

STUDIA INFORMATICA

Formerly: *Zeszyty Naukowe Politechniki Śląskiej, seria INFORMATYKA*
Quarterly

Volume 28, Number 3B (73)

Andrzej CHYDZIŃSKI

CHARAKTERYSTYKI KOLEJKOWANIA
MARKOWSKICH MODELI RUCHU
W SIECIACH PAKIETOWYCH



Silesian University of Technology Press
Gliwice 2007

Editor in Chief

Dr. Marcin SKOWRONEK
Silesian University of Technology
Gliwice, Poland

Editorial Board

Dr. Mauro CISLAGHI
Project Automation
Monza, Italy

Prof. Bernard COURTOIS
Lab. TIMA
Grenoble, France

Prof. Tadeusz CZACHÓRSKI
Silesian University of Technology
Gliwice, Poland

Prof. Jean-Michel FOURNEAU
Université de Versailles - St. Quentin
Versailles, France

Prof. Jurij KOROSTIL
IPME NAN Ukraina
Kiev, Ukraine

Dr. George P. KOWALCZYK
Networks Integrators Associates, President
Parkland, USA

Prof. Peter NEUMANN
Otto-von-Guericke Universität
Barleben, Germany

Prof. Olgierd A. PALUSINSKI
University of Arizona
Tucson, USA

Prof. Svetlana V. PROKOPCHINA
Scientific Research Institute BITIS
Sankt-Petersburg, Russia

Prof. Karl REISS
Universität Karlsruhe
Karlsruhe, Germany

Prof. Jean-Marc TOULOTTE
Université des Sciences et Technologies de Lille
Villeneuve d'Ascq, France

Prof. Sarma B. K. VRUDHULA
University of Arizona
Tucson, USA

Prof. Hamid VAKILZADIAN
University of Nebraska-Lincoln
Lincoln, USA

Prof. Adam WOLISZ
Technical University of Berlin
Berlin, Germany

Dr. Lech ZNAMIROWSKI
Silesian University of Technology
Gliwice, Poland

STUDIA INFORMATICA is indexed in INSPEC/IEE (London, United Kingdom)

© Copyright by Silesian University of Technology Press, Gliwice 2007
PL ISSN 0208-7286, QUARTERLY
Printed in Poland

ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ**OPINIODAWCY:**

Dr hab. inż. Jan Magott, prof. Politechniki Wrocławskiej
Prof. dr hab. inż. Maciej Stasiak

KOLEGIUM REDAKCYJNE

REDAKTOR NACZELNY – Prof. dr hab. inż. Andrzej Buchacz
REDAKTOR DZIAŁU – Dr inż. Marcin Skowronek
SEKRETARZ REDAKCJI – Mgr Elżbieta Leško

Majce

SPIS TREŚCI

WSTĘP	9
1. METODA POTENCJAŁU	14
1.1. Przypadek wielowymiarowy	14
1.2. Przypadek jednowymiarowy	17
1.3. Nota bibliograficzna do Rozdziału 1	18
2. KOLEJKOWANIE PROSTYCH I ZŁOŻONYCH STRUMIENI POISSONA	20
2.1. Model systemu kolejkowego	20
2.2. Kolejowanie procesu Poissona	21
2.2.1. Długość kolejki	23
2.2.2. Współczynnik strat	28
2.2.3. Opóźnienie	28
2.2.4. Okres przepełnienia bufora	31
2.2.5. Liczba strat w okresie przepełnienia bufora	38
2.3. Kolejowanie złożonego procesu Poissona	43
2.3.1. Okres przepełnienia bufora	46
2.3.2. Czas do przepełnienia bufora	53
2.4. Nota bibliograficzna do Rozdziału 2	56
3. KOLEJKOWANIE STRUMIENI MMPP	57
3.1. Definicja i własności MMPP	57
3.2. Charakterystyki kolejowania MMPP	66
3.2.1. Długość kolejki	67
3.2.2. Współczynnik strat pakietów	74
3.2.3. Opóźnienie	82
3.2.4. Okres przepełnienia bufora	86
3.2.5. Liczba strat w okresie przepełnienia bufora	97

