1)}(x,s;\xi)$, $\bar{\gamma}(s,\xi)$, $\bar{\gamma}(s;b)$ są określone odpowiednio przez (4.25), (4.27), (4.29).

Przypadki szczególne

a) Wykładniczy rozkład czasów obsługi, $B(x) = 1 - e^{-\mu x}$. Z (4.32) otrzymujemy

$$\bar{\varkappa}(x,s) = \bar{g}(s)H(s)\left[e^{D(s)x} - e^{-\mu x}\right] \qquad (4.39)$$

gdzie $H(s) = \frac{1}{A(s)}\left[\frac{\mu}{D(s) + \mu} - \frac{\mu}{E(s) + \mu}\right]$

oraz

$$\bar{\mathcal{H}}(x,s) = \int_0^x \varkappa(x,s)dx = g(s)H(s)\left\{-\frac{1}{D(s)} \cdot \left[1 - e^{-D(s)x}\right] - \right.$$

$$\left. - \frac{1}{\mu}\left(1 - e^{-\mu x}\right)\right\}. \qquad (4.40)$$

Powtórzmy uwagę poczynioną przy rozwiązaniu stanu ustalonego: w przypadku prawie ogólnego rozkładu czasów obsługi sprowadzamy ten rozkład do sumy rozkładów wykładniczych, a $\bar{\mathcal{H}}(x,s)$ jest w tym przypadku ważoną sumą rozwiązań (4.39).

b) W chwili początkowej proces ma wartość $X(0) = 0$, tj. $p_0(0) = 1$; $\Psi(x) = 0$ dla $x > 0$, co odpowiada aproksymacji zachowania się stanowiska pustego w chwili początkowej:

$$\begin{aligned} f(x,t) &= \varkappa(x,t) + p_0(t)\delta(x), \\ F(x,t) &= \mathcal{H}(x,t) + p_0(t), \\ \bar{f}(x,s) &= \bar{\varkappa}(x,s) + \bar{p}_0(s)\delta(x), \\ \bar{F}(x,s) &= \bar{\mathcal{H}}(x,s) + \bar{p}_0(s), \qquad x \geq 0 \end{aligned} \qquad (4.41)$$

$$\bar{g}(s) = \frac{\bar{I}_0(s)}{1 - \bar{I}_0(s)\bar{\vartheta}(s)}, \tag{4.42}$$

$$\bar{p}_0(s) = \frac{1}{s}\left[1 + \frac{\bar{I}_0(s)\bar{\vartheta}(s) - \bar{I}_0(s)}{1 - \bar{I}_0(s)\bar{\vartheta}(s)}\right] = \frac{1}{s}\,\frac{1 - \bar{I}_0(s)}{1 - \bar{I}_0(s)\bar{\vartheta}(s)}. \tag{4.45}$$

Jeżeli dodatkowo rozkład B(x) jest wykładniczy, to zmienna w czasie wartość średnia procesu jest wyrażona transformatę

$$E[X(s)] = \int_0^\infty x\bar{f}(x,s)dx = \bar{g}(s)H(s)\left[\frac{1}{D^2(s)} - \frac{1}{\mu^2}\right]. \tag{4.44}$$

(4.44) po obliczeniu transformaty odwrotnej aproksymuje średni czas czekania klientów przed pustym w chwili t = 0 stanowiskiem M/M/1.

c) W chwili $t < 0$ panuje stan ustalony, funkcja gęstości procesu ma postać (4.12). Współczynniki równania dyfuzji mają wartość $\beta_{(1)}$, $\alpha_{(1)}$; $z_{(1)} = 2\beta_{(1)}/\alpha_{(1)}$. W chwili t = 0 współczynniki równania dyfuzji zmieniają wartość na $\beta_{(2)}$, $\alpha_{(2)}$.
Mamy więc:

$$\psi(x) = f_{(1)}(x) = \frac{2\lambda_{(1)}[1 - \rho_{(1)}]}{\alpha_{(1)}[\mu_{(1)} + z_{(1)}]}\left[e^{z_{(1)}x} - e^{-\mu_{(1)}x}\right] + \left[1 - \rho_{(1)}\right]\delta(x)$$

i jeżeli $B_{(2)}(x) = 1 - e^{-\mu_{(2)}x}$, to

$$\bar{f}(x,s) = \bar{g}(s)\int_0^\infty \bar{\varphi}_{(2)}(x,s;\xi)\bar{b}_{(2)}(\xi)d\xi + \int_0^\infty \bar{\varphi}_{(2)}(x,s;\xi)f_{(1)}(\xi)d\xi +$$

$$+ \bar{p}_0(s)\delta(x) =$$

$$= \bar{g}(s)H(s)\left[e^{D_{(2)}(s)x} - e^{-\mu_{(2)}x}\right] + H_1(s)\left[H_2(s)e^{z_{(1)}x} + \right.$$

$$\left. + H_3(s)e^{D_{(2)}(s)x} + H_4(s)e^{-\mu_{(1)}x}\right] + \bar{p}_0(s)\delta(x), \tag{4.45}$$

gdzie:

$$\bar{g}(s) = \frac{p_0(0) + \int_0^\infty \bar{\vartheta}_{(2)}(s,\xi)f_{(1)}(\xi)d\xi}{1 - \bar{I}_{(2)}(s)\bar{\vartheta}_{(2)}(s)} = \frac{1 - \rho_{(1)} + \hat{f}_{(1)}[E_{(2)}(s)]}{1 - \bar{I}_2(s)\bar{\vartheta}_{(2)}(s)},$$

$$\hat{f}_{(1)}(r) = \frac{2\lambda_{(1)}[1 - \rho_{(1)}]}{\alpha_{(1)}[1 + z_{(1)}]}\left[1 + \frac{1}{z_{(1)} + r} - \frac{1}{\mu_{(1)} + r}\right],$$

$$H(s) = \frac{1}{A_{(2)}(s)}\left[\frac{\mu_{(2)}}{D_{(2)}(s) + \mu_{(2)}} - \frac{\mu_{(2)}}{E_{(2)}(s) + \mu_{(2)}}\right],$$

$$H_1(s) = \frac{1}{A_{(2)}(s)}\,\frac{2\lambda_{(1)}[1 - \rho_{(1)}]}{\alpha_{(1)}[1 + z_{(1)}]},$$

$$H_2(s) = \frac{1}{-D_{(2)}(s) + z_{(1)}} + \frac{1}{E_{(2)}(s) - z_{(1)}},$$

$$H_3(s) = \frac{1}{D_{(2)}(s) - z_{(1)}} - \frac{1}{D_2(s) + \mu_{(1)}} - \frac{1}{E_{(2)}(s) - z_{(1)}} + \frac{1}{E_{(2)}(s) + \mu_{(1)}},$$

$$H_4(s) = \frac{1}{\mu_{(1)} + D_{(2)}(s)},$$

$$\bar{p}_0(s) = \frac{g(s)}{s}\,\frac{1 - \bar{I}_{(2)}(s)}{\bar{I}_{(2)}(s)}.$$

Zmienna w czasie średnia wartość czasu czekania ma transforamtę $E[\bar{X}(s)]$:

$$E[\bar{X}(s)] = \bar{g}(s)H(s)\left[\frac{1}{D^2_{(2)}(s)} - \frac{1}{\mu^2_{(2)}}\right] +$$

$$+ H_1(s)\left[\frac{H_2(s)}{z^2_{(1)}} + \frac{H_3(s)}{D^2_{(2)}(s)} + \frac{H_4(s)}{\mu^2}\right]. \tag{4.46}$$

Przykłady numeryczne

Przykłady numeryczne odnoszę się do stanowiska M/M/1 - jedynego, w którym czasowa postać $f_u(x,t)$, $F_u(x,t)$ jest znana w jawnej postaci.

Dla stanowiska M/M/1 rozkład prawdopodobieństwa $p(k,t;i)$ liczby klientów obecnych w stanowisku

$$p(k,t;i) = P[N(t) = k \mid N(0) = i]$$

jest znany i ma podaną już postać (3.2).

Korzystając z dość oczywistej dla stanowiska M/M/1 zależności pomiędzy N(t) i U(t), por. [Taka 62b], możemy napisać

$$f_u(x,t;0) = p(0,t;0)\delta(x) + \sum_{k=1}^{\infty} p(k,t;0)b^{*k}(x)$$

$$F_u(x,t;0) = p(0,t;i) + \sum_{k=1}^{\infty}\int_0^x p(k,t;0)b^{*k}(x)dx,$$

gdzie $b^{*k}(x)$ jest k-krotnym splotem rozkładu wykładniczego $b(x)$ z samym sobą, a więc rozkładem Erlanga, czyli

$$f_u(x,t;0) = p(0,t;0)\delta(x) + \sum_{k=1}^{\infty} p(k,t;0)\frac{(\mu x)^{k-1}}{(k-1)!}\mu \qquad (4.47)$$

$$F_u(x,t;0) = p(0,t;0) + \sum_{k=1}^{\infty} p(k,t;0)\, e^{-\mu\xi}\frac{(\mu\xi)^{k-1}}{(k-1)!}\mu d\xi. \qquad (4.48)$$

Oznaczmy

$$C_k(x) = e^{-\mu\xi}(\mu\xi)^k d\xi, \quad C_k = \int_0^{\infty} e^{-\mu\xi}(\mu\xi)^k d\xi = C_k(x)\Big|_0^{\infty}$$

i przepiszmy (4.48) jako

$$F_u(x,t;0) = p(0,t;0) + \sum_{k=1}^{\infty} p(k,t;0)\frac{\mu^k}{(k-1)!}\left[C_{k-1}(x) + C_{k-1}\right]. \qquad (4.49)$$

Obliczmy zmienną w czasie średnią wartość czasu czekania:

$$E[U(t)] = \int_0^{\infty} x f_u(x,t;0)dx = \sum_{k=1}^{\infty} p(k,t;0)\frac{\mu^k}{(k-1)!}C_k\,; \qquad (4.50)$$

$C_k(x)$ i C_k można obliczyć za pomocą następujących wzorów rekurencyjnych [Bron 70]

$$C_k(x) = \frac{1}{\mu}x^k e^{-\mu x} + \frac{k}{\mu}C_{k-1}(x), \quad C_o(x) = \frac{1}{\mu}e^{-\mu x}$$

$$C_k = \frac{k}{\mu}C_{k-1}, \quad C_o = \frac{1}{\mu}.$$

W przypadku stanowiska M/M/1 można więc stosunkowo łatwo porównać wyniki aproksymacji Gavera i aproksymacji włączającej powroty elementarne z wynikami dokładnymi (4.49), (4.50).

W przypadku aproksymacji Gavera, zgodnie z (4.7), przy zerowych warunkach początkowych

$$\bar{f}_G(x,s;0) = -\frac{D(s)}{s}e^{D(s)x}$$

oraz

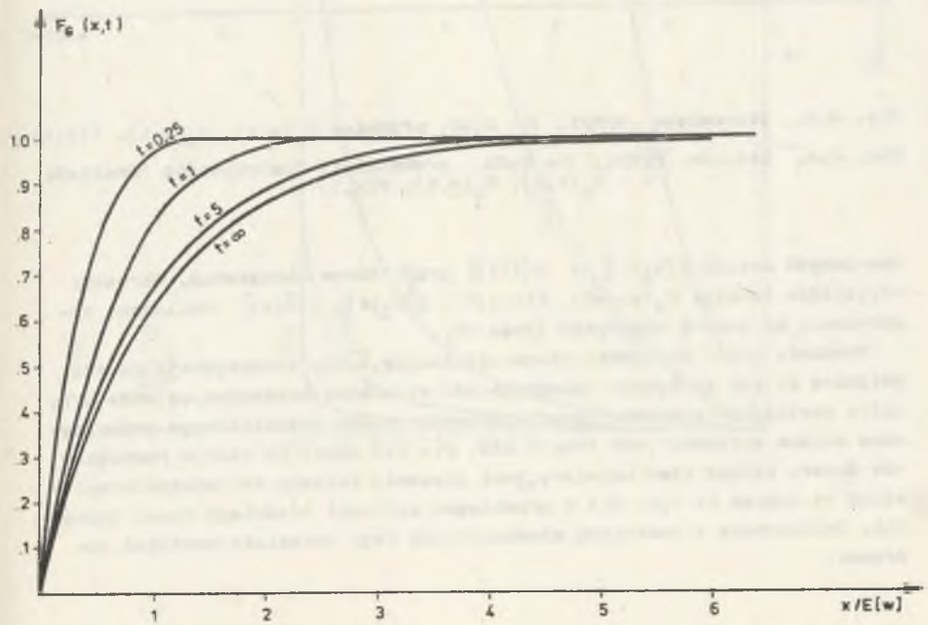
$$\bar{F}_G(x,s;0) = \frac{1}{s}\left[1 - e^{D(s)x}\right],$$

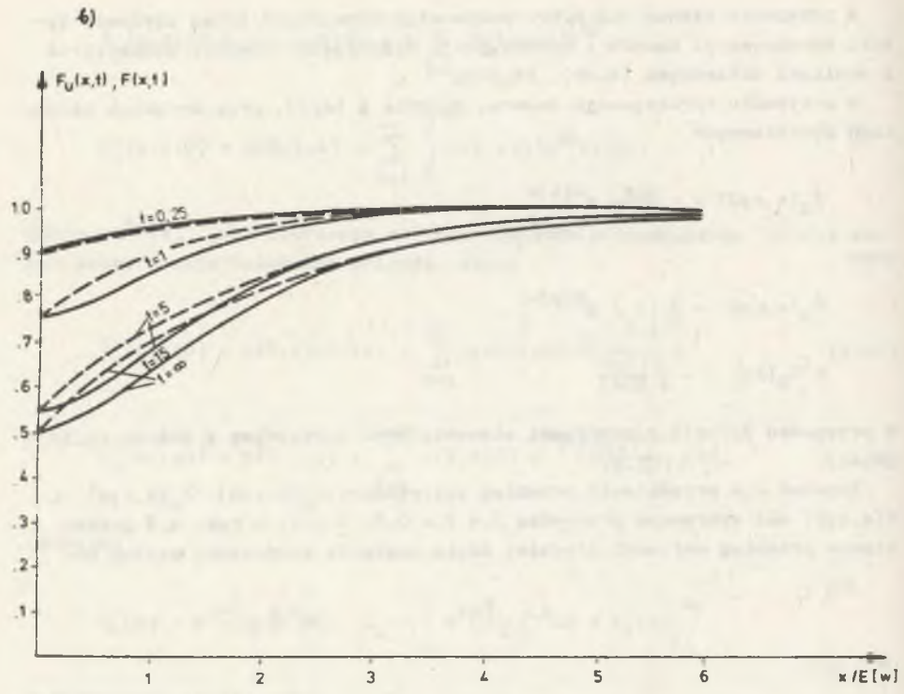
$$E\left[\bar{X}_G(s)\right] = -\frac{1}{s}\frac{1}{D(s)}$$

W przypadku dyfuzji z powrotami elementarnymi korzystamy z wzorów (4.39)-(4.41).

Rysunek 4.4 przedstawia przebieg dystrybuant $F_G(x,t;0)$, $F_u(x,t;0)$ i $F(x,t;0)$ dla wybranego przypadku $\lambda = \rho = 0{,}5$, $\mu = 1$, a rys. 4.5 przedstawia przebieg wartości średniej czasu czekania szacowanej według obu

a)



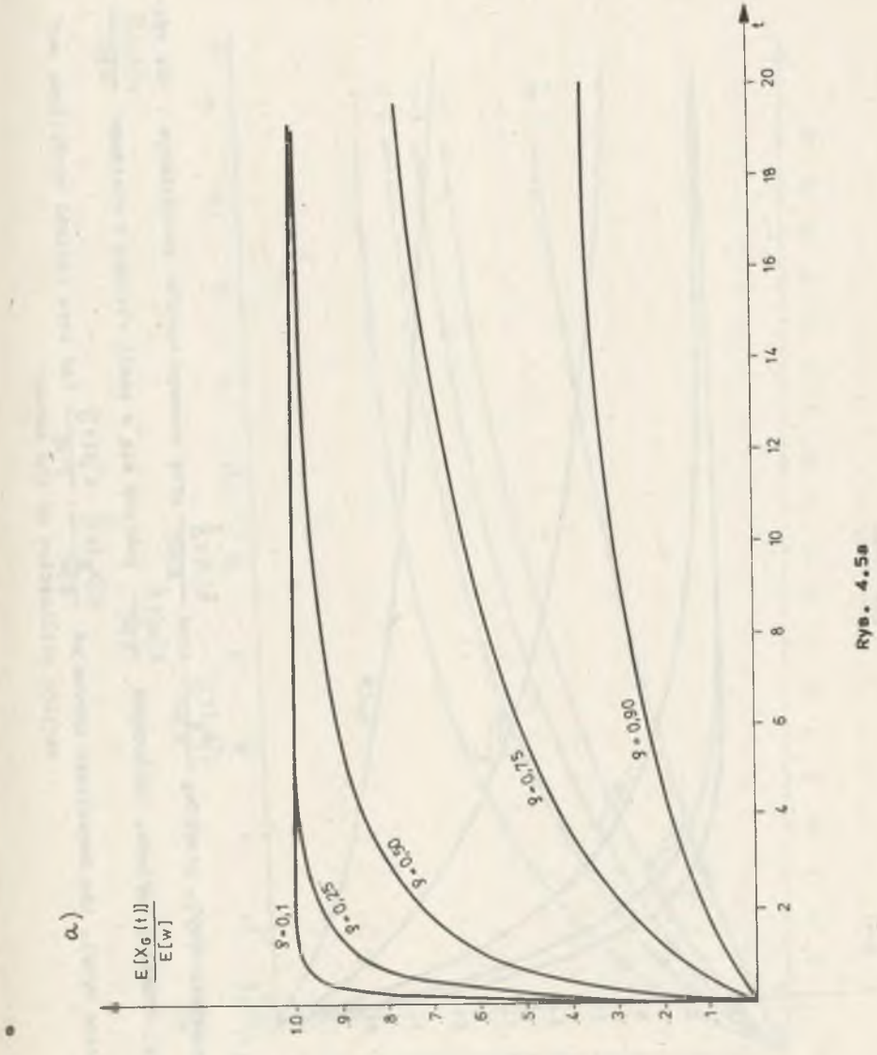


Rys. 4.4. Stanowisko M/M/1, $\rho = 0{,}50$, przebieg $F_u(x,t)$, $F_G(x,t)$, $F(x,t)$

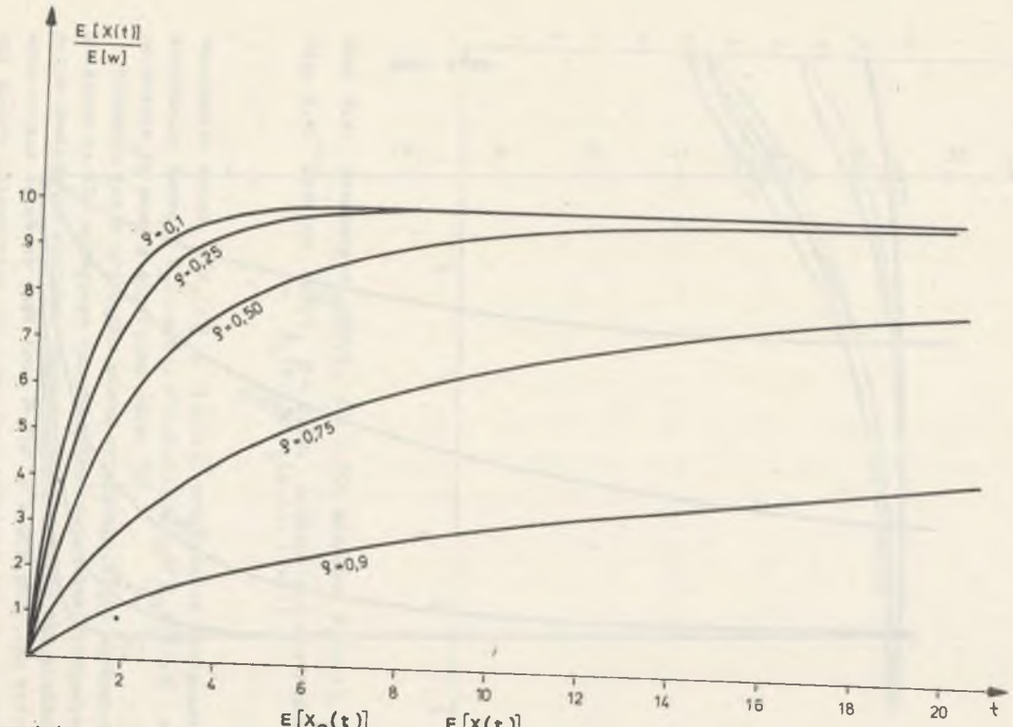
Fig. 4.4. Station M/M/1, $\rho = 0{,}50$, probability distribution functions $F_u(x,t)$, $F_G(x,t)$, $F(x,t)$

omawianych metod, $E[X_G(t)]$ i $E[X(t)]$ przy różnym obciążeniu. Wartości oryginałów funkcji $F_G(x,s;0)$, $\bar{F}(x,s;0)$, $E[\bar{X}_G(s)]$, $E[\bar{X}(s)]$ obliczono numerycznie za pomocą algorytmu [Steh 79].