3.2.6. Czas do przepełnienia bufora	100
3.2.7. Złożoność obliczeniowa	105
3.3. Nota bibliograficzna do Rozdziału 3	107
4. KOLEJKOWANIE STRUMIENI BMAP	108
4.1. Definicja i własności BMAP	108
4.2. Charakterystyki kolejki BMAP	116
4.2.1. Długość kolejki	116
4.2.2. Opóźnienie	122
4.2.3. Czas do przepełnienia bufora	130
4.2.4. Złożoność obliczeniowa	133
4.3. Nota bibliograficzna do Rozdziału 4	133
5. NARZĘDZIA	136
5.1. Odwracanie transformat	136
5.1.1. Odwracanie transformaty Laplace'a	136
5.1.2. Odwracanie dwuwymiarowej transformaty Laplace'a	139
5.1.3. Odwracanie funkcji tworzących	141
5.2. Metoda uniformizacji	143
5.3. Estymacja parametrów MMPP i BMAP	145
5.4. Prawdopodobieństwo całkowite	151
5.5. Symulacja BMAP	154
PODSUMOWANIE	156
BIBLIOGRAFIA	158
STRESZCZENIE	173
INDEKS	176
SPIS OZNACZEŃ	179
SPIS RYSUNKÓW I TABEL	183

CONTENTS

INTRODUCTION	9
1. POTENTIAL METHOD	14
1.1. Multi-dimensional potential method	14
1.2. One-dimensional potential method	17
1.3. Bibliographical note	18
2. QUEUEING OF THE POISSON PROCESS	20
2.1. Model of the queue	20
2.2. Poisson process queueing characteristics	21
2.2.1. Queue length	23
2.2.2. Loss ratio	28
2.2.3. Delay	28
2.2.4. Buffer overflow period	31
2.2.5. Number of losses in the overflow period	38
2.3. Queueing of the compound Poisson process	43
2.3.1. Buffer overflow period	46
2.3.2. Time to buffer overflow	53
2.4. Bibliographical note	56
3. QUEUEING OF THE MMPP	57
3.1. Definition and properties of the MMPP	57
3.2. MMPP queueing characteristics	66
3.2.1. Queue length	67
3.2.2. Loss ratio	74
3.2.3. Delay	82
3.2.4. Buffer overflow period	86
3.2.5. Number of losses in the overflow period	97
3.2.6. Time to buffer overflow	100

3.2.7. Numerical complexity	105
3.3. Bibliographical note	107
4. QUEUEING OF THE BMAP	108
4.1. Definition and properties of the BMAP	108
4.2. BMAP queueing characteristics	116
4.2.1. Queue length	116
4.2.2. Delay	122
4.2.3. Time to buffer overflow	130
4.2.4. Numerical complexity	133
4.3. Bibliographical note	133
5. TOOLS	136
5.1. Transform inversion	136
5.1.1. Laplace transform inversion	136
5.1.2. Two-dimensional Laplace transform inversion	139
5.1.3. Generating function inversion	141
5.2. Uniformization method	143
5.3. MMPP and BMAP parameter estimation	145
5.4. Total probability theorem	151
5.5. Simulation of the BMAP	154
SUMMARY	156
BIBLIOGRAPHY	158
ABSTRACT	174
INDEX	176
SYMBOLS	179
FIGURES AND TABLES	183

WSTĘP

Sieci pakietowe, a w szczególności sieci IP (*ang.* Internet Protocol), pełnią obecnie dominującą rolę wśród sieci komputerowych i telekomunikacyjnych. Na ich wydajność istotnie wpływają zjawiska związane z kolejkowaniem ruchu pakietów w buforach urządzeń sieciowych (ruterach IP, przełącznikach ATM), takie jak opóźnienia kolejkowania spowodowane oczekiwaniem pakietów w buforach czy straty pakietów spowodowane przepełnieniem bufora.

Umiejętność precyzyjnego obliczania parametrów kolejkowania ruchu sieciowego pozwala projektować sieci w sposób bardziej świadomy (tzn. zwiększać wydajność lub/i obniżać koszty produkcji i użytkowania urządzeń sieciowych). Na przykład, rozmiary buforów ruterów internetowych do tej pory nie są dobierane na podstawie jakichkolwiek precyzyjnych wyliczeń, lecz jedynie opierają się na intuicyjnych „regułach kciuka”¹. Czasami reguły te są wzajemnie sprzeczne [166], ponadto coraz częściej uważa się, że znacznie przeszacowują one wymagane rozmiary buforów [99, 21, 82] i nawet stukrotne zmniejszenie obecnie stosowanych buforów w ruterach nie miałyby destrukcyjnego wpływu na ich pracę [14, 86, 204].