Podobnie jak w przypadku stanu ustalonego, błędy aproksymacji Gavera, związane ze źle oddającym rzeczywistość warunkiem brzegowym, są duże przy małym obciążeniu systemu. Ponadto dynamika stanu nieustalonego przewidywana obiema metodami jest inna - dla $\rho = 0{,}5$ czas, po którym rozwiązanie Gavera osiąga stan ustalony, jest znacznie krótszy od rzeczywistego - widać to dobrze na rys. 4.5 z przebiegów wartości średniego czasu czekania. Rozwiązanie z powrotami elementarnymi daje natomiast wartości poprawne.



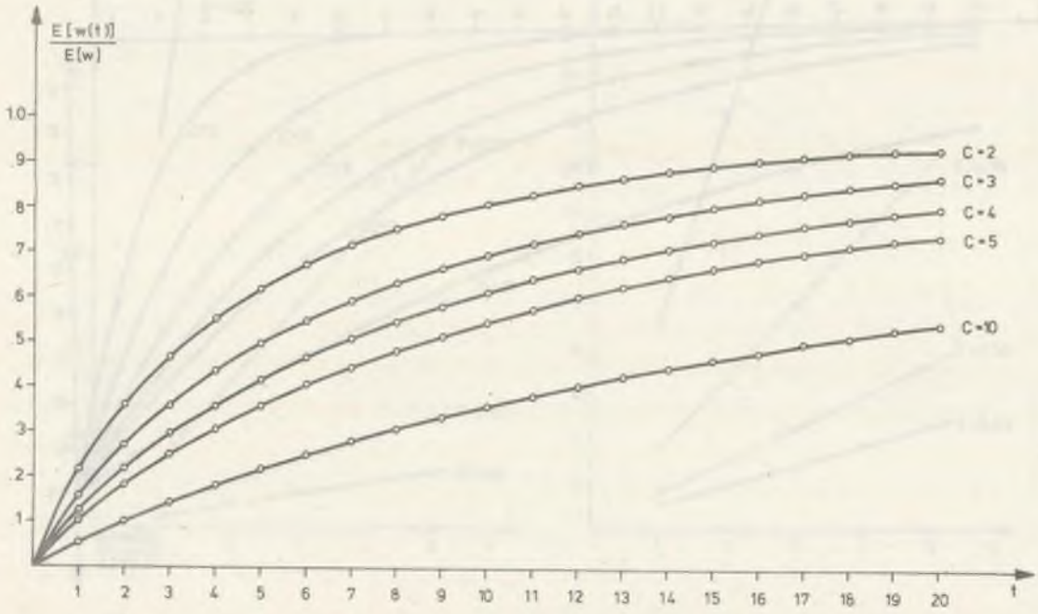
Rys. 4.5a



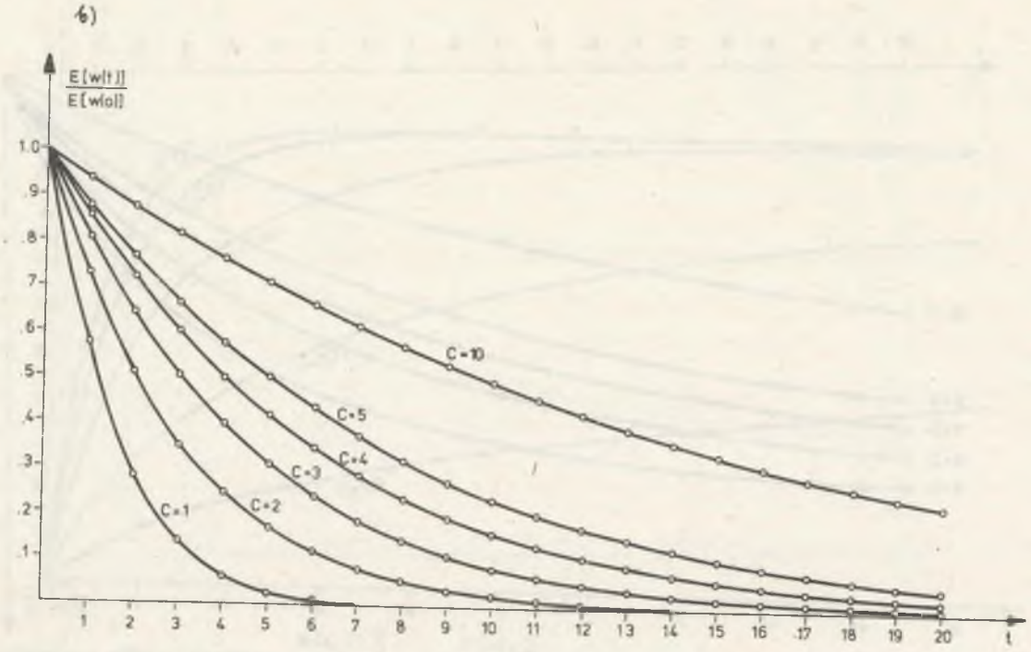
Rys. 4.5. Stanowisko M/M/1, przebieg $\frac{E[X_C(t)]}{E[w]}$ oraz $\frac{E[X(t)]}{E[w]}$ przy zerowym warunku początkowym i dla różnego obciążenia. Przebieg wartości dokładnych $\frac{E[U(t)]}{E[w]}$ pokrywa się w skali rysunku z wykresem $\frac{E[X(t)]}{E[w]}$

Fig. 4.5. Station M/M/1, the normalized curves of $\frac{E[X_C(t)]}{E[w]}$, $\frac{E[X(t)]}{E[w]}$ for zero initial conditions and varying utilization of the server

a)

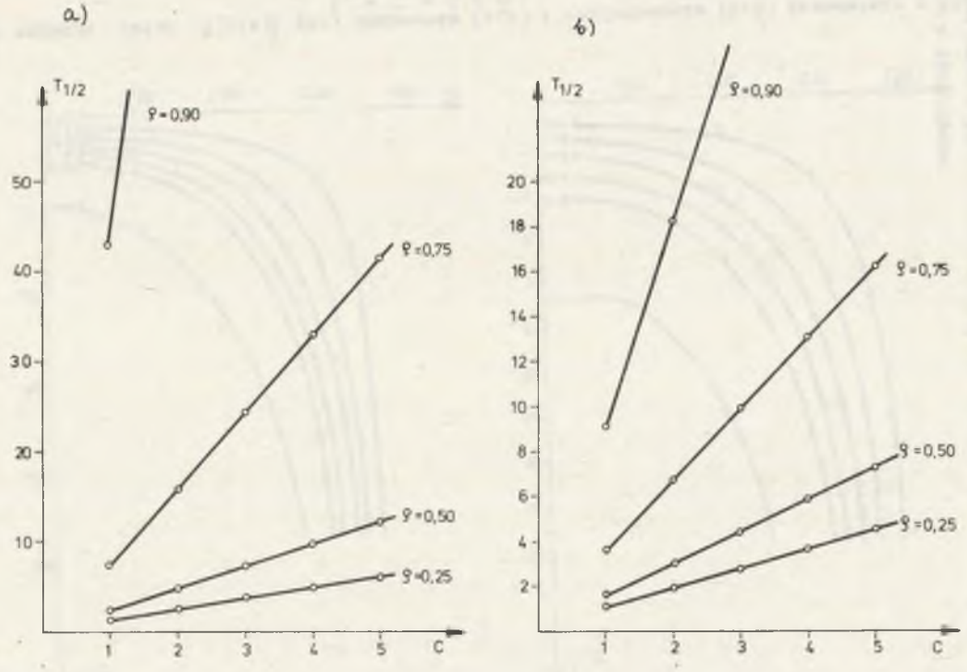


Rys. 4.6a



Rys. 4.6. Przybliżony średni wirtualny czas czekania przy ładowaniu (a) i rozładowaniu (b) stanowiska, $\rho = 0,5$, $C_a = C_b = C$

Fig. 4.6. Approximate mean virtual waiting time during charging (a) and discharging (b) of a station; $\rho = 0,5$, $C_a = C_b = C$



Rys. 4.7a,b

Rys. 4.7. Stała czasowa zmian $E[U(t)]$ przy ładowaniu (a,c) i rozładowaniu (b,d) stanowiska w funkcji $C_a = C_b = C$ i ρ

Fig. 4.7. Time constant of the changes of $E[U(t)]$ during charging (a,c) and discharging (b,d) of the station as a function of $C_a = C_b = C$ and ρ

Wykresy na rys. 4.6 przedstawiają zależność wartości średniego czasu czekania przy ładowaniu pustego stanowiska i przy rozładowaniu stanowiska będącego początkowo w stanie ustalonym, od wariancji rozkładów A(x) i B(x).

Rysunek 4.7 przedstawia stałą czasową tych przebiegów, wyznaczoną w ten sam sposób jak w rozdziale poprzednim dla zmian średniej liczby klientów w stanowisku.

Rozdział V

STANOWISKO OBSŁUGI ZE SKOŃCZONĄ POPULACJĄ KLIENTÓW - APROKSYMACJA DYFUZYJNA MODELU KONSERWATORA

Problem konserwatora jest klasycznym zadaniem teorii masowej obsługi. Znany od lat kilkudziesięciu [Chin 33, Palm 47], był też pierwszym modelem probabilistycznym zastosowanym do analizy pracy systemu komputerowego [Sher 61], a sukces tego zastosowania ("Jest rzeczą zadziwiającą, jak wiele cech, wydawałoby się istotnych dla pracy rzeczywistego, działającego w podziale czasu systemu komputerowego, może być pominiętych w tym modelu, który mimo to poprawnie przewiduje średni czas odpowiedzi systemu"- pisał A. Sherr [Sher 67]) przyczynił się do rozwoju bardziej złożonych modeli systemów komputerowych.

Literatura poświęcona zadaniu konserwatora jest obszerna, jej przegląd i porównanie różnych wariantów modelu zawiera m.in. [Izyd 81]. Model podstawowy zakłada skończoną populację N jednakowych klientów i pojedyncze stanowisko, obsługujące klientów zgodnie z kolejnością zgłoszeń. Rozkład czasu obsługi będziemy oznaczać jak dotychczas przez $B(x)$, natomiast $A(x)$ będzie oznaczać rozkład czasu pobytu klienta w źródle, tj. czasu pomiędzy zakończeniem obsługi klienta a momentem, gdy ustawi się on znów w kolejce. Dokładne rozwiązanie analityczne problemu, podające rozkład $p(n)$ liczby klientów w stanowisku obsługi, istnieje w przypadku, gdy jeden z wymienionych rozkładów $A(x)$, $B(x)$ jest dowolny, a drugi wykładniczy. Zastąpienie rozkładu wykładniczego rozkładem Erlanga pociąga już za sobą konieczność obliczeń numerycznych [Kopo 73].

Poniżej przedstawimy aproksymację dyfuzyjną problemu konserwatora (najpierw dla jednej, następnie dla wielu klas klientów), w przypadku gdy oba rozkłady $A(t)$, $B(t)$ są dowolne, a więc przybliżony model wieloklasowego stanowiska $G/G/1//\underline{N}$ w stanie ustalonym i nieustalonym. W analizie stanu ustalonego rozwiązanie uwzględnia również przypadek wielokanałowego stanowiska obsługi - model $G/G/k//\underline{N}$ - oraz model, w którym tylko m spośród $N-n$ obecnych w źródle klientów może zażądać obsługi.

Nie jest to pierwsza próba rozwiązania problemu konserwatora w sposób przybliżony, por. [Igle 74], ani zastosowania w tym celu aproksymacji dyfuzyjnej [Igle 65, Gave 73]; wymienione prace dotyczą jednak przypadków szczególnych (bardzo duża populacja klientów, dwufazowa obsługa), a ich wyniki nie są tu wykorzystane.

5.1. Analityczny opis stanów nieustalonych i ustalonych w modelu konserwatora

Zebrane tu dokładne wyniki posłużą do oceny błędów aproksymacji. Punkt (a) przedstawia podstawowy model konserwatora z wykładniczym rozkładem czasu pobytu klientów w źródle i dowolnym rozkładem czasu obsługi, w postaci uzyskanej przez Takácsa [Taká 62b]; punkt (b) korzysta z wyników [Mari 79a] i zawiera model konserwatora ograniczony do analizy stanów ustalonych, w którym rozkład $B(x)$ ma formę prawie ogólną (rozkład Coxa), a stanowisko obsługi objęte jest pętlą sprzężenia zwrotnego.

a) Stanowisko obsługi bez sprzężenia zwrotnego

Niech rozkład prawdopodobieństwa czasu pobytu klienta w źródle ma funkcję gęstości $a(x) = \lambda e^{-\lambda x}$, a rozkład czasu obsługi jest dowolny.

Proces $N(t)$ liczby klientów obecnych w stanowisku, obserwowany w momentach zakończenia obsługi kolejnych klientów, tworzy włożony łańcuch Markowa, w którym prawdopodobieństwo

$$p(n,t;0) = \Pr[N(t) = n \mid N(0) = 0]$$

ma transformatę Laplace'a

$$\bar{p}(n,s;0) = \int_0^\infty p(n,t;0)e^{-st}dt$$

w postaci [Taká 62]

$$\bar{p}(n,s;0) = \sum_{r=N-n}^{N} (-1)^{r-N+n}\binom{r}{N-n}\Theta_r(s) \qquad n = 0,1,\ldots,N, \tag{5.1}$$

gdzie:

$$\Theta_r(s) = \frac{N\lambda C_{r-1}(s)}{(s+r\lambda)\left[N\lambda \sum_{j=0}^{N-1}\binom{N-1}{j}\frac{1}{C_j(s)} + s\sum_{j=0}^{N}\binom{N}{j}\frac{1}{C_{j-1}(s)}\right]} \cdot$$

$$\cdot\left\{\left[\sum_{j=r}^{N-1}\binom{N-1}{j}\frac{1}{C_j(s)}\right]\left[\sum_{j=0}^{r}\binom{N-1}{j}\frac{1}{C_{j-1}(s)}\right] -\right.$$

$$- \left[\sum_{j=0}^{r-1} \binom{n}{j} \frac{1}{C_j(s)} \right] \left[\sum_{j=r+1}^{N-1} \binom{N-1}{j} \frac{1}{C_{j-1}(s)} \right] \Bigg\} +$$

$$+ \frac{\binom{N}{j} \sum_{j=0}^{N} \binom{N}{j} \frac{1}{C_j(s)}}{N\lambda \sum_{j=0}^{N-1} \binom{N-1}{j} \frac{1}{C_j(s)} + s \sum_{j=0}^{N} \binom{N}{j} \frac{1}{C_{j-1}(s)}},$$

$$C_j(s) = \prod_{i=0}^{j} \frac{\bar{b}(s+i)}{1 - \bar{b}(s+i)}, \qquad C_{-1} = 1,$$

$$\bar{b}(s) = \int_0^\infty b(t) e^{-st} dt.$$

Wyznaczania oryginału $p(n,t;0)$ wymaga numerycznej transformaty odwrotnej wzoru (5.1). Jeżeli $E[b] = \frac{1}{\mu} < \infty$, to istnieje niezależny od warunku początkowego stan ustalony:

$$p(n) = \lim_{t\to\infty} p(n,t;n_0) = \lim_{s\to 0} s\bar{p}(n,s;n_0) =$$

$$= \sum_{r=N-n}^{N} (-1)^{r-N+n} \binom{r}{N-n} \theta_r, \tag{5.2}$$

gdzie:

$$\theta_0 = 1, \quad \theta_r = \frac{NC_{r-1}}{r} \frac{\sum_{i=r-1}^{N-1} \binom{N-1}{j} \frac{1}{C_j}}{1 + N\frac{\lambda}{\mu} \sum_{j=0}^{N-1} \binom{N-1}{j} \frac{1}{C_j}}, \qquad r = 1,2,\ldots,N,$$

$$C_0 = 1, \quad C_j = \sum_{i=1}^{j} \frac{\bar{b}(i)}{1 - \bar{b}(i)}.$$

Jeżeli rozkład czasu obsługi ma postać sumy lub splotu rozkładów wykładniczych, to proces $N(t)$ można sprowadzić do łańcucha Markowa. Algorytm numerycznego rozwiązywania układu równań Chapmana-Kołmogorowa, wiążących stacjonarne prawdopodobieństwa stanów tego procesu, opracowano w [Stew 79]. W najprostszym przypadku, gdy $b(x) = \mu e^{-\mu x}$ prawdopodobieństwa stanów mają postać

$$p(n) = p(0) \left(\frac{\lambda}{\mu}\right)^n \frac{N!}{(N-n)!}, \qquad n = 1,\ldots,N,$$

$$p(0) = 1 + \left[\sum_{n=1}^{N} \left(\frac{\lambda}{\mu}\right)^n \frac{N!}{(N-n)!} \right]^{-1}. \tag{5.3}$$

b) <u>Stanowisko obsługi ze sprzężeniem zwrotnym</u>

Niech $B(t)$ jest rozkładem Coxa złożonym z F faz:

$$b(x) = \sum_{f=1}^{F} g_f \mu_1 e^{-\mu_1 x} * \ldots * \mu_f e^{-\mu_f x}, \tag{5.4}$$

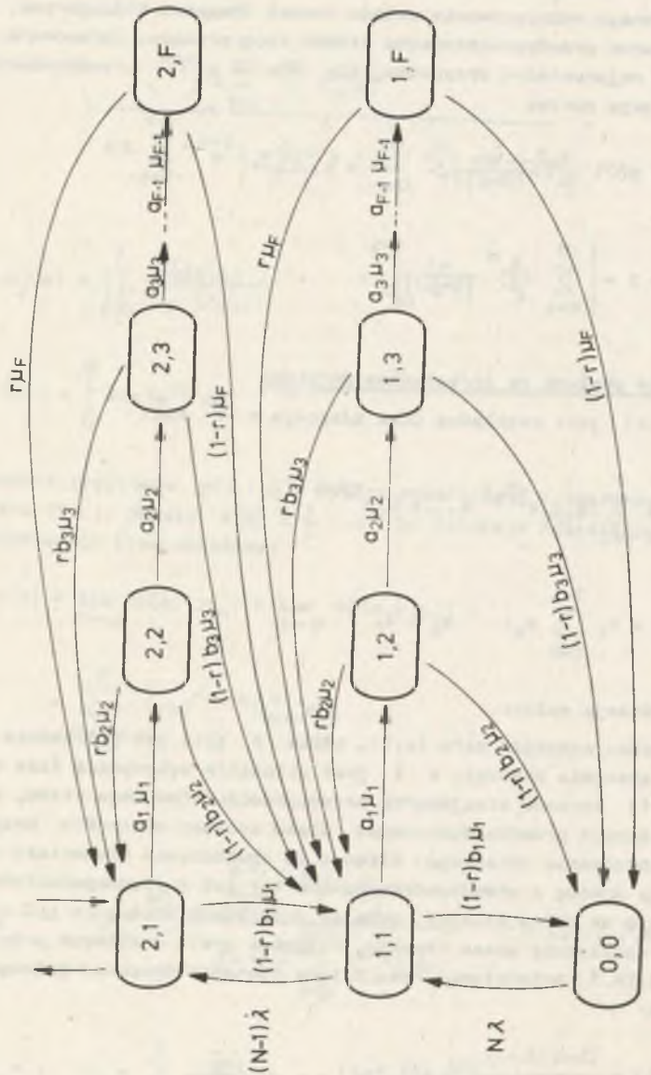
gdzie: $g_f = b_f \sum_{i=0}^{f-1} a_i, \quad a_0 \equiv 1,$

$*$ - jest operacją splotu.

Stan systemu wyznacza para (n,f), gdzie n jest jak poprzednio liczbą klientów w systemie obsługi, a f jest aktualnie wykonywaną fazę obsługi. Niech $p(n,f)$ oznacza stacjonarne prawdopodobieństwo tego stanu. W równaniach wiążących prawdopodobieństwa stanu systemu nietrudno uwzględnić przypadek sprzężenia zwrotnego: klienci po opuszczeniu stanowiska obsługi powracają do źródła z prawdopodobieństwem $1-r$ lub z prawdopodobieństwem r ustawiają się na końcu kolejki, żądając dodatkowej obsługi o tym samym co poprzednio rozkładzie czasu trwania. Fragment grafu możliwych przejść między stanami (n,f) przedstawia rys. 5.1, a równania Chapmana-Kołmogorowa mają postać

$$c(1,j) = \frac{(N-1)\lambda + \mu_{j+1}}{\mu_j a_j} c(1,j+1),$$

$$N\lambda = (1-r) \sum_{j=1}^{F} \mu_j (1-a_j) c(1,j) \qquad \text{dla} \quad n = 1,$$



Rys. 5.1. Fragment grafu przejść między stanami procesu N(t) w przypadku, gdy b(x) ma postać (5.4)
Fig. 5.1. A part of the graph of states of the process N(t) in the case when b(x) has the form (5.4)

oraz

$$c(n,j) = \frac{(N-n)\lambda + \mu_{j+1}}{\mu_j a_j}\, c(n,j+1) - \frac{N-n+1}{\mu_j a_j}\, c(n-1,\, j+1),$$

$$\left[(N-n+1)\lambda + \mu_1\right] c(n-1,1) = (1-r) \sum_{j=1}^{F} \mu_j (1-a_j) c(n,j) +$$

$$+ r \sum_{j=1}^{F} \mu_j (1-a_j) c(n-1,j) + (N-n+2)\lambda c(n-2,1)$$

$$\text{dla} \quad 2 \leq n \leq N \tag{5.5}$$

gdzie: $c(n,j) = \frac{p(n,j)}{p(0)}$.