Klasyczna teoria kolejek², zapoczątkowana przez duńskiego inżyniera A. Erlanga na początku dwudziestego wieku i potem rozwijana przez całe stulecie, jedynie w ograniczonym stopniu może być wykorzystana do wyznaczania charakterystyk kolejkowania ruchu w sieciach pakietowych. Stało się to jasne, kiedy odkryto [135, 77, 203, 163], że ruch ten ma skomplikowany statystycznie charakter, tzn. występują w nim takie zjawiska, jak samopodobieństwo (*ang.* self-similarity), dalekosiężność (*ang.* long-range dependence) czy spiętrzenia (*ang.* burstiness). Stosowane w klasycznej teorii kolejek procesy odnowy nie potrafią dobrze modelować tych zjawisk, zaś zjawiska te pogarszają charakterystyki kolejkowania (dłuższe kolejki i opóźnienia, większe straty pakietów). Dlatego obliczenia otrzymane przy

¹Jedna z popularnych reguł mówi, że rozmiar bufora powinien wynosić $b = RTT \times \text{przepustowość}$, co np. dla przepustowości 1 Gbps i typowego $RTT = 250$ ms daje $b = 250$ Mb (por. [199, 83]).

²Popularne podręczniki teorii kolejek to np. [117, 76, 74, 193], w języku polskim [90, 78].

użyciu klasycznych wzorów dają bardzo optymistycznie przeszacowane charakterystyki – często o wiele rzędów wielkości.

Po dokonaniu tego odkrycia rozpoczęto poszukiwania nowych lub adaptację znanych typów procesów stochastycznych do potrzeb modelowania tych negatywnych zjawisk w ruchu sieciowym. Badano między innymi takie typy procesów, jak: ułamkowy ruch Browna [158], mapy chaotyczne [87], procesy FARIMA [186], falki multifraktalne [175], miary multifraktalne [111] czy wreszcie procesy oparte na łańcuchach Markowa, w szczególności MMPP (*ang.* Markov-modulated Poisson process, [91]) i BMAP (*ang.* batch Markovian arrival process, [137]).

Wszystkie wymienione powyżej procesy mają swoje wady i zalety, jednak za wymienionymi na końcu procesami markowskimi przemawia kilka silnych argumentów. Modele takie są intuicyjnie dość proste, łatwe w symulacji, zaś metody analityczne związane z łańcuchami Markowa są stosunkowo dobrze rozwinięte.

Ponadto – co szczególnie ważne – **odpowiednio sparametryzowane modele markowskie pozwalają na dokładne wyznaczenie charakterystyk kolejkowania ruchu, który modelują** [181]. Stąd też wynika duża popularność procesów markowskich w modelowaniu ruchu w sieciach pakietowych.

Po zaadaptowaniu procesów markowskich do modelowania ruchu sieciowego rozpoczęto badania nad charakterystykami kolejkowania dla takich modeli ruchu. Początkowo większość prowadzonych badań poświęcona była jedynie charakterystykom stanu ustalonego oraz systemom z nieskończonym buforem, czyli takim, w którym kolejka może być dowolnie długa.

Założenie nieskończonego rozmiaru bufora ułatwia analizę, jednak praktyczna przydatność otrzymanych w ten sposób wyników jest ograniczona. W rzeczywistych urządzeniach bufony mają skończony rozmiar i to właśnie ich przepełnienia oraz związane z nimi straty pakietów istotnie wpływają na funkcjonowanie sieci. Ponadto w obecności silnie skorelowanego, samopodobnego ruchu nawet duże rozmiary buforów nie eliminują przepełnień i strat. Dlatego też z praktycznego punktu widzenia istotne jest, by przy wyznaczaniu charakterystyk kolejkowania ruchu w sieciach pakietowych brać pod uwagę skończone rozmiary buforów. W niniejszej pracy wszystkie charakterystyki wyznaczone są przy założeniu skończonych rozmiarów buforów.