Wzór (5.5) można po przekształceniach doprowadzić do zależności określającej rozwiązanie iteracyjne [Mari 79a]

$$p(n) = G(n)c(n,F) - H(n)p(n-1,F) \tag{5.6}$$

gdzie:

$$p(n) = \sum_{j=1}^{F} p(n,j), \qquad G(n) = \sum_{j=1}^{F} g(n,j),$$

$$g(n,j) = \frac{(N-n)\lambda + \mu_{j+1}}{\lambda_j a_j}\, g(n,j+1), \qquad g(n,F) = 1,$$

$$H(n) = \sum_{j=1}^{F} h(n,j), \quad h(n,j) = \frac{(N-n+\mu_{j+1})h(n,j+1) + (N-n+1)\lambda d(n-1,j+1)}{\mu_j a_j},$$

$$h(n,F) \equiv 0, \qquad d(n-1,j+1) = \frac{c(n-1,j+1)}{c(n-1,F)}.$$

5.2. Aproksymacja dyfuzyjna podstawowego modelu konserwatora - stan ustalony

Opis procesu dyfuzji

Proces $N(t)$ jest różnicą dwóch procesów: procesu wejściowego o wartości odpowiadającej liczbie klientów, którzy nadeszli do systemu w przedziale czasu $[0,t]$ i procesu wyjściowego o wartości odpowiadającej liczbie klientów, którzy w tym samym przedziale czasu opuścili system. Jeżeli oba procesy nie zależą od wartości n, jak ma to miejsce w modelu G /G /1, zmiany

$$\Delta N(t) = N(t + \Delta t) - N(t)$$

nie zależą od wartości N i w aproksymacji dyfuzyjnej procesu $N(t)$ przyjmowaliśmy stałe współczynniki β i α.

W przypadku ograniczonej populacji klientów proces wejściowy zależy oczywiście od liczby klientów już obecnych w stanowisku obsługi.

Przyjmiemy więc - podobnie jak ma to miejsce w opisie wielokanałowych stanowisk obsługi [Hala 78, Pujo 78, Kimu 83, Fili 84], gdzie z kolei proces wyjściowy zależy od wartości N - że parametry równania dyfuzji są zmienne, $\beta = \beta(x)$, $\alpha = \alpha(x)$. Przyjmiemy, że zmieniają się one w sposób skokowy:

$$\beta(x) = \beta_n, \quad \alpha(x) = \alpha_n \text{ dla } x \in (n-1,n], \quad n = 1,\ldots,N \tag{5.7}$$

Rozwiązanie równania dyfuzji jest mimo to ciągłe (por. [Mand 68]) i ma postać

$$f(\cdot) = f_n(\cdot), \quad x \in (n-1,n], \quad n = 1,\ldots,N.$$

W rozpatrywanym modelu podstawowym z populacją N jednakowych klientów i jednokanałowym stanowiskiem obsługi przyjmiemy

$$\begin{aligned} \beta_n &= (N-n+1)\lambda - \mu, \\ \alpha_n &= (N-n+1)\lambda C_a + \mu C_b, \qquad n = 1,\ldots,N. \end{aligned} \tag{5.8}$$

Jeżeli stanowisko obsługi jest k-kanałowe i np. $k < N$, to

$$\begin{aligned} \beta_n &= (N-n+1)\lambda - n\mu \\ \alpha_n &= (N-n+1)\lambda C_a + n\mu C_b \qquad 1 \leq n \leq k \\ \beta_n &= (N-n+1) - k\mu \\ \alpha_n &= (N-n+1)\lambda C_a + k\mu C_b \qquad k \leq n \leq N. \end{aligned} \tag{5.8a}$$

W przypadku gdy co najwyżej m spośród $N-n$ klientów obecnych w źródle może zażądać obsługi, a reszta tworzy pasywną w danym momencie rezerwę, przyjmiemy (dla stanowiska jednokanałowego)

$$\begin{aligned} \beta_n &= m\lambda - \mu, \qquad \alpha_n = m\lambda C_a + \mu C_b, \qquad 1 \leq n \leq N-m+1, \\ \beta_n &= (N-n+1)\lambda - \mu, \quad \alpha_n = (N-n+1)\lambda C_a + \mu C_b, \quad n \geq N-m+1. \end{aligned} \tag{5.8b}$$

Ponieważ $0 \leq N(t) \leq N$, przyjmiemy jako warunki brzegowe, podobnie jak przy opisie stanowiska G/G/1/N, bariery w punktach $x = 0$, $x = N$ i powroty elementarne odpowiednio do punktów $x = 1$, $x = N-1$; proces $X(t)$ po osiągnięciu wartości granicznej 0 (lub N) zachowuje ją przez czas określony rozkładem $l_0(x)$ (lub $l_N(x)$), po czym przeskakuje do punktu $x = 1$ ($x = N-1$) i znów jest procesem dyfuzji. Zakładając wykładnicze rozkłady $l_0(x) = \lambda_0 e^{-\lambda_0 x}$, $l_N(x) = \lambda_N e^{-\lambda_N x}$, piszemy

$$\frac{\partial f(x,t;x_0)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\alpha(x) f(x,t;x_0)\right] - \frac{\partial}{\partial x}\left[\beta(x) f(x,t;x_0)\right] +$$

$$+ p_0(t)\lambda_0\delta(x-1) + p_N(t)\lambda_N(x-N+1),$$

$$\frac{dp_0(t)}{dt} = \lim_{x\to 0}\left\{\frac{1}{2}\frac{\partial}{\partial x}\left[\alpha(x) f(x,t;x_0)\right] - \beta(x) f(x,t;x_0)\right\} - \lambda_0 p_0(t),$$

$$\frac{dp_N(t)}{dt} = -\lim_{x\to N}\left\{\frac{1}{2}\frac{\partial}{\partial x}\left[\alpha(x) f(x,t;x_0)\right] - \beta(x) f(x,t;x_0)\right\} - \lambda_N p_N(t), \tag{5.9}$$

oraz

$$\lim_{x\to 0} f(x,t;x_0) = \lim_{x\to N} f(x,t;x_0) = 0, \tag{5.10}$$

gdzie:

$$p_0(t) = \Pr[X(t) = 0]; \quad p_N(t) = \Pr[X(t) = N].$$

Przyjmijmy też, że parametry λ_0, λ_N (odwrotności średniego czasu pobytu w punktach $x = 0$, $x = N$) mają wartość

$$\lambda_0 = N\lambda = \Lambda, \quad \lambda_N = \mu. \tag{5.11}$$

W stanie ustalonym

$$\lim_{t\to\infty} f(x,t;x_0) = f(x), \quad \lim_{t\to\infty} p_0(t) = p_0, \quad \lim_{t\to\infty} p_N(t) = p_N,$$

a równania (5.9) po uwzględnieniu (5.7), (5.10), (5.11) przybierają postać

$$0 = \frac{\alpha_n}{2}\frac{d^2 f_n(x)}{dx^2} - \beta_n \frac{df_n(x)}{dx} + \Lambda p_0 \delta(x-1) + \mu p_N \delta(x-N+1), \quad n = 1,\ldots,N$$

$$\frac{\alpha_1}{2}\lim_{x\to 0}\frac{df_1(x)}{dx} = \Lambda p_0,$$

$$-\frac{\alpha_N}{2}\lim_{x\to N}\frac{df_N(x)}{dx} = \mu p_N. \qquad (5.12)$$

Dla ciągłej poza punktami $x = 0$, $x = N$ funkcji $f(x)$,

$$f_n(n) = f_{n+1}(n), \quad n = 1,\ldots,N-1,$$

przyjmując brak przepływu masy prawdopodobieństwa dla $x \in (1,N-1)$:

$$\frac{\alpha_n}{2}\frac{df_n(x)}{dx} - \beta_n f_n(x) = 0, \qquad \begin{array}{l} x \in (1,N-1), \\ n = 2,\ldots,N-1, \end{array}$$

otrzymujemy rozwiązanie

$$f_1(x) = \begin{cases} \dfrac{\Lambda p_0}{-\beta_1}\left(1 - e^{z_1 x}\right) & \text{dla } \beta_1 \neq 0 \\ \dfrac{2\Lambda p_0}{\alpha_1}\, x & \text{dla } \beta_1 = 0 \end{cases} \qquad 0 < x \leq 1$$

$$f_n(x) = \begin{cases} f_{n-1}(n-1)\, e^{z_n(x-n+1)} & \text{dla } \beta_n \neq 0 \\ f_{n-1}(n-1) & \text{dla } \beta_n = 0 \end{cases} \qquad n-1 \leq x \leq n, \quad n = 2,\ldots,N-1$$

$$f_N(x) = \begin{cases} \dfrac{\mu p_N}{\beta_N}\left[1 - e^{z_N(x-N)}\right] & \text{dla } \beta_N \neq 0 \\ \dfrac{-2\mu p_N}{\alpha_N}\,(x-N) & \text{dla } \beta_N = 0 \end{cases} \qquad N-1 \leq x < N$$

gdzie:

$$z_n = \frac{2\beta_n}{\alpha_n}. \qquad (5.13)$$

Ostatni z warunków ciągłości, $f_{N-1}(N-1) = f_N(N-1)$ oraz warunek normalizacyjny

$$p_0 + \sum_{n=1}^{N}\int_{n-1}^{n} f_n(x)dx + p_N = 1$$

pozwalają wyznaczyć p_0 i p_N.

Wzór (5.13) podaje przybliżone rozwiązanie zadania konserwatora; jako aproksymację rozkładu $p(n)$ liczby klientów w stanowisku obsługi przyjmiemy

$$p(0) = p_0, \quad p(n) = f_n(n), \quad n = 1,\ldots,N-1, \quad p(N) = p_N. \qquad (5.14)$$

W przybliżonych metodach analizy sieci zamkniętych często (np.: [Chan 75, Mari 79b, Shum 67]) rozpatruje się niezależnie każde ze stanowisk, przedstawiając resztę sieci jako jedno wypadkowe stanowisko obsługi lub zastępcze źródło klientów. Jeżeli rozpatrywane stanowisko objęte jest pętlą sprzężenia zwrotnego, zaleca się [Mari 79a] pozostawienie tej pętli w sposób jawny, nie uwzględniając jej w charakterystyce zastępczego źródła czy stanowiska. Ponieważ wydaje się, że przedstawiony model dyfuzyjny jest lepiej przystosowany do analizy pracy poszczególnych stanowisk w sieci niż modele stosowane do tej pory - M/M/1 w [Chan 75], M/G/1/N w [Shum 67], M_n/K/1 w [Mari 79b] - bierze bowiem jednocześnie pod uwagę skończoną liczbę klientów w sieci i dowolny charakter strumieni klientów między stanowiskami, rozpatrzymy poniżej przypadek stanowiska ze sprzężeniem zwrotnym.

Istnienie pętli sprzężenia można uwzględnić w modelu poprzez zmianę parametrów związanych z czasem obsługi. Jeżeli po wyjściu ze stanowiska klient zawraca do kolejki z prawdopodobieństwem r, to pomiędzy kolejnymi pobytami w źródle zgłoszeń przechodzi on przez stanowisko obsługi l-krotnie z prawdopodobieństwem r^{l-1}, całkowity czas obsługi ma więc rozkład

$$b_c(t) = (1-r)\sum_{l=1}^{\infty} r^{l-1} b(t)^{*l},$$

gdzie $*l$ oznacza l-krotną operację splotu.

Transformata Laplace'a rozkładu $b_c(t)$ ma postac

$$\bar{b}_c(s) = (1-r)\sum_{l=1}^{\infty} r^{l-1}\bar{b}(s)^l = \frac{(1-r)\bar{b}(s)}{1-r\,\bar{b}(s)},$$

skąd obliczamy parametry tego rozkładu :

$$\mu_c = -\left[\frac{d\bar{b}_c(s)}{ds}\bigg|_{s=0}\right]^{-1} = \mu r,$$

$$C_{bc} = \sigma^2_{bc}\mu_c^2 = E[b_c^2]\mu_c^2 - 1 = \frac{d^2\bar{b}_c(s)}{ds^2}\bigg|_{s=0}\mu_c^2 - 1 = C_b(1-r) + r,$$

czyli parametry równania dyfuzji mają w przypadku sprzężenia zwrotnego r postac

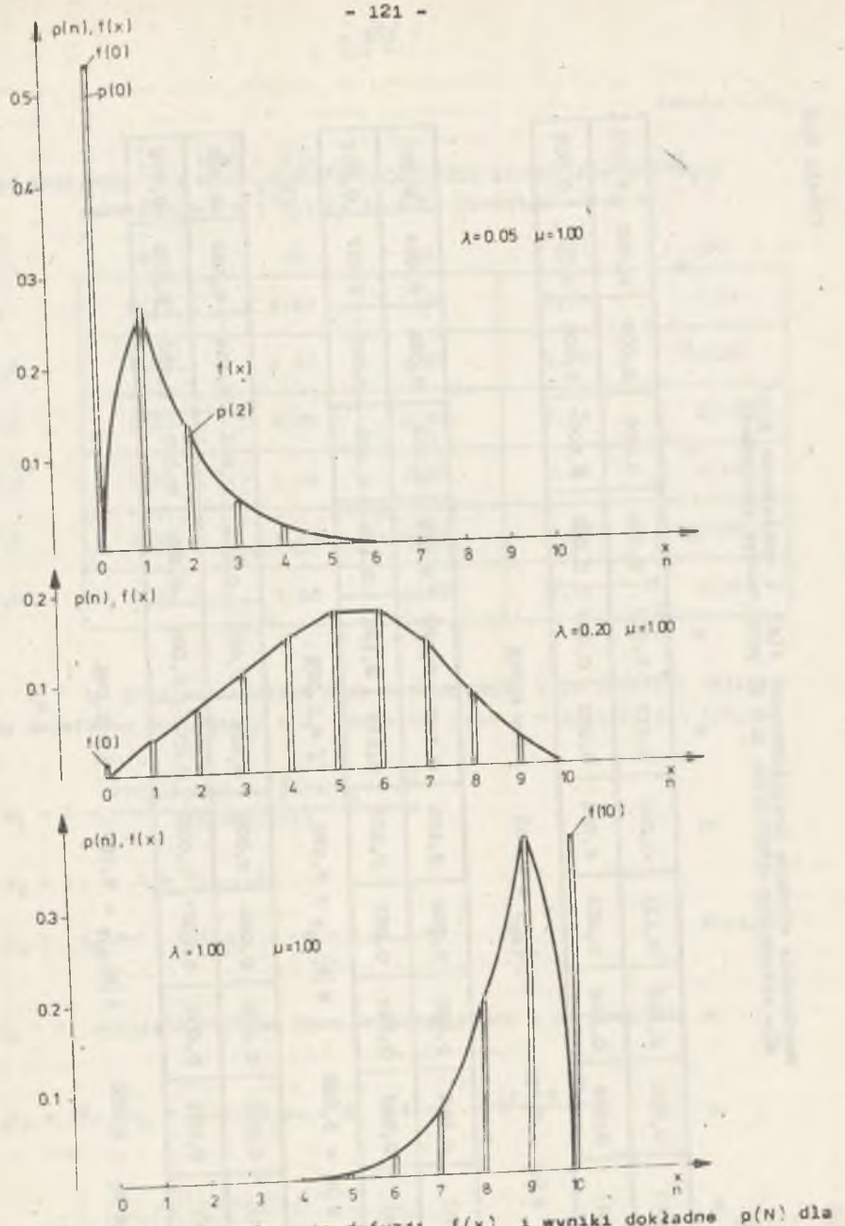
$$\beta_n = (N-n+1)\lambda - r,$$

$$\alpha_n = (N-n+1)\lambda C_a + r\mu\left[C_b(1-r) + r\right]. \qquad (5.15)$$

5.3. Wyniki numeryczne

Wykresy na rys. 5.2 oraz dane w tabeli 5.I zestawiają wyniki przybliżone (5.13), (5.14) z wynikami dokładnymi w przypadku wykładniczych rozkładów A(t), B(t). Wybrane przypadki wskazują, że aproksymacja dyfuzyjna nie tylko określa poprawnie wartości średnie $E[n]$, lecz bardzo dobrze dostosowuje się do zmiennego wraz z obciążeniem kształtu rozkładu $p(n)$. Dzięki zastosowaniu procesu elementarnych powrotów stopień wykorzystania stanowiska nie ma wpływu na błąd aproksymacji.

Tabela 5.II ilustruje wpływ wzrostu współczynnika zmienności rozkładu B(t) oraz liczebności populacji klientów na błąd aproksymacji. Wzrost wariancji rozkładu czasu obsługi powoduje wzrost błędu aproksymacji, co jest typowe dla metod przybliżonych - por. [Czac 77, 79], jednak wzrost ten jest znacznie powolniejszy niż w przypadku aproksymacji dyfuzyjnej stanowiska z nieskończonym źródłem klientów. Im większa populacja N klientów, tym mniejszy błąd względny aproksymacji. Zaznaczmy, iż wykorzystanie ogólnego rozwiązania analitycznego (5.2) dla modelu M/G/1//N było dla N = 50, 100 niemożliwe z powodu zbyt dużych błędów obliczeniowych - mimo stosowania podwójnej precyzji przy 35-bitowym słowie maszynowym. Ponieważ przyjęto rozkład Coxa jako reprezentację rozkładu dowolnego, wykorzystano w tym przypadku rozwiązanie (5.6).



Rys. 5.2. Rozwiązanie równania dyfuzji f(x) i wyniki dokładne p(N) dla stanowiska M/M/1//10, μ = 1 przy różnym obciążeniu

Fig. 5.2. Solution f(x) of the diffusion equation and the exact values of p(n) for the station M/M/1//10, μ = 1, for various utilizations of the station

Tabela 5.I

Porównanie wyników przybliżonych f(n) i dokładnych p(n) dla stanowiska M/M/1//10, μ = 1, przy różnym obciążeniu

	n	0	1	2	3	4	5	6	7	8	9	10
λ = 0,05	p(n)	0,538	0,269	0,121	0,048	0,017	0,005	0,001	0,000	0,000	0,000	0,000
	f(n)	0,538	0,262	0,123	0,052	0,020	0,007	0,002	0,000	0,000	0,000	0,000

$E[n] = 0,759 \qquad E[n]_{dyf} = 0,793 \qquad \varepsilon = 4,48\%$

λ = 0,20	p(n)	0,018	0,037	0,066	0,106	0,148	0,178	0,178	0,142	0,085	0,034	0,007
	f(n)	0,020	0,038	0,067	0,107	0,149	0,179	0,179	0,143	0,087	0,037	0,011

$E[n] = 5,092 \qquad E[n]_{dyf} = 5,195 \qquad \varepsilon = 2,02\%$

λ = 1,00	p(n)	0,000	0,000	0,000	0,000	0,001	0,003	0,015	0,061	0,184	0,368	0,368
	f(n)	0,000	0,000	0,000	0,000	0,001	0,006	0,021	0,070	0,190	0,370	0,370

$E[n] = 9,000 \qquad E[n]_{dyf} = 9,187 \qquad \varepsilon = 2,08\%$

Tabela 5.II

Błąd względny $\varepsilon = \dfrac{E[n] - E[n]_{dyf}}{E[n]}$ 100% dla stanowiska M/G/1//10 jako funkcja N i C_b; μ = 1, Nλ = constans = 0,2

C_b \ N	5	10	20	50	100
0,1	8,75	4,07	1,70	0,64	0,33
0,5	5,00	2,92	1,25	0,70	0,28
1,0	3,08	2,02	0,93	0,34	0,17
2,0	2,22	1,35	0,60	0,23	0,18
5,0	2,96	1,74	0,56	0,22	0,26
10,0	4,43	3,50	1,59	0,88	0,45

Dla $C_b < 1$ przyjęto rozkład Coxa n-tego rzędu o parametrach obliczanych na podstawie wartości μ i C_b [Saue 75] jako n = entier $(1 + 1/C_b)$.

$$a_1 = 1 - \frac{2nC_b + n - 2 - \sqrt{4 - 4nC_b + n^2}}{2(C_b+1)(n-1)}$$

$$a_i = 1, \qquad i = 2,\ldots,n$$

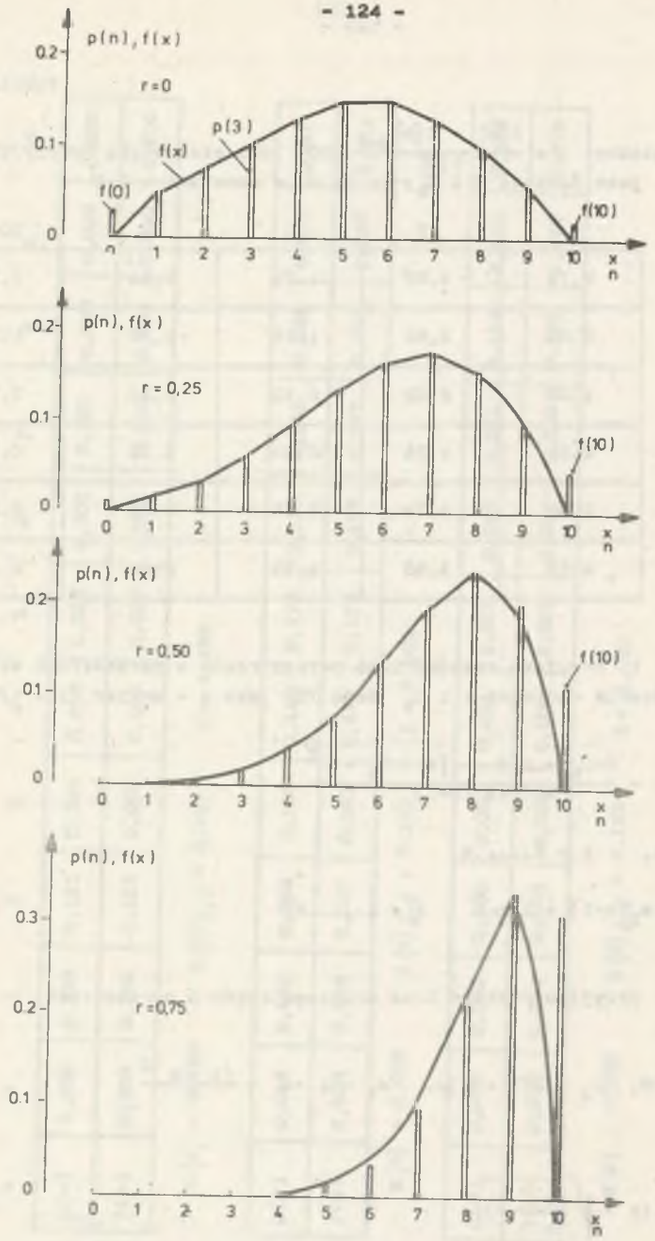
$$\mu_1 = [a_1(n-1) + 1]\mu \qquad i = 1,\ldots,n; \tag{5.16}$$

dla $C_b > 1$ przyjęto rozkład Coxa drugiego rzędu o parametrach

$$\mu_1 = 2\theta, \quad \mu_2 = 2(1-\theta)\mu, \quad a_1 = 1 - \theta - \frac{(1-\theta)^2}{\ }$$

$$\theta = \frac{1}{2}\left(1 - \sqrt{\frac{C_b - 1}{C_b + 1}}\right). \tag{5.17}$$

Rysunek 5.3 i tabela 5.III ilustrują wpływ sprzężenia zwrotnego i błąd aproksymacji w tym przypadku.



Rys. 5.3. Wyniki dokładne p(n) i rozwięzanie f(x) równania dyfuzji dla stanowiska M/M/1//10 ze sprzężeniem zwrotnym r

Fig. 5.3. The exact results p(n) and the solution f(x) of the diffusion equation for the station M/M/1//10 with a feedback r

Tabela 5.III

Porównanie wyników przybliżonych f(n) i dokładnych p(n) dla stanowiska M/G/1//10 (λ = 0,2, μ = 1, C_b = 2) ze sprzężeniem zwrotnym r

r	n	0	1	2	3	4	5	6	7	8	9	10
r = 0,25	p(n)	0,007	0,015	0,031	0,056	0,092	0,134	0,169	0,181	0,159	0,108	0,046
	f(n)	0,009	0,017	0,033	0,059	0,095	0,136	0,169	0,177	0,157	0,105	0,055

E [n] = 6,277 E [n] dyf = 6,371 ε = 1,50%

r	n	0	1	2	3	4	5	6	7	8	9	10
r = 0,50	p(n)	0,001	0,002	0,005	0,014	0,034	0,072	0,129	0,194	0,233	0,207	0,109
	f(n)	0,001	0,002	0,006	0,015	0,036	0,075	0,132	0,196	0,231	0,199	0,116

E [n] = 7,501 E [n] dyf = 7,641 ε = 1,87%

r	n	0	1	2	3	4	5	6	7	8	9	10
r = 0,75	p(n)	0,000	0,000	0,000	0,001	0,002	0,010	0,033	0,095	0,213	0,340	0,306
	f(n)	0,000	0,000	0,000	0,001	0,003	0,012	0,036	0,101	0,218	0,336	0,312

E [n] = 8,750 E [n] dyf = 8,932 ε = 2,08%

5.4. Model z wieloma klasami klientów

Opis modelu

Załóżmy obecnie, że zbiór N klientów jest podzielony na K klas; populację klientów opisuje wektor $\underline{N} = [N^1,\ldots,N^K]$, gdzie N^k oznacza liczbę klientów klasy k, $\sum_{k=1}^{K} N^k = N$. Klienci klasy k mają sobie właściwy dowolny rozkład $A^k(x)$ czasu pobytu w źródle i rozkład $B^k(x)$ czasu obsługi. Klasa klienta nie ma wpływu na kolejność obsługi, wszyscy tworzą wspólną kolejkę zarządzaną regulaminem naturalnym.

Wyniki analityczne dotyczące modelu konserwatora z wieloma klasami klientów są nieliczne; brak rozwiązania w jawnej postaci nawet w przypadku wykładniczych rozkładów $A^k(x)$, $B^k(x)$. [Kame 82] zawiera przegląd istniejących rezultatów i dowód najważniejszego chyba z nich, dotyczącego współczynnika wykorzystania stanowiska obsługi

$$(1-\rho)^{-1} = \sum_{d^1=0}^{1} \sum_{d^2=0}^{1} \cdots \sum_{d^K=0}^{1} \Big(\sum_{k=1}^{K} d^k\Big)! \prod_{k=1}^{K} (x^k)^{-d^k}, \quad x^k = \mu^k/\lambda,$$

ograniczony do przypadku wykładniczych rozkładów $A^k(x)$, $B^k(x)$, przy czym $N^k = 1$, $\lambda^k = \lambda$, dla $k = 1,\ldots,K$.

Dla wykładniczych rozkładów czasów obsługi można otrzymać rozwiązanie przybliżone stosując metodę wartości średnich przeznaczoną dla sieci typu BCMP [Reis 79], opisaną krótko w rozdz. I. Zaadaptujmy ją tu do potrzeb modelu konserwatora, tworząc sieć złożoną z K reprezentujących źródło stacji typu IS (stacji z nieskończoną liczbą kanałów obsługi - nie tworzy się w nich kolejka, czas pobytu jest równy czasowi obsługi) i z obsługującego klientów wszystkich klas stanowiska z regulaminem naturalnym. To ostatnie stanowisko nie mieści się w założeniach sieci typu BCMP, dlatego otrzymane wyniki są przybliżone. Klienci danej klasy odwiedzają sobie tylko właściwe stanowisko IS i stanowisko z regulaminem naturalnym. Algorytm metody polega na rekurencyjnym wykonaniu, poczynając od populacji $1,0,\ldots,0$ i powiększając ją każdorazowo o jednego klienta, aż do pełnej populacji $\underline{N}$, następujących kroków:

(1) W momencie, gdy w zamkniętej sieci BCMP o populacji $\underline{N}$ klient klasy k nadchodzi do stanowiska obsługi, stacjonarny rozkład kolejki, którą zastaje w stanowisku, jest taki sam jak stacjonarny rozkład tej kolejki w sieci, w której brak rozpatrywanego klienta, a więc o populacji $\underline{N} - \underline{e}^k$, gdzie $\underline{e}^k$ jest wektorem posiadającym K składowych, z których k-ta jest równa jedności, a pozostałe są zerowe. Czas pobytu klienta w stanowisku z regulaminem naturalnym obliczamy więc jako:

$$T^k(\underline{N}) = \frac{1}{\mu^k} + \sum_{c=1}^{K} \frac{1}{\mu^c} E[n^c(\underline{N} - \underline{e}^k)]$$

zakładając jako warunek początkowy $E[n^c(\underline{0})] = 0$;

(2) Przepustowość (liczba klientów obsłużonych w jednostce czasu) systemu dla klientów klasy k, $\mathcal{P}^k(\underline{N})$, wynosi zgodnie z prawem Little'a odniesionym do źródła i stanowiska łącznie

$$\mathcal{P}^k(\underline{N}) = \frac{N^k}{\frac{1}{\lambda^k} + T^k(\underline{N})v^k}$$

gdzie mianownik wyraża czas obiegu klienta w układzie źródło - stanowisko obsługi, v^k jest liczbą wizyt klienta w stanowisku obsługi, przypadających na jedną wizytę w źródle; jeżeli stanowisko objęte jest pętlą sprzężenia zwrotnego r^k, to $v^k = 1/(1 - r^k)$;

(3) Zgodnie z prawem Little'a dla stanowiska obsługi średnia długość kolejki wynosi

$$E[n^k(\underline{N})] = \mathcal{P}^k(\underline{N})T^k(\underline{N})$$

oraz współczynnik wykorzystania stanowiska

$$\rho^k = \mathcal{P}^k(\underline{N})v^k.$$

Inne możliwe aproksymacje rozwiązania, to zastąpienie stanowiska obsługi z regulaminem naturalnym dla wielu klas klientów:

- stanowiskiem z regulaminem podziału procesora, dla którego istnieje rozwiązanie analityczne w ramach modelu BCMP, lub
- stanowiskiem, w którym klienci są losowo wybierani z kolejki; wymagane jest wtedy numeryczne rozwiązanie układu równań Chapmana-Kołmogorowa dla prawdopodobieństw stanów w tym stanowisku, lecz liczba równań jest znacznie mniejsza niż dla stanowiska z regulaminem naturalnym (por. [Rubi 84]).

W przedstawionym poniżej modelu dyfuzyjnym proces dyfuzji służy jako aproksymacja liczby $n = \sum_{k=1}^{K} n^k$ klientów wszystkich klas obecnych w stanowisku obsługi. Niech q^k oznacza prawdopodobieństwo, że klient znajdujący się w stanowisku obsługi należy do klasy k, a ψ^k jest prawdopodobieństwem, że klient znajdujący się w źródle należy do klasy k. Zakładamy, że wariancja ogólnej liczby klientów nadchodzących do stanowiska

w jednostkowym czasie jest sumą wariancji liczby klientów poszczególnych klas przybyłych w tym czasie:

$$\lambda^3\sigma_a^2 = \sum_{k=1}^{K} \vartheta^k(\lambda^k)^3(\sigma_a^k)^3$$

lub

$$\lambda c_a = \sum_{k=1}^{K} \vartheta^k\lambda^k c_a^k,$$

a rozkład czasu obsługi jest wazoną sumą rozkładów właściwych poszczególnym klasom

$$b(x) = \sum_{k=1}^{K} q^k b^k(x),$$

skąd obliczamy wartość średnią i wariancję, a następnie współczynnik zmienności tego rozkładu

$$\frac{1}{\mu} = E[b] = \sum_{k=1}^{K} q^k E[b^k] = \sum_{k=1}^{K} q^k \mu^k,$$

$$c_b = \mu^2\sigma_b^2 = \mu^2 \sum_{k=1}^{K} \frac{q^k}{(\mu^k)^2}(c_b^k + 1) - 1.$$

Jako współczynniki równania dyfuzji (5.12) przyjmiemy więc

$$\beta_i = (N-i+1)\sum_{k=1}^{K} \vartheta^k\lambda^k - \mu$$

$$\alpha_i = (N-i+1)\lambda c_a + \mu c_b . \qquad (5.18)$$

W celu określenia wartości q^k, ϑ^k proponujemy następującą procedurę iteracyjną:

(1) przyjąć jako pierwsze przybliżenie $q^k = \vartheta^k = N^k/N$, co jest wartością dokładną, jeżeli wszystkie klasy są charakteryzowane tymi samymi rozkładami;

(2) określić za pomocą aproksymacji dyfuzyjnej (5.13), (5.14) ze współczynnikami (5.18) rozkład p(n) i współczynnik wykorzystania systemu $\rho = 1 - p(0)$;

(3) stosując wzór Little'a obliczyć jednakowy dla klientów wszystkich klas średni czas czekania w kolejce

$$E[w] = \frac{E[n] - \rho}{\rho\mu},$$

gdzie $E[n] - \rho$ jest średnią długością kolejki bez klienta aktualnie obsługiwanego i $\rho\mu = \lambda$ jest przepustowością stanowiska obsługi;

(4) obliczyć nowe wartości ϑ^k, q^k

$$\vartheta^k = \frac{N^k\frac{1}{\lambda^k}}{\frac{1}{\lambda^k} + E[w] + \frac{1}{\mu^k}} \Bigg/ \left[\sum_{k=1}^{K} \frac{N^k\frac{1}{\lambda^k}}{\frac{1}{\lambda^k} + E[w] + \frac{1}{\mu^k}}\right],$$

$$q^k = \frac{(E[w] + \frac{1}{\mu^k})N^k}{\frac{1}{\lambda^k} + E[w] + \frac{1}{\mu^k}} \Bigg/ \left[\sum_{k=1}^{K} \frac{(E[w] + \frac{1}{\mu^k})N^k}{\frac{1}{\lambda^k} + E[w] + \frac{1}{\mu^k}}\right];$$

(5) powrócić do punktu (2), jeżeli wyniki otrzymane w kolejnych iteracjach nie są podobne.

Brak dowodu zbieżności tej procedury - podobnie jak nie udowodniono zbieżności wielu innych stosowanych w kolejkowych modelach systemów komputerowych algorytmów iteracyjnych, np. [Chan 82, Mari 79, Schw 79, Zaho 80], których użyteczność i poprawne funkcjonowanie potwierdziła praktyka obliczeniowa. W każdym z około stu obliczonych przykładów, przy doborze bardzo różnych parametrów, liczba iteracji powyższego algorytmu, po których obliczone w dwóch kolejnych iteracjach wartości

$$\beta_1 = N\sum_{k=1}^{K} \vartheta^k\lambda^k - \left(\sum_{k=1}^{K} q^k/\mu^k\right)^{-1}$$

różniły się mniej niż o 10^{-6}, co przyjęto jako kryterium końca obliczeń, nie przekraczała dziesięciu. Rozkład liczby klientów poszczególnych klas obecnych w stanowisku obliczamy na podstawie p(n) i rozkładu dwumianowego

$$p^k(\vartheta) = \sum_{n=\vartheta}^{N^k} p(n) \sum_{k=1}^{K} (q^k)^{\vartheta}(1-q^k)^{n-\vartheta}.$$

Wartości średnie charakterystyczne dla klasy k można określić w prosty sposób, np.:

$$\rho^k = q^k \rho, \quad E[n^k] = q^k E[n],$$

$$T^k = E[w] + \frac{1}{\mu^k}, \quad \mathcal{P}^k = E[n^k]/T^k.$$

Wyniki numeryczne

Dwa przykłady numeryczne ilustrują dokładność metody wartości średnich i aproksymacji dyfuzyjnej w przypadku wielu klas klientów. Wyniki dokładne otrzymano rozwiązując numerycznie układ równań Chapmana-Kołmogorowa wiążących prawdopodobieństwa stanu systemu.

Zwróćmy uwagę na fakt, iż w obu przykładach liczba stanów systemu, a więc liczba niewiadomych i liczba równań, wynosiła kilkaset.

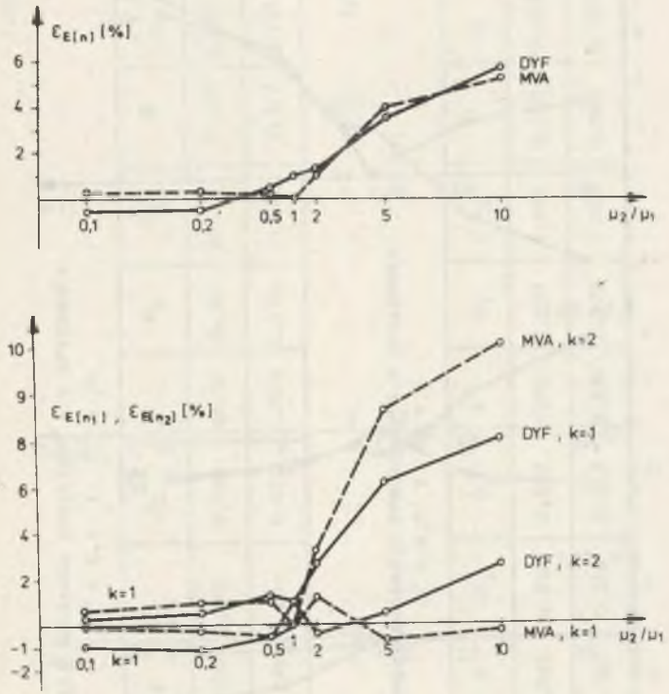
Rozwiązanie takiego układu równań, a także samo ich ułożenie, jest już problemem samym w sobie (korzystano z pakietu obliczeniowego QNAP [Vera 84]) i zmusza do wyboru przykładowych systemów z małą populacją klientów, gdzie, jak świadczą o tym wyniki poprzedniego rozdziału, błędy aproksymacji dyfuzyjnej są największe. Zamieszczone przykłady są więc raczej surowym sprawdzianem aproksymacji dyfuzyjnej.

Przykład 1

Populacja obejmuje dwie klasy klientów, k = 2, $\underline{N} = [5,5]$. Czas pobytu w źródle i czas obsługi mają rozkład wykładniczy dla obu klas klientów, przy czym $\lambda^1 = \lambda^2 = 0{,}2$, $\mu^1 = 1$. Obliczenia przeprowadzono dla wartości $\mu^2 = 0{,}1,\ 0{,}2,\ 0{,}5,\ 1,\ 2,\ 5,\ 10$ - a więc wzajemny stosunek średnich czasów obsługi klientów obu klas zmienia się w granicach 1-100. Pociąga to za sobą duże zmiany parametrów pracy systemu. Tabela 5.IV podaje dla skrajnych wartości parametru podstawowe wyniki dokładne charakteryzujące pracę systemu: średnią liczbę klientów obecnych w systemie $E[n]$, czas pobytu w systemie T, wykorzystanie stanowiska obsługi ρ, przepustowość $\mathcal{P}$. Wyniki podane są dla obu klas klientów łącznie i dla każdej klasy osobno. Wykresy na rys. 5.4 i 5.5 przedstawiają jako funkcję μ^2/μ^1 błąd względny $\varepsilon = (x_{przybl} - x_{dokł})/x_{dokł}$, popełniony przy obliczaniu tych wielkości metodą aproksymacji dyfuzyjnej i metodą wartości średnich.

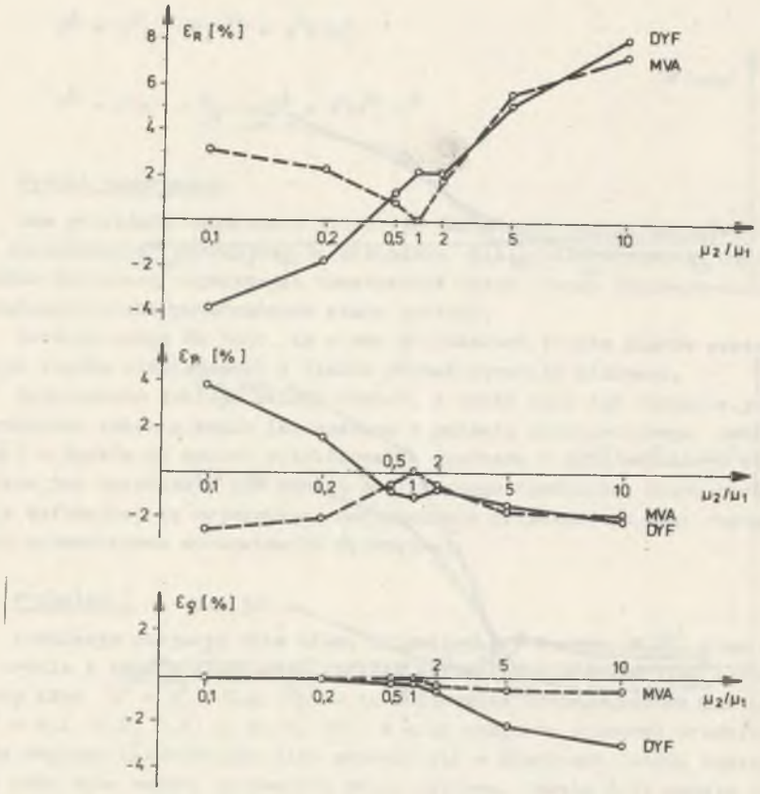
Przykład 2

W systemie są dwie równoliczne klasy klientów, $\underline{N} = [4,4]$, ich czas pobytu w źródle opisany jest tym samym rozkładem wykładniczym o parametrze $\lambda = \lambda^1 = \lambda^2 = 0{,}1$. Czas obsługi klientów klasy pierwszej ma rozkład wykładniczy, $\mu^1 = 1$, czas obsługi klientów klasy drugiej ma rozkład



Rys. 5.4. Błąd względny oceny średniej długości kolejki dla dwóch klas klientów łącznie (a) i dla każdej klasy osobno (b), dokonanej za pomocą aproksymacji dyfuzyjnej (linia ciągła) i metody wartości średnich (linia przerywana) dla danych z Przykładu 1

Fig. 5.4. Example 1; the relative error of the mean queue length for two classes of customers taken together (a) and for each class separately (b) - results of the diffusion approximation (solid line) and of the mean value analysis (broken line)



Rys. 5.5. Błąd względny oceny średniego czasu pobytu w stanowisku R, przepustowości $\mathfrak{P}$ i wykorzystania stanowiska ρ dla klientów obu klas łącznie, dokonanej za pomocą aproksymacji dyfuzyjnej (linia ciągła) i metody wartości średnich (linia przerywana) dla danych z Przykładu 1