Charakterystyki stanu ustalonego są oczywiście podstawą do oceny efektywności mechanizmów kolejkowania ruchu, jednak dają one tylko uśredniony obraz z ich długoterminowego działania. W ocenie efektywności mechanizmów kolejkowania ważne są również charakterystyki stanu nieustalonego pokazujące krótkookresowe zachowanie systemu. Nabierają one szczególnego znaczenia, gdy model ruchu ma skomplikowaną statystycznie strukturę. Wówczas krótkookresowe zachowanie systemu może znacznie odbiegać od uśrednionego stanu ustalonego. W tej monografii

wszystkie charakterystyki wyznaczane są zarówno dla stanu ustalonego, jak i nieustalonego.

W klasycznej teorii kolejek bada się zwykle kilka podstawowych charakterystyk, takich jak: długość kolejki, opóźnienie, okres zajętości systemu. Do ilościowego opisu mechanizmów buforowania ruchu w sieciach pakietowych, oprócz tych klasycznych parametrów, przydatne są również inne, rzadziej studiowane charakterystyki, takie jak: okres przepełnienia bufora, rozkład liczby strat pakietów w okresie przepełnienia, współczynnik strat czy czas do przepełnienia bufora. Jak same nazwy wskazują, charakterystyki te opisują dokładnie proces przepełniania się bufora oraz proces związanych z tym strat pakietów. W niniejszej pracy sporo miejsca poświęcono tym charakterystykom.

Większą część materiału tej książki stanowią oryginalne wyniki badań autora. W szczególności, Rozdziały 2, 3 i 4 niemal w całości zawierają oryginalne rezultaty, Rozdział 1 – w dużej części. Badania te prowadzone były przez autora w Instytucie Informatyki Politechniki Śląskiej, w Zakładzie Teorii i Projektowania Systemów Komputerowych. Większość prezentowanych w książce wyników została też opublikowana w czasopiśmie branżowych i materiałach konferencyjnych renomowanych wydawnictw, takich jak np. Springer, Elsevier, IEEE czy ACM.

Prezentowane tu rezultaty zostały otrzymane głównie metodami analitycznymi i są przedstawione w postaci twierdzeń matematycznych z dowodami, jednakże ich poprawność była także weryfikowana przy pomocy symulatora OMNeT++ [160] w ramach prowadzonego przez autora laboratorium oceny wydajności systemów i sieci komputerowych.

Oprócz wyników teoretycznych, w pracy pokazane zostały też liczne przykłady obliczeniowe dla konkretnych parametrów systemów.

Podsumowując:

Monografia ta poświęcona jest wyznaczaniu charakterystyk kolejkowania ruchu w sieciach pakietowych z uwzględnieniem

- **wiarygodnych modeli ruchu, umożliwiających naśladowanie skomplikowanych zjawisk statystycznych obserwowanych w ruchu sieciowym,**
 - **skończonych rozmiarów buforów w urządzeniach sieciowych,**
 - **charakterystyk zarówno dla stanu ustalonego, jak i nieustalonego,**
 - **charakterystyk opisujących proces przepełniania się bufora i strat pakietów.**
-

W książce tej nie podano konkretnych wskazówek dotyczących doboru parametrów urządzeń sieciowych, gdyż dobór ten zależy od bardzo wielu czynników, z których część nie jest tutaj dyskutowana. Przykładowo, projektując rozmiar bufora, należy wziąć pod uwagę fakt, że większy rozmiar bufora oznacza rzadziej pojawiające się przepełnienia, mniejsze straty pakietów i lepsze wykorzystanie łącza, jednak zwiększa opóźnienie kolejkowania. Dokładne wartości wszystkich tych parametrów zależą istotnie od statystycznej struktury ruchu, ta zaś od wielu innych czynników - np. dla ruterów internetowych od liczby obsługiwanych strumieni TCP³, strumieni UDP, mechanizmów kontroli przeciążeń itp. Dodatkowo bardzo istotny jest aspekt ekonomiczny. Dla urządzeń obsługujących wielogigabitowe łącza konieczne staje się stosowanie bardzo szybkich pamięci SRAM zamiast DRAM. Nie dość, że pamięci te są drogie, to jeszcze przy dużych rozmiarach buforów powstają poważne problemy konstrukcyjne związane z umieszczeniem dużej liczby układów scalonych na płycie oraz zapewnieniem im odpowiedniego chłodzenia [172].