Fig. 5.5. Example 1: the relative error of the reaction time R, of the throughput $\mathfrak{P}$ and of the utilization of the server ρ for the customers of both classes taken together - results of the diffusion approximation (solid line) and of the mean value analysis (broken line)

Tabela 5.IV

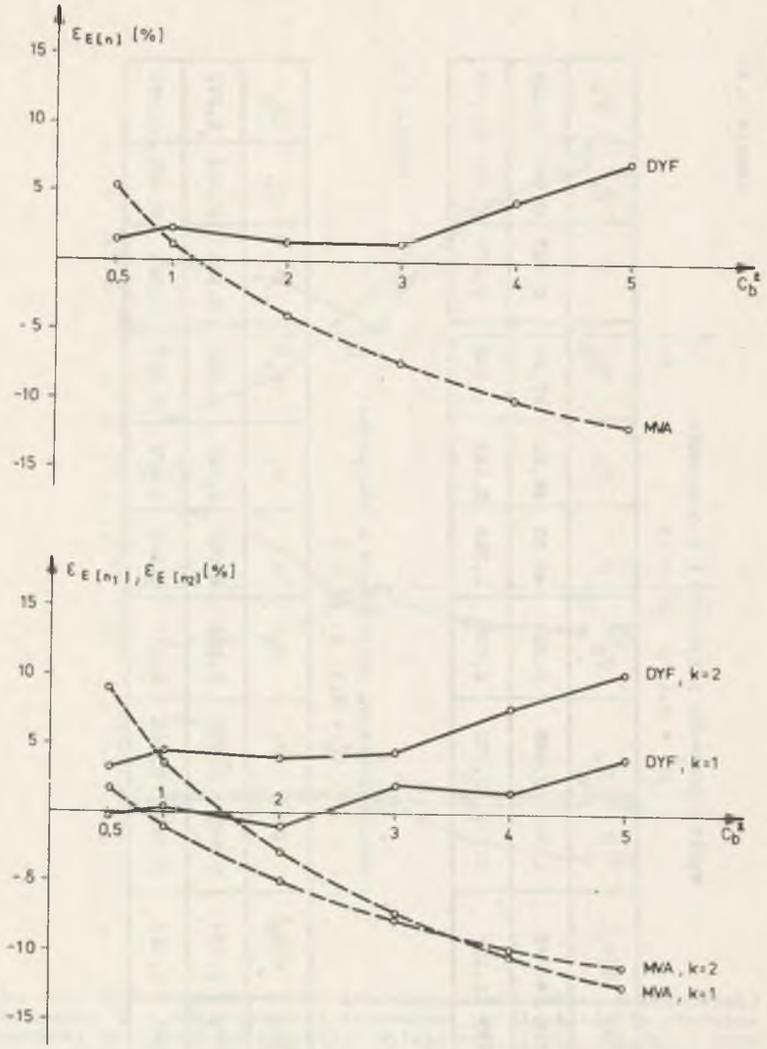
Wyniki dokładne przykładu 1 w przypadku
$\mu_1^2 = 0{,}1$ 1 $\mu_2^2 = 10$

μ_2^2	E[n]	E[n_1]	E[n_2]	ρ	ρ_1	ρ_2	R	R_1	R_2	$\mathfrak{P}$	$\mathfrak{P}_1$	$\mathfrak{P}_2$
0,1	9,066	4,518	4,548	1,000	0,096	0,904	48,52	46,82	50,34	0,187	0,096	0,090
10,0	2,613	1,495	1,117	0,779	0,701	0,078	1,768	2,134	1,438	1,477	0,701	0,777

Tabela 5.V

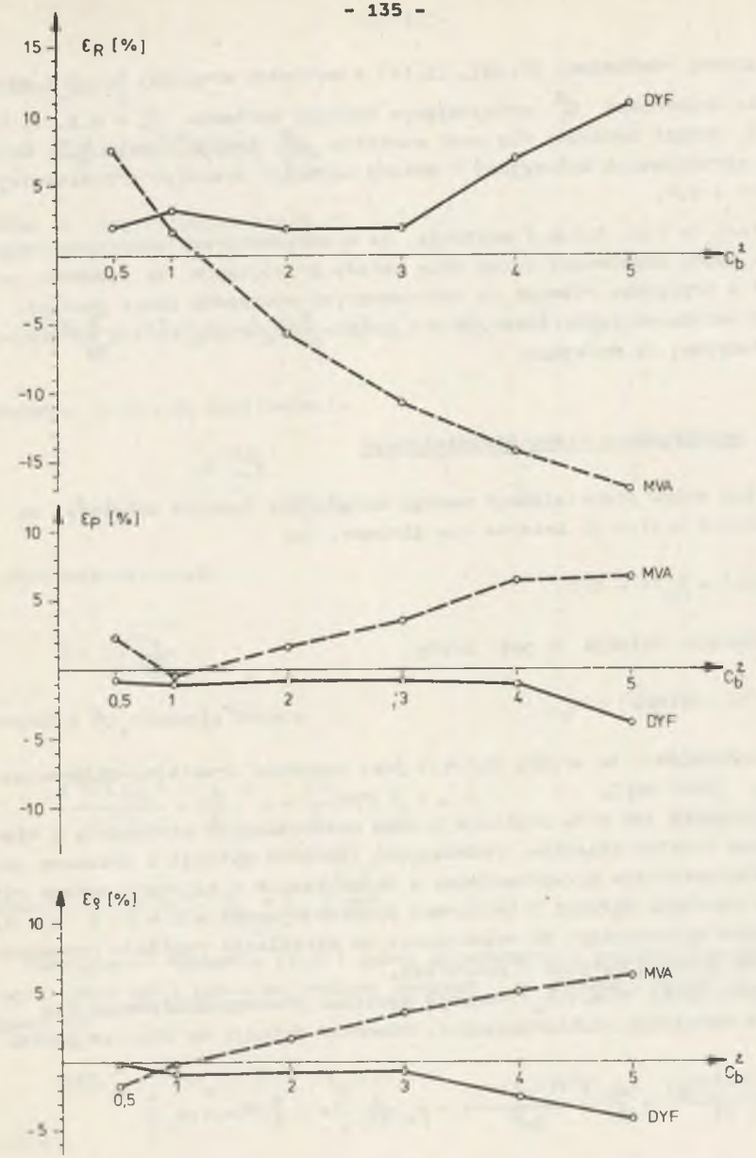
Wyniki dokładne przykładu 2 w przypadku
$C_b^2 = 0{,}5$ 1 $C_b^2 = 5$

C_b	E[n]	E[n_1]	E[n_2]	ρ	ρ_1	ρ_2	R	R_1	R_2	$\mathfrak{P}$	$\mathfrak{P}_1$	$\mathfrak{P}_2$
0,5	2,313	1,079	1,234	0,845	0,292	0,553	4,067	3,693	4,462	0,569	0,292	0,277
5,0	2,761	1,347	1,413	0,783	0,265	0,517	5,269	5,079	5,463	0,524	0,265	0,259



Rys. 5.6. Błąd względny oceny średniej długości kolejki dla dwóch klas klientów łącznie (a) i dla każdej klasy osobno (b), dokonanej za pomocą aproksymacji dyfuzyjnej (linia ciągła) i metody wartości średnich (linia przerywana) dla danych z Przykładu 2

Fig. 5.6. Example 2 : the relative error of the mean queue length for both classes of customers taken together (a) and considered separately (b) - results of the diffusion approximation (solid line) and of the mean value analysis (broken line)



Rys. 5.7. Błąd względny oceny średniego czasu pobytu w stanowisku R, przepustowości $\mathcal{P}$ i wykorzystania stanowiska ϱ dla klientów obu klas łącznie, dokonanej za pomocą aproksymacji dyfuzyjnej (linia ciągła) i metody wartości średnich (linia przerywana) dla danych z Przykładu 2

Fig. 5.7. Example 2: the relative error of the mean reaction time R, of the thronghput $\mathcal{P}$ and of the utilisation of the server ϱ for two classes of customers taken together - results of the diffusion approximation (solid line) and of the mean value analysis (broken line)

Coxa opisany równaniami (5.16), (5.17) o wartości średniej $\frac{1}{\mu} = 2$ i współczynniku zmienności C_b^2 przyjmującym kolejno wartości $C_b^2 = 0{,}5,\ 1,\ 2,\ 3,\ 4,\ 5$. Wyniki dokładne dla tych wartości C_b^2 podaje tabela 5.V, błędy metody aproksymacji dyfuzyjnej i metody wartości średnich przedstawiają rys. 5.6 i 5.7.

Wykresy na rys. 5.4-5.7 wskazują, że w przypadku wykładniczych czasów obsługi błędy popełnione przez obie metody przybliżone są podobne, natomiast w przypadku różnych od wykładniczych rozkładów czasu obsługi, których metoda wartości średnich nie uwzględnia, błędy metody aproksymacji dyfuzyjnej są mniejsze.

5.5. Aproksymacja stanu nieustalonego

Analiza stanu nieustalonego wymaga uproszczeń modelu: założymy, że współczynnik dryftu β zmienia się liniowo, np.

$$\beta(x) = \beta_0(x - N/2),$$

a współczynnik dyfuzji α jest stały

$$\alpha(x) = \alpha(N/2) = \alpha_0,$$

a więc przyjmiemy, że proces dyfuzji jest procesem Ornsteina-Uhlenbecka (por. np. [Bhar 60]).

Analogicznie jak przy analizie stanów nieustalonych stanowiska z nieskończonym źródłem klientów, rozwiązanie równania dyfuzji z procesem powrotów elementarnych przeprowadzimy w dwóch fazach - najpierw podamy rozwiązanie równania dyfuzji z barierami pochłaniającymi w $x = 0$ i $x = N$, a następnie wykorzystamy to rozwiązanie do określania rozkładu prawdopodobieństwa procesu dyfuzji z powrotami.

Oznaczmy przez $\varphi(x,t;x_0)$ funkcję gęstości prawdopodobieństwa dla procesu z barierami pochłaniającymi. Równanie dyfuzji ma obecnie postać

$$\frac{\partial\varphi(x,t;x_0)}{\partial t} = \frac{\alpha_0}{2}\frac{\partial^2\varphi(x,t;x_0)}{\partial x^2} - \beta_0\frac{\partial}{\partial x}\left[(x - \tfrac{N}{2})\varphi(x,t;x_0)\right],$$

$$\varphi(0,t;x_0) = \varphi(N,t;x_0) = 0. \qquad (5.19)$$

Po podstawieniu $\xi = N/2$ poszukujemy rozwiązania równania dyfuzji typową dla parabolicznych równań cząstkowych metodą rozdzielenia zmiennych [Bhar 60]. Zakładamy

$$\varphi(\xi,t;\xi_0) = X(\xi;\xi_0)T(t)$$

otrzymując

$$\frac{dT(t)}{dt} + D^2\frac{\alpha_0}{2}T(t) = 0,$$

gdzie D jest dowolną stałą,
oraz

$$\frac{\alpha_0}{2}\frac{d^2}{d\xi^2}X(\xi;\xi_0) - \beta_0\xi\frac{d}{d\xi}X(\xi;\xi_0) + (\beta_0 + D^2)X(\xi;\xi_0) = 0 \qquad (5.20)$$

Równanie (5.20) po podstawieniu

$$X(\xi;\xi_0) = e^{-\xi^2\frac{2\beta_0}{\alpha_0}}Y(\xi;\xi_0)$$

i zamianie zmiennych

$$\xi = \zeta\sqrt{\frac{\alpha_0}{2\beta_0}}$$

prowadzi do równania Webera

$$\frac{d^2Y(\zeta;\zeta_0)}{d\zeta^2} + (\tfrac{1}{2} + \lambda - \tfrac{\zeta^2}{4})Y(\zeta,\zeta_0) = 0, \qquad (5.21)$$

gdzie: $\lambda = D^2/\beta_0$, $\zeta_0 = \xi_0\sqrt{\frac{2\beta_0}{\alpha_0}}$.

Rozwiązanie równania (5.21) można przedstawić w postaci funkcji Webera (por. [Bhar 60]) lub - co zrobimy poniżej - za pomocą funkcji y_e, y_o [Swee 70]

$$Y(\zeta;\zeta_0) = Ay_e(\lambda,\zeta) + By_o(\lambda,\zeta),$$

gdzie:

$$y_e(\lambda,\zeta) = e^{-\zeta^2/4}K(-\tfrac{\lambda}{2},\tfrac{1}{2},\tfrac{\zeta^2}{2}),$$

$$y_o(\lambda,\zeta) = e^{-\zeta^2/4}K(\tfrac{\lambda+1}{2},\tfrac{3}{2},\tfrac{\zeta^2}{2}),$$

a K(.) jest funkcją Kummera [Magn 49],

$$K(c,d,e) = 1 + \frac{c}{d}\frac{e}{1!} + \frac{c(c+1)e^2}{d(d+1)2!} + \frac{e(c+1)(c+2)e^3}{d(d+1)(d+2)3!} + \ldots$$

Po uwzględnieniu warunków brzegowych, które w przypadku barier ustawionych symetrycznie wokół zera mają postać

$$y_o(l,\zeta_{N/2})y_e(l,\zeta_{N/2}) = 0, \tag{5.22}$$

rozwiązanie równania (5.21) przybiera postać [Swee 79]

$$\varphi(x,t;x_o) = \sum_{j=1}^{\infty} A_j(x,x_o)e^{\beta_o l_j t} + \sum_{k=1}^{\infty} B_k(x,x_o)e^{-\beta_o l_k t} \tag{5.23}$$

lub po transformacji Laplace'a

$$\bar{\varphi}(x,s;x_o) = \sum_{j=1} \frac{A_j(x,x_o)}{\beta_o l_j + s} + \sum_{k=1} \frac{B_k(x,x_o)}{\beta_o l_k + s}. \tag{5.24}$$

gdzie:

l_j są pierwiastkami równania $y_e(l,\zeta_{N/2}) = 0$,

l_k są pierwiastkami równania $y_o(l,\zeta_{N/2}) = 0$,

$$A_j(x,x_o) = \exp\left[-\tfrac{1}{4}(\zeta^2 - \zeta_o^2)\right] y_e(l_j,\zeta_o)y_e(l_j,\zeta)\left[\int_{-\zeta_{N/2}}^{\zeta_{N/2}} y_e^2(l_j,\zeta)d\zeta\right]^{-1},$$

$$B_k(x,x_o) = \exp -\tfrac{1}{4}(\quad^2 - \quad_o^2)\; y_o(l_k,\ _o)y_o(l_k,\ )\left[\int_{-\zeta_{N/2}}^{\zeta_{N/2}} y_o^2(l_k,\zeta)d\zeta\right]^{-1},$$

$$-\zeta_{N/2} \le \zeta \le \zeta_{N/2}, \quad -\zeta_{N/2} \le \zeta_o \le \zeta_{N/2},$$

$$\zeta = (x - \tfrac{N}{2})\sqrt{\frac{2\beta_o}{\alpha_o}}, \quad \zeta_o = (x_o - \tfrac{N}{2})\sqrt{\frac{2\beta_o}{\alpha_o}}.$$

Rozwiązanie (5.23) możemy teraz wykorzystać do opisu procesu dyfuzji z powrotami elementarnymi z punktu x = 0 do x = 1 i z punktu x = N do x = N-1. Znamy już obieg masy prawdopodobieństwa w układzie: punkt x = 0,

przedział (0,N), punkt x = N. Postać strumieni $\gamma_0(t)$, $\gamma_N(t)$ masy prawdopodobieństwa wpływającej do barier w x = 0 i x = N jest opisana równaniami (3.27), należy w nich tylko uwzględnić $\varphi(x,t;x_o)$ daną przez (5.23), a postać strumieni $g_1(t)$, $g_{N-1}(t)$ masy prawdopodobieństwa pojawiającej się w punktach x = 1, x = N-1 w wyniku procesu powrotów określają równania (3.26).

Znając $g_1(t)$, $g_{N-1}(t)$ określamy rozwiązanie: gęstość prawdopodobieństwa $f(x,t;x_o)$ dla procesu dyfuzji z powrotami

$$f(x,t;x_o) = \varphi(x,t;x_o) + \int_0^t g_1(\tau)\varphi(x,t-\tau;1)d\tau$$

$$+ \int_0^t g_{N-1}(\tau)\varphi(x,t-\tau;N-1)d\tau. \tag{5.25}$$

W praktyce, podobnie jak poprzednio, trzeba równania (3.26), (3.27) poddać transformacie Laplace'a, wyznaczyć $\bar{g}_1(s)$, $\bar{g}_{N-1}(s)$ i poszukiwać oryginału funkcji

$$\bar{f}(x,s;x_o) = \bar{\varphi}(x,s;x_o) + \bar{g}_1(s)\bar{\varphi}(x,s;1) + \bar{g}_{N-1}(s)\bar{\varphi}(x,s,N-1)$$

na drodze numerycznej.

Podobnie jak w rozdz. III i IV, można też uwzględnić warunek początkowy w postaci funkcji $\Psi(x)$ oraz opisać reakcję systemu na zmianę warunków pracy (zmianę wartości współczynników równania dyfuzji).

UWAGI KOŃCOWE

W pracy przedstawiono metodę rozwiązywania równań dyfuzji z powrotami elementarnymi oraz zbadano na przykładzie kilku podstawowych w teorii kolejek modeli, jak wprowadzenia do aproksymacji dyfuzyjnej warunków brzegowych w postaci powrotów elementarnych zwiększa dokładność tej aproksymacji. Zaproponowana metoda polega na przedstawieniu poszukiwanego rozwiązania jako superpozycji rozwiązań równania dyfuzji z warunkiem brzegowym w postaci ekranu pochłaniającego i pozwala na analizę stanów nieustalonych, która zawsze stwarzała duże trudności teorii masowej obsługi.

Dla uzyskania wyników liczbowych metoda ta wymaga numerycznej inwersji transformat Laplace'a - co jednak nie stwarza trudności obliczeniowych ani nie zmienia dokładności aproksymacji.

Analiza stanów nieustalonych nie jest tylko problemem teoretycznym: zmiany warunków pracy systemu komputerowego na skutek zmian jego obciążenia lub zmian konfiguracji, krótkotrwałe awarie powodujące spiętrzenie oczekujących zadań i następne stopniowe rozładowanie kolejek utworzonych w tym czasie, adaptacyjny wybór drogi pakietu w sieci telekomunikacyjnej w zależności od aktualnego obciążenia poszczególnych węzłów komutacji - to przykłady procesów, których modelowanie wymaga analizy stanów nieustalonych. Opis stanów nieustalonych w stanowiskach obsługi jest jednak znacznie bardziej złożony od opisu stanów ustalonych, co jest powodem słabego rozwoju odpowiednich modeli.

Rozpatrywane w pracy przykłady stanów nieustalonych dotyczą odpowiedzi systemu obsługi na skokową zmianę warunków pracy - badano dynamikę tych zjawisk i wyznaczano ich stałą czasową w funkcji obciążenia stanowiska i w funkcji wariancji rozkładu czasów między zgłoszeniami klientów i czasu obsługi.

Aproksymacja dyfuzyjna jest z założenia metodą przybliżoną - zastępuje proces stochastyczny, którego opis analityczny stwarza trudności, łatwiejszym w opisie procesem dyfuzji. Błąd tej metody zależy istotnie od przyjętego wariantu aproksymacji, a w szczególności od wyboru warunków brzegowych równania dyfuzji.

W zbadanych w pracy przypadkach stanowisk G/G/1 i G/G/1/N uwzględnienie powrotów elementarnych w procesie dyfuzji zwiększa dokładność aproksymacji, zarówno przy opisie procesu N(t) - liczby klientów w stanowisku obsługi, jak przy aproksymacji U(t) - wirtualnego czasu czekania. Redukcja błędu jest najbardziej widoczna przy małym obciążeniu systemu, kiedy

nie można pominąć prawdopodobieństwa pustego stanowiska i gdy model dyfuzyjny z barierą odbijającą w x = 0 daje niezadawalające wyniki.

Rozpatrując aproksymację dyfuzyjną procesu N(t) uzyskano też aproksymację czasu trwania okresu czynnego stanowiska obsługi oraz aproksymację rozkładu maksymalnej długości kolejki w okresie czynnym o bardzo dobrej zgodności z wynikami rzeczywistymi.

Natomiast wprowadzenie aproksymacji wyższego rzędu (uwzględnienie trzeciego momentu rozkładu czasu obsługi) tylko w ograniczonym stopniu zwiększa dokładność aproksymacji.

Opisywane w pracy modele odnoszą się do pojedynczego stanowiska obsługi. Wykorzystanie ich w opisie sieci jest możliwe, lecz wymaga dalszej pracy. W przypadku sieci otwartej i stanów nieustalonych można np. wykorzystać metodę zaproponowaną w [Duda 84], obliczającą w kwantach czasu zmienne parametry strumieni wewnętrz sieci, lub wyciągnąć wnioski ze zbadanej w tej pracy dynamiki zmian kolejek i czasu czekania w stanowiskach G/G/1.

W przypadku sieci zamkniętej i stanów ustalonych model konserwatora z rozdz. V może służyć do opisu każdego stanowiska z osobna - resztę sieci uważa się wtedy za źródło zgłoszeń, podobnie jak ma to miejsce w metodach iteracyjnych [Chan 75, Shum 67, Mari 79b].

Przedstawiony w rozdz. V model jest nie tylko rozwiązaniem problemu konserwatora, lecz jest to model stanowiska z czasem obsługi i strumieniem wejściowym zależnym od liczby klientów w stanowisku - i jako taki może być wykorzystany w technikach redukujących rozmiary sieci, gdzie część sieci jest zastępowana jednym tego typu stanowiskiem (por. [Zaho 83]) oraz w modelach opisujących synchronizację klientów w sieci poprzez zmiany czasu obsługi w funkcji stanu kolejki np. [Wilk 86].

W pracy rozpatrywano powroty elementarne z krańców odcinka, na którym określony jest proces dyfuzji, do punktu wewnętrz tego odcinka. Można opracowaną metodę wykorzystać do opisu przeskoków wewnętrz przedziału dyfuzji, co odpowiadało by np. grupowym odejściom klientów z kolejki, w której następuje ich synchronizacja.

LITERATURA

Adam 77 C. Adams, E. Gelenbe, J. Vicard: An experimentally validated model of a paging drum, IRIA Rapport de Recherche no. 229, Rocquencourt 1977.

Agra 80 A.K. Agrawala, S.T. Tripathi: Transient solution of the virtual waiting time of a single server queue and its applications, Information Siences, vol. 21, 1980, pp. 141-158.

Alle 80 A.O. Allen: Queueing Models of Computer Systems, Computer, vol. 13, no. 3, 1980, pp. 13-24.

Arat 75 M. Arató: Diffusion approximation for multiprogrammed computer systems, in: Computation and Mathematics with Applications, vol. 1, no. 3/4, 1975, pp. 315-326.

Arno 51 W.E. Arnoldi: The Principle of Minimized Iterations in the Solutions of the Matrix Eigenvalue Problem, Quart. Appl.Math., vol. 19, 1951, pp. 17-29.

Asgh 78 K. Asgharzadeh, G.F. Newell: Optimal Dispatching Strategies for Vehicles Having Exponential Distributed Trip Times, Naval Research Logistics Quaterly, vol. 25, 1978, pp. 489-509.

Awen 82 О.И. Авен, Н.Н. Гурин, Я.А. Коган: Оценка качества и оптимизация вычислительных систем, Наука, Москва 1982.

Bade 75 M. Badel, A.V.Y. Shum: Accuracy of an approximate computer system model, IRIA Rapport de Recherche no. 130, Rocquencourt 1975.

Bail 54 N.T.J. Bailey: A continous time treatment of a simple queue using generating furictions, J.R. Statist. Soc. B16, 1954, pp. 288-291.

Baio 83 C. Baiocchi, A.C. Capelo, V. Comincioli, G. Serazzi: A mathematical model for transient analysis of computer systems, Performance Evaluation, vol. 3, 1983, pp. 247-264.

Balb 79 G. Balbo, S.C. Bruell, H.D. Schwetman: Customer Classes and Closed Network Models - a Solution Technique, Proc. IFIP Congress 77, North Holland, Amsterdam 1979, pp. 559-564.

Balb 82 G. Balbo: Approximate Solutions of Queueing Network Models of Computer Systems, Ph.D. Thesis, Purdue Univ., Was. Lafayette, 1979.

Bals 82 S. Balsamo, G. Iazeolla: An Extention of Norton's Theorem for Queueing Networks, IEEE Trans. on Softwsre Eng., vol. SE-8, no. 4, 1982, pp. 298-305.

Banh 78 Banh - Tri-An: Reseaux d'Ordinateurs à Commutation des Paquets, Ph. D. Thesis, Université de Liège, Liège 1978.

Bard 79 Y. Bard: Some Extentions to Multiclass Queueing Network Analysis, in: M. Arató et al. (Edts.), Performance of Computer Systems, North Holland, Amsterdam 1979.

Bask 75 F. Baskett, M. Chandy, R. Muntz, J. Palacios: Open, Closed and Mixed Networks of Queues with Different Classes of Customers, J.ACM, vol. 22, no. 2, 1975, pp. 248-260.

Bard 81 Y. Bard: A Simple Approach to System Modeling, Performance Evaluation, vol. 1, no. 3, 1981, pp. 225-248.

Basz 83 Г.П. Башарин, А.Л. Толмачев: Теория сетей массового обслуживания и её приложения к анализу информационно - вычислительных систем, теория вероятностей - математическая статистика теоретическая кибернетика, Т 20, Итоги науки и техн. ВИНИ ТИ АН СССР, Москва 1983, сс. 3-119.

Beth 68 I.A. Bather: A Diffusion Model for the Control a Dam, J. of Appl. Probab. vol. 5, 1968, pp. 55-71.

Bela 81 В. Беляков, Ю.И. Митрофанов: К исследованию замкнутых сетей массового обслуживания большой размерности, Автоматика и телемеханика, но. 7, 1981, сс. 61-69.

Bene 56 V.E. Beneš: On Queues with Poisson Arrivals, Annals of Mathematical Statistics, vol. 28, 1956, pp. 670-677.

Bene 63 V.E. Beneš: General Stochastic Processes in the Theory of Queues, Addison Wesley, Reading, Mass., 1963.

Bere 81 B. Bere: Influence du moment d'ordre 3 sur les files d'attente à lois de service générales, Thèse de docteur 3e cycle, Univ. de Rennes I, 1981.

BEST 78 BEST/1 User's Guide, BGS Systems, Inc., BE78-020-01, Lincoln, Ma., 1978.

Bhar 80 K. Bharath-Kumar: Discrete-time queueing systems and their networks, IEEE Trans. Communs., vol. 28, no. 2, 1980, pp. 260-263.

Bhar 60 A.T. Bharuha-Reid: Elements of the Theory of Markov Processes and Their Epplications, Mc Graw-Hill, New York 1960.

Blum 83 A. Blum, L. Donatiello, P. Heidelberger, S.S. Lavenberg, E.A. MacNair: Experiments with Decomposition of Extended Queueing Network Models, IBM Research Report RC10213, Yorktown Heights, NY 1983.

Boel 83 R. Boel: Boundary conditions for diffusion aproximations to queueing problems. Proc. Conf. Modeling and Performance Evaluation, Paris 1983.

Boro 80 А.А. Боровков Асимптотические методы в теории массового обслуживания, Наука, Москва 1980.

Booy 84 M. Booyens et al: SNAP: An Analytic Multiclass Queueing Network Analyser, in Preprints of In. Conf. on Modelling Techniques and Tools for Performance Analysis, Paris 1984.

Boye 79 C. Boyer de Bouillane: Sur l'approximation de modèles de files d'attente par des processus de diffusion associés à une condition frontière, thèse du docteur en 3ème cycle, l'Universite Pierre et Marie Curie, Paris 1979.

Bran 82 A. Brandwajn: Fast approximate solution of multiprogramming models, in: Proc. 1982 ACM SIGMETRICS Conf. Meas. Modeling Comput. Syst. 1982, pp. 141-149.

Bron 70 I.N. Bronsztejn, K.A. Siemiendiajew: Matematyka, poradnik encyklopedyczny, PWN, Warszawa 1970.

Brue 80 S.C. Bruell, G. Balbo: Computational Algorithms for Closed Queueing Networks, Elsevier-North Holland, New York 1980.

Brya 83 R.M. Bryant, A.E. Krzesiński, P. Teunissen: The MVA preempt-resume priority approximation, in: Perform. Eval. Rev., Special Issue, Proc. 1983 ACM SIGMETRICS Conf. Meas. Modeling Comput. Syst., 1983, pp. 12-27.

Buze 73 J.P. Buzen: Computational Algorithms for Closed Queueing Networks with Exponential Servers, Comm. ACM, vol. 16, no. 9, 1973, pp. 527-531.

Cach 81 F. Cachard: Logiciel numérique associé à une modélisation de systèmes informatiques, Thèse de doctorat, Univ. de Grenoble 1981.

Cham 56 D.C. Champernowne: An elementary method of solution of the queueing problem with a single server and constant parameters, J.R. Statist. Soc. B18, 1956, pp. 125-128.

Chan 43 S. Chandrasekhar: Stochastic Problems in Physics and Astronomy, Rev. Mod. Phys., vol. 15, 1943, pp. 57-59.

Chan 72 K.M. Chandy: The Analysis and Solutions for General Queueing Networks, Proc. 6th Annual Princeton Conf. on Inform. Sci. and Systems, Princeton 1972.

Chan 75 K.M. Chandy, U. Herzog, L.Woo: Parametric Analysis of Queueing Networks and Approximate Analysis of General Queueing Networks, IBM J. Res. Develop., vol. 10, no. 1, 1975, pp. 36-49.

Chan 77 K.M. Chandy, J.H. Howard, D.F. Towsley: Product Form and Local Balance in Queueing Networks, J. ACM, vol. 24, no. 2, 1977, pp. 250-263.

Chan 82a K.M. Chandy, C.H. Sauer: Computational Algorithms for Product Form Queueing Networks, Comm. ACM, vol. 23, no. 10, 1980, pp. 573-583.

Chan 32b K.M. Chandy, D. Neuse: Linearizer: A Heuristic Algorithm for Queueing Network Models of Computing Systems, Comm. ACM, vol. 25, no. 2, 1982, pp. 126-134.

Chan 77 S.S.L. Chang: Simulation of transient and time varying conditions in queueing networks, Proc. 7th Annual Pittsburgh Conference, Modelling and Simulation, vol. 8, Part. 2, Pittsbourgh, 1977.

Chin 33 A.Y. Chinczyn: On the average stopping time of machines (in Russian) Mat. Sbornik, vol. 40, 1933, pp. 119-123.

Chow 83 W.M. Chow: Approximations for Large Scale Closed Queueing Networks, Performance Evaluation, vol. 3, no. 1, 1983, pp.1-12.

Clar 53 A.B. Clarke: On Time-dependent Waiting Line Processes, Ann. Math. Statist., vol. 24, 1953, pp. 491-492.

Coff 83 E.G. Coffman, M.I. Reiman: Diffusion approximations for computer/communication systems, Proceedings of the international Workshop on Applied Mathematics and Performance Reliability Models of Computer Communication Systems, University of Pisa 1983.

Cohe 69 J.W. Cohen: The Single Server Queue, North-Holland, Amsterdam 1969.

Cono 58 B.W. Conolly: A Difference Equation Technique to the Simple Queue, J.Rm Statist. Soc. B20, 1958, pp. 165-175.

Conw 85 A.E. Conway, N.D. Georganas: RECAL - A New Efficient Algorithm for the Exact Analysis of Multiple - Chain Closed Queueing Networks, INRIA Rapport de Recherche no. 373, Rocquencourt 1985.

Cox 65 R.P. Cox, H.D. Miller: The Theory of Stochastic Processes, Chapman and Hall, London 1965.

Cour 71 P.J. Courtois, J. Georges: On a Single Server Finite Queueing Model with State-Dependent Arrival and Service Processes, Operating Research, vol. 19, 1971, pp. 424-429.

Cour 77 P.J. Courtois: Decomposability, Queueing and Computer Systems Applications, Academic Press, New York 1977.

Czac 77,79 T. Czachórski: Przybliżone rozwiązania dokładnych modeli i dokładne rozwiązania modeli przybliżonych w ocenie efektywności złożonych układów komputerowych; cz. I, Podstawy Sterowania t. 7, z. 1, 1977, ss. 11-24, cz. II: Model oparty o symulację, Podstawy Sterowania, t. 9, z. 2, 1979, ss. 125-138.

Czac 80 T. Czachórski: A Multiqueue Approximate Computer Performance Model with Priority Scheduling and System Overhead, Podstawy Sterowania, t. 10, z. 3, 1980, pp. 223-240.

Czac 81 T.Czachórski, M. Kowalówka: The GI/G/1 busy period - Investigation via diffusion model, w: Modelisation et Evaluation des Systèmes Informatiques, Collection Seminaires INRIA, Rocquencourt 1981.

Czac 82 T. Czachórski, M. Kowalówka, A. Wilk: A Software for Computer Systems Analysis - One More Effort, Proc. 7th Conf. on Operating Systems, Vységrad 1982.

Czac 83a T. Czachórski: Stany nieustalone stanowisk obsługi - opis przy pomocy aproksymacji dyfuzyjnej, Podstawy Sterowania, t. 13, z. 3, 1983, pp. 191-206.

Czac 83b T. Czachórski: Numeryczne i przybliżone metody obliczeniowe w statystycznych modelach systemów komputerowych, t. 13, z.4, 1983, pp. 293-314.

Czac 84 T. Czachórski: Zadanie konserwatora - opis przy pomocy aproksymacji dyfuzyjnej, Podstawy Sterowania, tom 14, z. 4, 1984, pp. 301-328.

Czac 85a T. Czachórski: Aproksymacja dyfuzyjna rozkładu czasu czekania w stanowisku M/G/1, Podstawy Sterowania. t. 15, z. 1-2, 1985, pp. 65-90.

Czac 85b T. Czachórski, M. Kowalówka, Z. Szczerbiński, A. Wilk: Analizator Modeli Kolejkowych AMOK - narzędzie modelowania pracy systemów komputerowych; Podręcznik użytkownika (w przygotowaniu).

Czac 85c T. Czachórski: A Third-Order Approximation of the Virtual Waiting Time of the M/G/1 Queue, Rapport de Recherche no. 43, Laboratoire ISEM, Universite de Paris-Sud, Orsay 1985.

Czac 86 T. Czachórski, A. Duda: A Diffusion Model of the Fork and Join Synchronization Primitives, Rapport de Recherche, Laboratoire ISEM, Université de Paris-Sud, Orsay 1986 (w druku).

Darl 53 D.A. Darling, A.J.F. Siegert: The First Passage Problem for a Continuous Markov Process, Ann. Math. Statist., vol. 24, 1953, pp. 823-841.

Denn 78 P.J. Denning, J.P. Buzen: The operational analysis of queueing network models, Comput. Surveys, vol. 10, 1978, pp. 225-261.

Ditk 64 W.A. Ditkin, A.P. Prudnikow: Przekształcenia całkowe i rachunek operatorowy. PWN, Warszawa 1964.

Dodd 81 S.L. Dodd, D.F. McAllister, W.J. Stewart: An Alternative Method for the Exact Solution of Coxian Queueing Networks, Proc. ACM SIG-METRICS Conf. on Measuring and Modeling of Computer Systems, Las Vegas 1981.

Drix 82 P. Drix: Algorithme de reseaux de files d'attente fermés BCMP et heuristiques associées, Thèse de docteur ingénieur, Univ. Paris VI, 1982.

Duda 83 A. Duda: Performance Analysis of the Checkpoint -Rollback-Recovery System via Diffusion Approximation, Proc. of International Workshop on Applied Mathematics and Performance Reliability Models of Computer Communication Systems, Pisa 1983.

Duda 84 A. Duda: Analyse et modélisation des régimes transitoires dans des systèmes informatiques, Thèse de docteur ingénieur, Université Paris Sud 1984.

Eage 83 D.L. Eager, K.C. Sevcik: Performance Bound Hierarchies for Queueing Networks, ACM Trans. Comput. Syst., vol. 1, 1983, pp. 99-115.

Eaga 84 D.L. Eager: Bounding Algorithms for Queueing Network Models of Computer Systems, Ph. D. Thesis, Dept. of Computer Sci. Univ. Toronto, Toronto 1984.

ElAf 83 M.A. El-Affendi, D.D. Kouvatsos: A Maximum Entropy Analysis of the M/G/1 and G/M/1 Queueing Systems at Equilibrium, Acta Informatica, fasc. 19, 1983, pp. 339-355.

Fan 76 W. Fan: Simulation of queueing network with time varying arrival rates, Mathematics and Computers in Simulation, vol. 18, 1976, pp. 165-170.

Fell 36 W. Feller: Zur Theorie der Stochastischen Prozesse, Math. Ann. 113, 1936, pp. 116-160.

Fell 51 W. Feller: Diffusion Processes in Genetics, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability 1951.

Fell 54 W. Feller: Diffusion process in one dimension, Trans. Amer. Math. Soc. vol. 77, 1954, pp. 1-31.

Fell 78 W. Feller: Wstęp do rachunku prawdopodobieństwa, t. 2, PWN, Warszawa 1978.

Ferr 83 H.D. Ferrari, G. Serazzi, A. Zeigner: Measurement and Tuning of Computer Systems, Prentice Hall. Englewood Cliffs. N.J. 1983.

Fili 83 J. Filipiak: Diffusion-equation model of slightly loaded M/M/1 queue, Operations Research Letters, vol. 2, 1983, pp. 134-139.

Fili 83a J. Filipiak: Dynamiczna teoria przepływów w sieciach, Zeszyty Naukowe Akademii Górniczo-Hutniczej - nr 898, Kraków 1983.

Fili 84 J. Filipiak, A.R. Pach: Selection of Coefficients for a Diffusion-Equation Model of Multi-Server Queue, PERFORMANCE'84 - Proc. of The 10th International Symposium on Comp. Perf., North Holland, Amsterdam 1984.

Fili 86 J. Filipiak: Diffusion Equation Model of Birth and-Death Congestion System, Podstawy Sterowania, t. 16, z. 1, 1986.

Fili J. Filipiak: Location of a Boundary for a Diffusion Equation of a Single Server Queue, praca zgłoszona do druku.

Fisc 77a M.J. Fischer: Analysis and Design of Loop Service System via a Diffusion Approximation, Operations Research, vol. 25, 1977, pp. 269-278.

Fisc 77b M.J. Fischer: An Approximation to Queueing Systems with Interruptions, Management Science, vol. 24, 1977, pp. 338-344.

Fosc 77 G.J. Foschini: On heavy traffic diffusion analysis and dynamic routing in packet switched networks, Computer Performance, K.M. Chandy and M. Reiser Eds., North Holland, 1977, pp. 499-513.

Fosc 78 G.J. Foschini, J. Salz: A basic dynamic routing problem and diffusion, IEEE Trans. Comm., vol. COM-26, 1978, pp. 320-327.

Gave 65 D.P. Gaver: Observing stochastic processes, and approximate transform inversion, Operations Research, vol. 13, 1965, pp. 444-459.

Gave 68 D.P. Gaver, Jr. Diffusion approximation and models for certain congestion problems, J. Appl. Prob. vol. 5, 1968, pp. 607-623.

Gave 71 D.P. Gaver: Analysis of Remote Terminal Backlogs under Heavy Demand Conditions, Journal of the ACM, vol. 18, 1971, pp.405-415.

Gave 73a D.P. Gaver, G.S. Shedler: Processor utilization in multiprogramming systems via diffusion approximations, Operations Research, vol. 21, 1973, pp. 569-576.

Gave 73b D.P. Gaver, G.S. Shedler: Approximate Models for Processor Utilization in Multiprogrammed Computer Systems, SIAM J. Comput., vol. 2, no. 3, 1973, pp. 183-192.

Gave 77 D.P. Gaver, J.P. Lehoczky: A Diffusion Approximation Solution for a Repairman Problem, with Two Types of Failure, Management Science, vol. 24, no. 1, 1977, pp. 71-81.

Gave 83 D.P. Gaver, P.A. Jacobs, G. Latouche: The Normal Approximation and Queue Control for Response Times in a Processor-Shared Computer System Model, Colloque International sur la Modélisation et les Outils d'Analyse de Performance, Paris 1983.

Gele 75 E. Gelenbe: On approximate computer system models, JACM, vol. 22, 1975, pp. 261-269.

Gele 76 E. Gelenbe, G. Pujolle: The behaviour of a single queue in a general queueing network, Acta Informatica, vol. 7, 1976, pp. 123-136.

Gele 77 E. Gelenbe, G. Pujolle: A diffusion model for multiple class queueing networks, in: Measuring, Modelling and Evaluating Computer Systems, North Hollend, Amsterdam 1977, pp. 189-199.

Gele 78 E. Gelenbe, A. Kurinckx: Random Injection Control of Multiprogramming in Virtual Memory, IEEE Transactions on Software Engineering, vol. SE-4, no. 1, 1978.

Gele 79 E. Gelenbe: Probabilistic models of computer systems, Part II, Acta Informatica, vol. 12, 1979, pp. 285-303.

Gele 80 E. Gelenbe, I. Mitrani: Analysis and synthesis of computer systems, Academic Press, London 1980.

Gele 82 E. Gelenbe, G. Pujolle: Introduction aux réseaux de filles d'attente, Eyrolles, Paris 1982.

Gned 62 B.V. Gnedenko: The Theory of Probability, Chelsea, New York 1962.

Gord 67 W.J. Gordon, G.F. Newell: Closed Queueing Systems with Exponential Servers, Operations Research, vol. 15, no. 2, 1967, pp. 252-267.

Gras 77 W.K. Grassmann: Transient solutions in markovian queueing systems, Comput. and Ops. Res. vol. 4, 1977, pp. 47-53.

Gras 81 W.K. Grassmann: Simulating transient solutions of markovian systems, Research Report 81-13, University of Saskatchewan 1981.

Gros 84 D. Gross, D.R. Miller: The Randomization Technique as a Modeling Tool and Solution Procedure for Transient Markov Processes, Operations Research, vol. 32, no. 2, 1984, pp. 343-361.

Hala 78 B. Halachmi, W.R. Franta: A Diffusion Approximation to the Multi-Server Queue, Management Sci. vol. 24, no. 5, 1978, pp. 522-529.

Harr 78 J.M. Harrison, A.J. Taylor: The Diffusion Approximation for Tandem Queues in Heavy Traffic, Advances in Applied Probability, vol. 10, 1978, pp. 886-905.

Harr 81a P.G. Harrison: Application of queueing network transient analysis, Research Report no. DOC 81/5, Department of Computing Imperial College of Science and Technology London, March 1981.

Harr 81b P.J. Harrison: Time-dependent properties of queueing networks, Performance 81, North Holland 1981.

Harr 81c P.M. Harrison, A.J. Lemoine: Sticky Brownien motion as the limit of storage processes, J. Appl. Prob., vol. 18, 1981, pp. 227-235.

Haso 63 A.M. Hasofer: On the integrability, continuity and differentiability of a family of functions introduced by L. Takács, Ann. Math. Statist., vol. 34, 1963, pp. 1045-1049.

Heid 83 P. Heidelberger, H.S. Trivedi: Analytic Queueing Models for Programs with Internal Concurrency, IEEE Trans. on Computers, vol. C-32, no. 1, 1983, pp. 73-83.

Heym 74 D.P.Heyman: An aproximation for the busy period of the M/G/1 queue using a diffusion model, J. Appl. Prob., vol. 11, 1974, pp. 159-169.

Heym 75 D.P. Heyman: A diffusion model approximation for the GI/G/1 queue in heavy traffic, The Bell System Technical Journal., vol. 54, 1975, pp. 1637-1646.

Heym 77 D.P. Heyman: Optimal Disposal Policies for a Single Item Inventory System with Returns, Naval Research Logistics Quaterly, vol. 24, 1977, pp. 385-405.

Hill 49 E. Hille: On the Integration Problem for Fokker-Planck's Equation in the Theory of Stochastic Processes, XI Congr. des Math. Scond. 1949, pp. 183-194.

Igle 65 D.L. Iglehart: Limiting Diffusion Approximations for the Many Server Queue and the Repairman Problem, JAP, vol. 2, 1965, pp. 429-441.

Igle 70 D. Iglehart, W. Whitt: Multiple Channel Queues in Heavy Traffic I-III, Advances in Applied Probability, vol. 2, 1970, pp. 150-177, 355-369.

Igle 73 D.L. Iglehart: Weak Convergence in Queueing Theory, Adv. Appl. Prob., vol. 5, 1973, pp. 570-594.

Igle 74 Iglehart, A.J. Lemoine: Approximations for the Repairman Problem with Two Repair Facilities, I: No Spares, II: Spares, Advances in Applied Probability, vol. 5, 1973, pp. 596-613, vol. 6, 1974, pp. 147-158.

Itô 65 K. Itô, H.P. McKean: Diffusion Processes and their Sample Path, Springer Verlag, Berlin, Heidelberg, New York 1965.

Izyd 81 J. Izydorczyk: Zadanie konserwatora - analiza wpływu zmian parametrów systemu obsługi na jego charakterystyki, Podstawy Sterowania, t. 11, z. 2, 1981, ss. 77-102.

Jack 57 J.R. Jackson: Networks of waiting lines, Oper. Res., vol. 5, no. 4, 1957, pp. 518-521.

Jack 63 J.R. Jackson: Jobshop-Like Queueing Systems, Management Sci., vol. 10, 1963, pp. 131-142.

Jaco 82 P.A. Jacobson, E.D. Lazowska: Analyzing Queueing Networks with Simultaneous Resource Possession, Comm. of the ACM, vol. 25, no. 2, 1982, pp. 142-151.

Jone 80 S.K. Jones, R.K. Cavin, D.A. Johnston: An Efficient Computational Procedure for the Evaluation of the M/M/1 Transient State Occupancy Probabilities, IEEE Trans. on Comm., vol. COM-28, no. 12, 1980, pp. 2019-2020.

Kame 82 H. Kameda: A Finite-Source Queue with Different Customers, J.ACM, vol. 29, no. 2, 1982, pp. 478-491.

Kami 80 В.Н. Каминский: Асимптотические свойства замкнутых сетевых моделей вычислительных систем, в сб.: Вычислительные системы и комплексы, Ленинград 1980, сс. 7-12.

Keil 64 J. Keilson: A review of transient behavior in regular diffusion and birth-death processes, Part I., J. Appl. Prob., vol. 1, 1964, pp. 247-266.

Keil 65 J. Keilson: A review of transient behavior in regular diffusion and birth-death processes. Part II., J. Appl. Prob. vol. 2, 1965, pp. 405-428.

Kend 53 D.G. Kendall: Stochastic Processes Occuring in the Theory of Queues and their Analysis by the Method of Imbedded Markov Chains, Annals of Math. Statist., vol. 24, 1953, pp. 338-354.

Khin 32 A.J. Khinchin: Mathematical Theory of Stationary Queues, Mat. Sbornik, vol. 39, 1932, pp. 73-84.

Kimu 79 T. Kimura: Deffusion Approximation for GI/G/1 Queueing Systems with Finite Cpacity: I - The First Overflow Time, II - The Stationary Behaviour, J. of Oper. Res. Soc. of Japan, vol. 22, no. 1, pp. 41-58, no. 4, 1979, pp. 301-319.

Kimu 80 T. Kimura: Studies on diffusion approximation for queueing systems, PhD thesis, Kyoto University, 1980.

Kimu 83 T. Kimura: Diffusion Approximation for an M/G/m Queue, Operations Research, vol. 31, no.2, 1983, pp. 304-321.

King 80 P.J.B. King, I. Mitrani: Numerical Methods for Infinite Markov Processes, Proc. Performance'80, Toronto 1980.

King 61 J.F.C. Kingman: The Single Server Queue in Heavy Traffic, Proc. Cambridge Phil. Soc., vol. 57, 1961, pp. 902-904.

King 62 J.F.C. Kingman: On Queues in Heavy Traffic, Journal of Royal Stat. Soc. Series B, vol. 24, 1962, pp. 383-392.

Klei 75 L. Kleinrock: Queueing Systems, vol. I: Theory, Wiley, New York 1975.

Klei 76 L. Kleinrock: Queueing Systems, vol. II: Computer Applications, Wiley, New York 1976.

Koba 74a H. Kobayashi: Application of the diffusion approximation to queueing networks I: Equilibrium queue distributions, JACM, vol. 21, 1974, pp. 316-328.

Koba 74b H. Kobayashi: Application of the diffusion approximation to queueing Networks II: Nonequilibrium distributions and applications to computer modelling, JACM, vol. 21, 1974, pp. 459-469.

Koba 77a H. Kobayashi, R. Konheim: Queueing models for computer communications system analysis, IEEE Trans. Comm. COM-25, no. 1, 1977.

Koba 77b H. Kobayashi, Y. Onozato, D. Yuynh: An Approximate Method for and Analysis of an ALOHA System, IEEE Trans. Commun., vol. COM-25, no. 1, 1977, pp. 158-168.

Koba 78 H. Kobayashi: Modeling and Analysis: An Introduction to system Performance Evaluation Methodology Addison Wesley, Reading, Mass. 1978.

Kolm 31 A.N. Kolmogorov: Über die analytischen Methoden in der Wahrscheinlichkeiterechnung, Math. Ann. vol. 104, 1931, pp. 415-458.

Kopo 73 I. Kopocińska: The Repairman Problem, Dissertationes Mathematicae, GI, PWN, Warszawa 1973.

Koti 78 T.C.T. Kotiah: Approximate transient analysis of some queueing systems, Operations Research, vol. 26, no. 2, 1978, pp. 334-346.

Kouv 84 D.D. Kouvatsos: Maximum entropy methods for general queueing networks, Proc. Int. Conf. Modelling Techniques and Tools for Performance Analysis, May 1984, Paris.

Krit 82 P.S. Kritzinger, S. van Wyk, A.E. Krzesiński: A Generalization of Norton's Theorem for Multiclass Queueing Networks, Performance Evaluation, vol. 2, 1982, pp. 98-107.

Kueh 79 P.J. Kuehn: Approximate Analysis of General Queueing Networks by Decomposition, IEEE Trans. Commun., vol. COM-27, no. 1, 1979, pp. 113-126.

Koga 79a А.Я. Коган: Вероятностные модели сетей ЭВМ использующих радиоканалы связи, в: Проблемы управления в сетях ЭВМ, PWN, Warszawa-Łódź 1979.

Koga 79b Ya.A. Kogan, V.G. Litvin: Piecewise diffusion approximations for queueing problems with heterogeneous arrivals and service, Problems of Control and Information Theory, vol. 8, no. 5/6, 1979, pp. 433-443.

Koga 80 Я.А. Коган, В.Ф. Кондаков: Приближенные методы анализа марковских моделей сетей ЭВМ, в кн.: Тезисы докладов VIII Всесоюзного совещания по проблемам управления, кн. 3, Таллин 1980.

Krug 80 В.К. Кругликов: Применение диффузионной аппроксимации для исследования стохастической сети на уровне двух моментов распределений, в сб.: Вычислительные системы и комплексы, Ленинград 1980, сс. 41-47.

Labe 80 J. Labetoulle, G. Pujolle: Isolation Method in a Network of Queues, IEEE Trans. Software Eng., vol. 6, no. 4, 1980, pp. 373-381.

Labe 81 J. Labetoulle, G. Pujolle, C. Soula: Stationary Distributions of Flows in Jackson Networks, Math. Oper. Res., vol. 6, no. 2, 1981, pp. 173-185.

Lam 81 S.S. Lam, Y.L. Lien: Tree Convolution Algorithm, Research Report TR165, Dept. of Comp. Sci, Univ. of Texas, Austin 1981.

Lam 82 S.S. Lam: Dynamic Scaling and Growth Behavior of Queueing Network Normalization Constants, JACM, vol. 29, no. 2, 1982, pp. 492-513.

Lave 80 S.S. Lavenberg, M. Reiser: Stationary State Probabilities at Arrival instants for Closed Queueing Networks with Multiple Types of Customers, J. Appl. Probab., vol. 17, no. 4, 1980, pp. 1048-1061.

Lave 81 S.S. Lavenberg: Closed Multichain Product Form Queueing Networks with Large Population Sizes, Proc. Meeting on Appled Probability-Computer Science, Atlanta Univ., Florida 1981.

Lave 83 S.S. Lavenberg (Ed): Computer Performance Modeling Handbook, Academic Press, New York 1983.

Lazo 79 E.D. Lazowska, R.L. Garner: Effective Solution of General Queueing Networks Via Decomposability, Technical Report no. 79-04-03, Dept. of Comp. Sci., Univ. of Washington 1979.

Lazo 84 E.D. Lazowska, J. Zahorian, G. Scott Graham, K.C. Sevcik: Quantitative System Analysis Using Queueing Network Models, Prentice Hall, Englewood Cliffs, N.J. 1984.

LeBo 84 J.Y.Le Boudec: Analyse quantitative de reseaux de files d'attente Markoviens, thèse de docteur en troisième cycle, Université de Rennes, Rennes 1984.

Lede 54 W. Lederman: G.E.H. Reuter, Spectral theory for the differential equations of simple brith and death processes, Phil. Trans. R. Soc. A9, vol. 246, 1954, pp. 321-369.

Lemo 78 A.J. Lemoine: Networks of Queues - A Survey of Weak Convergence Results, Mangement Science, vol. 24, no. 11, 1978, pp. 1175-1193.

Lest 83 L.N, Lester: Accuracy of Approximating Queueing Network Departure Processes with Independent Renewal Processes, Information Processing Letters, vol. 16, no. 1, 1983, pp. 43-48.

Levy 48 P. Levy: Processus stochastiques et mouvement brownien, Paris 1948.

Loul 73 R. Loulou: Multi-Channel Queues in Heavy Traffic, JAP, vol.10, 1973, pp. 769-777.

Magn 49 W. Magnus, F. Oberhettinger: Fuctnions of Mathematical-Physics, Chelsea Publishing Company, New York 1949.

Mand 64 П. Мандл: Об управлении необрывающимися диффузионными процессами, Теория вероятностей и ее применения, т. IX, вып. 4, 1964.

Mand 68 P. Mandl: Analytical Treatment of One-Dimensional Markov Processes, Academia, Prague - Springer Verlag, Berlin 1968.

Mari 79a R. Marie: Algorithmes de calcul des probabilités asymptotiques pour les files d'attent $\lambda(n)/C_k/1/N$ avec rebouclage et cas particuliers, IRISA Publication Interne no. 125, Rennes 1979.

Mari 79b R. Marie: An Approximate Analytical Method for General Queueing Networks, IEEE Trans. on Software Eng., vol. SE-5, no. 5, 1979, pp. 530-538.

Mari 82 R. Marie, P.M. Snyder, W.J. Stewart: Extentions and Computational Aspects of an Iterative Method, Proc. ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems, Seattle 1982.

Mari 83a R. Marie, W.J. Stewart: On Exact and Approximate Iterative Methods for General Queueing Networks, IRISA Publication Interne no. 204, Rennes 1983.

Mari 83b M. Marie, G. Rubino: Sur l'existence et l'unicité du reseau homogène à un reseau fermée de files d'attente à lois générales, IRISA Publication Interne no. 207, Rennes 1983.

Mela 79 B. Melamed: Characterizations of Poisson-Traffic Streams in Jackson Queueing Networks, Adv. Appl. Probab., vol. 11, no. 2, 1979, pp. 422-438.

Merl 78 D. Merle: Contribution a l'etude d'un analyseur de modèles à réseaux de files d'attente, algorithmes de calcul et applications, Thèse de docteur ingénieur, Institut National Polytechnique de Lorraine 1978.

Mitr 82 Ю.И. Митрофанов, В.Г. Беляков, В.Х. Курбангулов: Методы и программные средства аналитического моделирования сетевых систем, Предворит. публ. Науч. сов. по комплекс. пробл. Кибернет. АН СССР, Москва 1982.

Moor 75 S.C. Moore: Approximating the behavior of non - stationary single - server queue, Oper. Res., vol. 23, 1975, pp. 1011-1032.

Mors 58 P.M. Morse: Queues, Inventories and Maintenance, Wiley, New York 1958.

McKe 81 J. McKenna, D. Mitra, K.G. Ramakrishnan: A Class of Closed Markovian Queueing Networks: Integral Representations, Asymptotic Expansions and Generalizations, Bell Syst. Tech. J. vol. 60, 1981, pp. 599-641.

McKe 82 J. McKenna, D. Mitra: Integral Representations and Asymptotic Expansions for Closed Markovian Queueing Networks: Normal Usage, Bell Syst. Tech. J., vol. 61, 1982, pp. 661-683.

Nain 81 Ph. Nain: Files d'attente avec priorité, Thèse de docteur 3ème cycle, sujet I, Univereité de Paris Sud, Centre d'Orsay 1981.

Newe 65 G.F. Newell: Approximation Methods for Queues with Application to the Fixed Cycle Traffic Light, SIAM Review, vol. 7, 1965, pp. 436-451.

Newe 68 G.F. Newell: Queues with time - dependent rates, Part I: The transition through saturation; Part II: The maximum queue and return to equilibrium; Part III: A mild rush hour, J. Appl. Prob., vol. 5, 1968, pp. 436-451, pp. 579-590, pp. 591-606.

Newe 71 G.F. Newell: Applications of Queueing Theory, Chapman and Hall. London 1971.

Newe 79 G.F. Newell: Approximate Behavior of Tandem Queues, Springer Verlag, Berlin, Heidelberg, New York 1979.

Neot 79 A.S. Noetzel: A Generalized Queueing Discipline for Product Form Network Solutions, J.ACM, vol. 26, no. 4, 1979, pp. 779-793.

Neus 82 D. Neuse, K.M. Chandy: HAM: The Heuristic Aggregation Method for Solving General Closed Queueing Network Models of Computer Systems, Proc. ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems, Seattle 1982.

Odon 83 A.R. Odoni, E. Roth: An empirical investigation of the transient behavior of stationary queueing systems, Ops. Res., vol. 31, 1983, pp. 432-455.

Palm 47 C. Palm: Arbetskraftens fördelning vid betjäning av automatmaskiner, Industritidningen Norden, vol. 75, no. 8, pp. 75-80, no. 9, pp. 90-94, no. 10, pp. 119-123, 1947; English translation: The Assignement of Workers in Servicing Automatic Machines, Journal of Industrial Engineering, vol. IX, no. 1, 1958, pp. 18-42.

PAWS 83 PAWS: A User Guide; Information Research Associates, Austin, TX, 1983.

PERF 72-84 PERFORMANCE - Proceedings of International Symposiums on Computer Performance, Measurement and Modelling, wydawane co roku, North Holland, Amsterdam 1972-1984.

Perr 83 H.G. Perros: On the $M/C_k/1$ Queue, Performance Evaluation, vol. 3, no. 2, 1983, pp. 83-93.

Pitt 79 Pittel B.: Closed Exponential Networks of Queues with Saturation, the Jackson type Stationary Distribution and Its Asymptotic Analysis, Math. Oper. Res., vol. 4, no. 4, 1979, pp. 357-378.

Poll 30 F. Pollaczek: Über eine Aufgabe der Wahrscheinlichkeitstheorie, I-II Math. Zeitschrift, vol. 32, 1930, pp. 64-100, 729-750.

Prab 65 N.U. Prabhu: Queues and Inventories, Wiley, New York 1965.

Pric 76 T.G. Price: A Note on the Effect of the Central Processor Service Time Distribution, J. of ACM, vol. 23, no. 2, 1976, pp. 342-346.

Pujo 78a G. Pujolle: Files d'attente avec dependance de l'etat et application à un système de communication multilignes, Congres SMF/AFCET Palaiseau 1978.

Pujo 78b G. Pujolle: Modélisation et évaluation de performances des réseaux à commutation de paquets, Thèse docteur es-sciences, Universite Paris Sud, Orsay 1978.

Rama 82 K.G. Ramakrishnan, D. Mitra: An overview of PANACEA,A software package for analysing Markovian queueing networks. Bell Syst. Tech. J., vol. 61, 1982, pp. 2849-2872.

Reim 83 M. Reiman: Some diffusion approximations with state space collapse, Proc. Int. Seminar Modelling and Performance Evaluation Methodology, vol. 1, Paris 1983, pp. 167-202.

Rein 82 A. Reinhard, K.J. Bueher: Characteristics of Cox Distribution, Performance Evaluation Revue, vol. 12, no. 1, winter 1982-83.

Reis 74 M. Reiser, H. Kobayashi: Accuracy of the diffusion approximation for some queueing systems. IBM Journal of Res. and Develop., vol. 18, 1974, pp. 110-124.

Reis 75 M. Reiser, H. Kobayashi: Queueing Network Models with Multiple Closed Chains: Theory and Computational Algorithms, IBM J.Res. Develop., vol. 19, 1975, pp. 283-294.

Reis 78 M. Reiser, C.H. Sauer: Queueing Network Models: Methods of Solution and their Program Implementation, in: K.M. Chandy, R.T.Yeh(Edts) Current Trends in Programming Methodology, vol.3, Prentice Hall, Englewood Cliffs 1978.

Reis 79a M. Reiser: A Queueing Network Analysis of Computer Communication Network with Window Flow Control, IEEE Trans. on Comm., vol. COM 27, no. 8, 1979, pp. 1199-1209.

Reis 79b M. Reiser: Mean Value Analysis of Queueing Networks - A New Look at an Old Problem, in: M. Arató et al. (Edts), Performance of Computer Systems, North Holland, Amsterdam 1979.

Reis 80a M. Reiser, S.S. Lavenberg: Mean Value Analysis of Closed Multichain Queueing Networks, J. ACM, vol. 27, no. 2, 1980, pp. 313-322.

Reis 80b M. Reiser, S.S. Lavenberg: Corrigendum to [Reis 80a], J. ACM, vol. 28, no. 3, 1980, p. 629.

Reis 81 M. Reiser: Mean-Value Analysis and Convolution Method for Queue-Dependent Servers in Closed Queueing Networks, Performance Evaluation, vol. 1, no. 1, 1981, pp. 7-18.

Remb 85 J. Rembowska: Modelowanie działania mechanizmu okienkowego sterowania w sieci komputerów, Archiwum Automatyki i Telemechaniki, Tom XXX, z. 2, Warszawa 1985.

Ride 76 K.L. Rider: A Simple Approximation to the Average Queue Size in the Time-Dependent M/M/1 Queue, J. ACM, vol. 23, no. 2, 1976, pp. 361-367.

Roth 83 E. Roth: A heuristic technique for the transient behavior of markovian queueing systems, Research Report GTE Laboratories Inc., November, 1983.

Rubi 84 G. Rubino: Contribution a l'étude de méthodes approchées pour reseaux de files d'attente, Thèse de doctorat en troisième cycle, Université de Rennes, Rennes 1984.

Saat 61 T.L. Saaty: Elements of Queueing Theory with Applications Mcraw-Hill, New York 1961.

Same 82 Samelson C.L., W.G. Bulgren: A Note on Product-Form Solution for Queueing Networks with Poisson Arrivals and General Service-Time Distributions with Finite Means, J. ACM, vol. 29, no. 3, 1982, pp. 830-840.

Saue 75 Ch.H. Sauer: Configuration of Computing Systems; An Approach Using Queueing Network Models, Ph. D. Thesis, Univ. of Texas, Austen 1975.

Saue 80 Ch.H. Sauer, K.M. Chandy: Approximate Solutions of Queueing Models, Computer, vol. 13, no. 4, 1980, pp. 25-31.

Saue 81 Ch.H. Sauer, K.M. Chandy: Computer Systems Performance Modeling, Prentice Hall, Englewood Cliffs, N.J. 1981.

Saue 82 Ch.H. Sauer, E.A. MacNair, J.F. Kurose: The Research Queueing Package Version 2: Introduction and Examples, IBM Research Report RA-138, Yorktown Heights, New York 1982.

Saue 83 Ch.H. Sauer, E.A. MacNair: Computer Communication Systems, Prentice Hall, Englewood Cliffs, N.J. 1983.

Schw 79 P. Schweitzer: Approximate Analysis of Multiclass Closed Networks of Queues, Proc. Int. Conf. on Stochastic Control and Optimization, Amsterdam 1979.

Sevc 77 K.C. Sevcik, A.I. Levy, S.K. Tripathi, J.L. Zahorjan: Improving Approximations of Aggregated Queueing Network Subsystems, in: K.M. Chandy, M. Reiser (Edts), Computer Performance, North Holland, Amsterdam 1977.

Sevc 79 K.C. Sevcik, I. Mitrani: The Distribution of Queueing Network States at Input and Output Instants, in: M. Arató et al.(Edts), Performance of Computer Systems, North Holland, Amsterdam 1979.

Sher 67 A.L. Sherr: An Analysis of Time-Shared Computer System, Ph.D. Thesis, Project MAC, MIT Press, Cambridge Mass. 1967.

Shum 76 A.W. Shum: Queueing Models for Computer Systems with General Service Time Distribution, Ph.D. Thesis, Harvard Univ., Cambridge, Mass. 1976.

SIGM 73-84 -: Proceedings of Annual ACM SIGMETRICS Conferences on Measurement and Modeling of Computer Systems, wydawane corocznie przez Performance Evaluation Revue - A Publication of the ACM Special Interest Group on Measurement and Evaluation.

Smol 13 M. von Smoluchowski: Einige Beispiele Brownscher Molekular bewegung unter Einfluss äusserer Kräfte, Bull de Acad. Sc. de Cracovie A, 1913, pp. 418-434.

Smol 15 M. von Smoluchowski: Über Brownsche Molekularbewegung unter Einwirkung äusserer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffusionsgleichung, Ann. der Physik, 1915, pp. 1103-1112.

Steh 79 H. Stehfest: Algorithm 368: Numeric Inversion of Laplace Transform., Comm. ACM, vol. 13, no. 1, 1979, pp. 47-49.

Ster 79 T.E. Tern: Approximations of Queue Dynamics and Their Application to Adaptive Routing in Computer Communication Networks IEEE Trans. Commun. vol. COM-27, no. 9, 1919, pp. 1331-1335.

Stew 77 W.J. Stewart: A New Approach to the Numerical Analysis of Markovian Models, in: K.M. Chandy, M. Reiser (Edts), Computer Performance, North Holland, Amsterdam 1977.

Stew 78 W.J. Stewart: A Comparison of Numerical Techniques in Markov Modelling, Comm. ACM, vol. 21, no. 2, 1978, pp. 144-152.

Stew 79 W.J. Stewart, R. Marie: A numerical solution for the $\lambda(n)/K/r$ queue, Rapport de IRISA no. 79, Rennes 1979.

Stew 80 W.J. Stewart, G.A. Zeiszler: On the Existence of Composite Flow Equivalent Markovian Servers, Proc. Performance 80, 7th IFIP W.G.7.3 International Symposium, Toronto 1980.

Stew 83 G.W. Stewart: Computable error bounds for aggregated Markov Chains, J.ACM, vol. 30, 1983, pp. 271-285.

Stew 84 W.J. Stewart, W.P. Stochs: Some Equivalence Results for Load-Independent Exponential Queueing Networks, IEEE Trans. on Software Engineering, vol. SE-10, no. 4, 1984, pp. 414-422.

Suna 78 T. Sunaga, E. Kondo, S.K. Biswas: An Approximation Method Using Continuous Models for Queueing Problems, J. Op. Res. Soc. Jap., vol. 21, 1978, pp. 30-44.

Swee 70 A.L. Sweet, J.C. Hardin: Solutions for Some Diffusion Processes with Two Barriers, JAP, vol. 7, 1970, pp. 423-431.

Szcz 82 H. Szczerbicka-Lipińska: Pewne rozszerzenia stosowalności metody średnich do rozwięzywania kolejkowych modeli systemów cyfrowych, Rozprawa doktorska, Politechnika Warszawska, 1982.

Taka 55 L. Takács: Investigation of Waiting Time Problems by Reduction to Markov Processes, Acta Math. Acad. Sci. Hung, vol. 6, 1955, pp. 101-129.

Taka 62a L. Takács: A Single Server Queue with Poisson Input, Operations Research, vol. 10, 1962, pp. 388-397.

Taka 62b L. Takács: Introduction to the Theory of Queues, Oxford University Press, New York 1962.

Taka 84 H. Takagi, L. Kleinrock: Diffusion Process Aproximation for the Queueing Delay in Contention Packet Broadcasting Systems; in: H. Rudin, W. Bux (editors): Performance of Computer - Communication Systems, Symposium IFIP WG 7.3/TC6, North Holland, Amsterdam 1984.

Thom 81 A. Thomasion, B. Nadji: Aggregation of Stations in Queueing Network Models of Multiprogrammed Computers, Proc. ACM SIGMETRICS Conf. on Measuring and Modeling of Computer Systems, Las Vegas 1981.

Thom 83 A. Thomasion , Bay P.: Analysis Techniques for queueing network models of multicomputer systems with shared ressources, Computer Performance, vol. 4, no. 3, 1983, pp. 151-166.

Tich 77 Б.У. Тихонов, М.А. Миронов: Марковские процессы, Сов. Радио, Москва 1977.

Trip 79 S.K. Tripathi: On Approximate Solution Techniques for Queueing Network Models of Computer Systems, Ph. D. Thesis, Univ. of Toronto 1979.

Trip 82 S.K. Tripathi, A.K. Agrawala: On characterizing the inter - departure process of a server, Performance Evaluation. vol. 2, 1982, pp. 251-255.

Trip 84 S.K. Tripathi: On Detecting Parallelism in software, ISEM Research Raport no. 31, Universite Paris-Sud, Orsay 1984.

Triv 80 K.S. Trivedi, R.E. Kinicki: A Model for Computer Configuration Design, Computer, vol. 13, no. 4, 1980, pp. 47-54.

Triv 82 K.S. Trivedi: Probability and Statistics with Reliability, Queueing and Computer Science Applications, Prentice Hall, Englewood Cliffs, NJ 1982.

Tucc 82 S. Tucci, E.A. MacNair: Implementation of Mean value Analysis for Open Closed and Mixed Networks, IBM Research Report RC 9392, Thomas Watson Res. Center, Yorktown Heights 1982.

Twee 73 R.L. Tweedie: The Calculation of Limit Probabilities for Denumerable Markov Processes from Infinitesimal Properties, J. Appl. Probab., vol. 10, 1973, pp. 84-99.

Upto 82 R.A. Upton, S.K. Tripathi: An approximate transient analysis of the M(t)/M/1 queue, Perform. Eval. vol. 2, 1982, pp. 118-132.

Upto 83 R.A. Upton, S.K. Tripathi: An analysis of integrated voice and data buffer/channel sharing schemes in a packet communication network, Research Report, System Design and Analysis Group, Department of Computer Science, University of Maryland, 1983.

Upto 84 R.A. Upton: A transient analysis of computer communication networks, PhD thesis, University of Maryland, juin 1984.

Vant 78 H.T. Vantilborgh: Exact Aggregation in Exponential Queueing Networks, J.ACM, vol. 25, no. 4, 1978, pp. 620-629.

Vant 60 H.T. Vantilborgh, R.L. Garner, E.D. Lazowska: Near-complete Decomposability of Queueing Networks with Clusters of Strongly Interacting Servers, Proc. Performance 80, 7th IFIP W.G.7.3 International Symposium, Toronto 1980.

Vera 84 M. Veran, D. Potier: QNAP2: Un Environnement Portable pour la Modélisation par Réseaux de Files d'Attente, Colloque International Sur la Modélisation et les Outils d'Analyse de Performance, Paris 1984.

Wien 23 N. Wiener: Differential Space, J. of Math. and Phys., vol. 2, 1923, pp. 131-174.

Whit 73 W. Whitt: A Diffusion Model for a Queue with Removable Server, Technical Report, Dept of Administrative Sciences, Yale University 1973.

Whit 74 W. Whitt: Heavy Traffic Limit Theorems for Queues: A Survey, in Proc. of a Conference on Mathematical Methods in Queueing Theory, Western Michigan Univ., 1973, Springer Verlag. Berlin 1974.

Whit 82a W. Whitt: On the heavy - traffic limit theorem for GI/G/ queues, Adv. Appl. Prob. vol. 14, 1982, pp. 171-190.

Whit 82b W. Whitt: Refining diffusion approximations for queues, Operations Research Letters, vol. 1, 1982, pp. 165-169.

Whit 83a W. Whitt: Approximations for networks of queues, Proceedings of ITC, 1983, pp. 234-240.

Whit 83b W. Whitt: The Queueing Network Analyzer, The Bell System Technical Journal, vol. 62, 1983, pp. 2779-2815.

Whit 83c W. Whitt: Performance of the Queueing Network Analyzer, The Bell System Technical Journal, vol. 62, 1983, pp. 2817-2843.

Woli 82 A. Wolisz: Wybrane niekonwencjonalne systemy masowej obsługi, Zeszyty Naukowe Politechniki Śląskiej, Gliwice 1982.

Yosh 57 K. Yoshida: Semi, group Theory and the Integration Problem of Diffusion Equations, Proc. Intern. Congr. of Mathematicians, Amsterdam 1954, vol. 1, 1957, pp. 405-420.

Zaho 79 J. Zahorjan: An Exact Solution Method for the General Class of Closed Separable Queueing Networks, Performance Evaluation Revue, vol. 8, no. 3, 1979, pp. 107-112.

Zaho 80 J. Zahorjan: The Approximate Solution of Large Queueing Network Models, Ph. D. Thesis, Univ. of Toronto 1980.

Wilk 86 A. Wilk: Geetion des resources par semaphores: Modélisation des mécanismes de contrôle de flux du protocole X25, MASI Rapport de Recherche no. 124, Paris 1986.

Zaho 81 J. Zahorjan, E. Wong: The Solution of Separable Queueing Network Models Using Mean Value Analysis, Pros. ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems, Las Vegas 1981.

Zaho 83 J. Zahorjan, E.D. Lazowska, R.L. Garner: A Decomposition Approach to Modeling High Service Time Variability, Performance Evaluation vol. 3, no. 1, 1983, pp. 35-54.

ZSAK 81-85 -: ZSAK PAN, Zespół Statystycznych Modeli Systemów Komputerowych, opracowania wewnętrzne 1981-1985.

PROCES POWROTÓW ELEMENTARNYCH W DOSKONALENIU APROKSYMACJI DYFUZYJNEJ JAKO METODY PROBABILISTYCZNEGO MODELOWANIA SYSTEMÓW KOMPUTEROWYCH

S t r e s z c z e n i e

Praca udoskonala aproksymację dyfuzyjną - metodę opisu działania stanowisk obsługi, często wykorzystywaną do modelowania pracy systemów komputerowych.

Sieć stanowisk obsługi, między którymi krążą klienci,to model stosowany już od około dwudziestu lat do analizy działania i oceny wydajności pracy systemów komputerowych. Stanowiska obsługi reprezentują elementy systemu komputerowego - procesor, pamięci zewnętrzne, terminale, linie transmisyjne - a klienci to wykonywane przez system zadania. Część początkowa niniejszej pracy zawiera monograficzny przegląd istniejących metod rozwiązywania takich modeli oraz dyskusję analitycznych i numerycznych trudności z nimi związanych. Trudności te powodują wykorzystanie metod przybliżonych, pozwalających na przyjęcie w modelu ogólniejszych założeń i redukujących nakład obliczeń do rozmiarów nie przekraczających możliwości maszyny cyfrowej.

Jedną z przybliżonych metod jest aproksymacja dyfuzyjna, zastępująca wartość procesu N(t) - liczbę klientów w stanowisku obsługi w chwili t - lub wartość procesu U(t) - wirtualny czas czekania klienta przybyłego do stanowiska w chwili t - wartością procesu dyfuzji. Dokładność aproksymacji zależy od wyboru warunków brzegowych dla równania dyfuzji. Proces dyfuzji musi być np. ograniczony do nieujemnej półosi liczb rzeczywistych - liczba klientów i czas czekania nie są nigdy ujemne. Wykorzystywana wcześniej w tym celu bariera w postaci ekranu odbijającego w punkcie z = = 0 jest dobrym przybliżeniem tylko w przypadku systemów mocno obciążonych, w których prawdopodobieństwo pustego stanowiska jest pomijalnie małe. Bliższy rzeczywistości jest zaproponowany w [Gele 75] warunek następujący: proces dyfuzji po dojściu do punktu z = 0 pozostaje w nim przez czas odpowiadający okresowi bezczynnemu stanowiska, poczym dyfuzja rozpoczyna się w punkcie z = 1 w przypadku opisu procesu N(t) lub w punkcie określonym przez rozkład czasu obsługi w przypadku opisu procesu U(t). Ciągły proces dyfuzji jest więc uzupełniony przeskokami, zwanymi powrotami elementarnymi.

W pracy przedstawiono nową metodę rozwiązywania układu równań charakteryzujących proces dyfuzji z procesem powrotów elementarnych. Uzyskane

rozwiązania pozwalają na opis stanów ustalonych i nieustalonych stanowiska obsługi. Przeanalizowano szczegółowo kilka podstawowych w teorii obsługi masowej problemów: aproksymację procesów N(t) i U(t) w stanowisku G/G/1, tj. w stanowisku z dowolnym wejściowym strumieniem klientów i dowolnym rozkładem czasu obsługi, aproksymację N(t) w stanowisku G/G/1/N, tj. z ograniczoną kolejką oraz w stanowisku G/G/1//N, tj. z ograniczoną populacją klientów. Zbadano zasadność zastąpienia procesu dyfuzji procesem Markowa wyższego rzędu. Przeprowadzono obliczenia numeryczne, które wskazują, że otrzymane rozwiązania są obciążone mniejszym błędem niż rozwiązania dotychczasowe, zwłaszcza w przypadku słabo obciążonych stanowisk obsługi. Dotyczy to zarówno stanów ustalonych jak i dynamiki zmian kolejki w stanach nieustalonych.

СКАЧКООБРАЗНЫЕ ГРАНИЧНЫЕ УСЛОВИЯ ДЛЯ УЛУЧШЕНИЯ ДИФФУЗИОННОЙ АППРОКСИМАЦИИ КАК МЕТОД МОДЕЛИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Р е з ю м е

В работе улучшено диффузионную аппроксимицию - метод описания действия обслуживающих устройств, который часто употребляется для моделирования вычислительных систем.

Сеть массового обслуживания, в которой циркулируют заявки переходящие из одной системы обслуживания в другую, это модель использована уже двадцать лет для анализа и оценки качества вычислительных систем. Обслуживающие устройства представляют собой ресурсы вычислительной системы - процессор, внешние запоминающие устройства, терминалы, линии передачи; заявки - это задания исполняемые системой. Начальная часть статьи включает монографический обзор существующих способов решения таких моделей, а также обсуждение аналитических и численных трудностей сопутствующих им. Эти проблемы являются причиной применения приближенных решений, в которых возможны более общие предположения и которые требуют более простых расчетов, возможных вычислительной системой. Одним из этих приближенных способов является диффузионная аппроксимация, которая заменяет случайный процесс N(t), представляющий собой длину очереди в момент времени t, или случайный процесс U(t), представляющий собой время ожидания заявки прибывшей в систему в момент t, марковском процессом диффузионного типа. Точность аппроксимации зависит от выбора граничных условий.

Диффузионный процесс надо на пример ограничить к неотрицательной части оси x - длина очереди и время ожидания всегда либо нулевые либо положительные. Простейшим граничным условием которое исполняет это требование является отражение в нуле. Оно удовлетворяет только в случае, когда обслуживающая система сильно загружена, т.е. когда вероятность пустой системы слишком мала. Ближе действительности другое граничное условие, предложенное в [Gelenbe 75], удерживающие траектории процесса на неотрицательной полуоси: после достижения нуля процесс задерживается здесь через время представляющие бездеятельный период системы, а потом совершает скачок в точку x = 1, если аппроксимация касается процесса N(t), или в точеу определенную распределением времени обслуживания, если аппроксимация касается процесса U(t). Таким образом беспрерывный стохастический процесс диффузии дополнен скачками - элементарными возвротами.

В статье представлен новый метод решения дифференциальных уравнений диффузии со скачкообразными граничными условиями. Результаты относятся к

переходному и стационарному режиму работы обслуживающего прибора. Анализируется подробно некоторые основные проблемы теории массового обслуживания: аппрокимацию процессов N(t) и U(t) в системе G/G/1, т.е. в системе с произвольным входящим потоком заявок и с произвольной длиностью обслуживания, процесс N(t) в системе G/G/1/N, т.е. в системе с ограниченной очередью, и в системе G/G/1//N, т.е. в системе с ограниченным числом заявок. Исследованы достоинства аппроксимации марковским процессом третьего порядка вместо процесса диффузии, т.е. процесса второго порядка. Даны численные примеры, которые показывают, что диффузионная аппроксимация со скачкообразными граничными условиями более точная чем существующие до сих пор результаты, особенно для систем с небольшой загрузкой, как в стационарном режиме работы, так и для динамики смен длины очереди в переходных периодах.

ELEMENTARY RETURN PROCESS IN REFINEMENT OF DIFFUSION APPROXIMATION VIEWED AS A METHOD OF PROBABILISTIC MODELLING OF COMPUTER SYSTEMS

S u m m a r y

The goal of this paper is to refine the diffusion approximation, one of the queueing theory methods used frequently to model computer systems and computer networks.

For the last two decades the queueing network models have been applied to analyse and evaluate computer systems performance. The servers within a queueing network represent the elements of a modelled system: processors, terminals, transmission lines, external memories; the customers served in them represent tasks performed by the system. The opening part of the paper revues existing methods of treating such models and discusses analytical and numerical problems they pose. Those problems cause the use of approximate methods which allow one to analyse larger and more general networks reducing the calculation effort to the amount bearable by a computer. The diffusion approximation is among these approximate methods. It replaces either the stochastic process N(t), the number of customers in the queue at the time t, or the process U(t), the virtual waiting time of a customer entering the queue at time t, with the diffusion process. The quality of this procedure depends on the choice of boundary conditions for the diffusion equation. The diffusion process should for example be limited to the nonnegative half of the axis x, as neither a queue nor a waiting time could be negative. The reflecting barrier put at the point z = 0, applied earlier, gives satisfactory results only in cases of heavy loaded servers where the probability of an empty system is negligible. The following condition, proposed in [Gele 75] is closer to reality: when the diffusion process attains zero, it stays there for a time interval which corresponds to the servers' idle period and then immediately recommences at z = 1 if N(t) is approximated or at a point x chosen with the service time probability distribution in the case of U(t). Hence, the continous diffusion process is supplemented with jumps called elementary returns.

The paper develops a proper mathematical method to solve diffusion equations with elementary return processes. Solutions which are obtained enable one to analyse the equilibrium and transient states of a server. Several typical problems are analysed in detail: approximation of N(t) and U(t) processes in the G/G/1 server, i.e. in the server with general

input stream and service time distribution, approximation of N(t) in the G/G/1/N server, i.e. in the server with a limited queue, approximation of N(t) in the G/G/1//N server, i.e. in the server with a finite customers' pupulation. A third-order Markovian approximation instead of the second-order diffusion approximation is investigated. The numerical results prove that the solutions which incorporate the elementary return process give much more exact results, especially when lightly loaded systems are considered. It refers both to equilbrium solutions and to the dynamics of queue changes during transient periods.

BIBLIOTEKA GŁÓWNA
Politechniki Śląskiej
P 41 93 /87/9

WYDAWNICTWA NAUKOWE I DYDAKTYCZNE POLITECHNIKI SLĄSKIEJ MOŻNA NABYĆ W NASTĘPUJĄCYCH PLACÓWKACH:

44-100 Gliwice — Księgarnia nr 096, ul. Konstytucji 14 b
44-100 Gliwice — Spółdzielnia Studencka, ul. Wrocławska 4 a
40-950 Katowice — Księgarnia nr 015, ul. Żwirki i Wigury 33
40-096 Katowice — Księgarnia nr 005, ul. 3 Maja 12
41-900 Bytom — Księgarnia nr 048, Pl. Kościuszki 10
41-500 Chorzów — Księgarnia nr 063, ul. Wolności 22
41-300 Dąbrowa Górnicza — Księgarnia nr 081, ul. ZBoWiD-u 2
47-400 Racibórz — Księgarnia nr 148, ul. Odrzańska 1
44-200 Rybnik — Księgarnia nr 162, Rynek 1
41-200 Sosnowiec — Księgarnia nr 181, ul. Zwycięstwa 7
41-800 Zabrze — Księgarnia nr 230, ul. Wolności 288
00-901 Warszawa — Ośrodek Rozpowszechniania Wydawnictw Naukowych PAN — Pałac Kultury i Nauki

Wszystkie wydawnictwa naukowe i dydaktyczne zamawiać można poprzez Składnicę Księgarską w Warszawie, ul. Mazowiecka 9.