Tak więc praca ta nie odpowiada na pytania typu „jaki ma być rozmiar bufora?”, lecz na pytania typu „jakie jest opóźnienie, współczynnik strat, długość kolejki, struktura strat dla bufora o rozmiarze 1MB?” oraz „jak zmieniają się te charakterystyki, gdy rozmiar bufora zmniejszymy (zwiększymy) dwukrotnie (dziesięciokrotnie)?”. Znając odpowiedzi na takie pytania, można zoptymalizować rozmiar bufora, ale do tego konieczne jest uwzględnienie wszystkich innych uwarunkowań.

Wykorzystywane w książce modele ruchu celowo podzielone zostały na trzy klasy: proste i złożone procesy Poissona (Rozdział 2), procesy MMPP (Rozdział 3), procesy BMAP (Rozdział 4). Klasy te mają coraz większe możliwości odwzorowywania własności ruchu w sieci kosztem coraz większej komplikacji modelu. Procesy Poissona są najłatwiejsze w opisie, parametryzacji i analizie, jednak stosunkowo rzadko mogą być wykorzystywane do modelowania ruchu ze względu na całkowity brak autokorelacji. Procesy MMPP są szczególnie ważne i godne polecenia, gdyż są rozsądnym kompromisem pomiędzy złożonością modelu a jego możliwościami. Innymi słowy, mają one najprostszą strukturę wśród modeli markowskich o dobrych (uwzględniających autokorelację) możliwościach modelowania własności ruchu. Wreszcie procesy BMAP mają największe możliwości i najbardziej złożoną budowę. Coraz bardziej skomplikowana struktura modeli ruchu pozwala prześledzić, jak metodologia obliczania charakterystyk kolejkowania rozwija się wraz z komplikacją modeli.

Szczegółowy układ treści dalszej części książki jest następujący.

W Rozdziale 1 zostanie pokazana tzw. metoda potencjału. Metoda ta służy do otrzymywania w zwartej postaci rozwiązań dużych układów równań liniowych

³Duża liczba strumieni TCP oznacza zwykle mniejsze fluktuacje zagregowanego ruchu.

o pewnej szczególnej budowie. Układy takiego typu będą się często pojawiać w dalszej części książki. Rozwiązania tych układów zostaną przedstawione przy pomocy specjalnego ciągu, zwanego potencjałem – stąd nazwa metody.

Rozdziały 2, 3 i 4 stanowią trzon tej książki i poświęcone są charakterystyce kolejkiowania dla markowskich modeli ruchu. W szczególności: Rozdział 2 poświęcony jest kolejkiowaniu prostych i złożonych procesów Poissona, Rozdział 3 – kolejkiowaniu procesów typu MMPP, zaś Rozdział 4 – kolejkiowaniu procesów typu BMAP. Rozdziały 2, 3, 4 mają następującą budowę: najpierw przedstawiona jest definicja oraz podstawowe własności modelu ruchu, potem pokazane są wzory na charakterystyki kolejkiowania z dowodami oraz przykładowymi obliczeniami numerycznymi (z wyjątkiem Rozdziału 2), wreszcie w końcowej części przedstawiona jest nota bibliograficzna, kierująca zainteresowanego Czytelnika do oryginalnych artykułów. Przykłady numeryczne do Rozdziału 2 można znaleźć w Rozdziałach 3 i 4, gdzie zostały umieszczone porównania charakterystyk kolejkiowania procesów Poissona z charakterystykami kolejkiowania MMPP i BMAP.

Rozdział 5 zawiera opis kilku niezbędnych narzędzi matematycznych i numerycznych koniecznych do praktycznego wykorzystania wyników analitycznych przedstawionych w poprzednich rozdziałach. W szczególności są to algorytmy stosowane do odwracania transformat Laplace'a oraz funkcji tworzących, metoda uniformizacji służąca do obliczania macierzy współczynników pojawiających się w charakterystykach kolejkiowania MMPP i BMAP, algorytm dopasowywania parametrów MMPP i BMAP do zarejestrowanych śladów ruchu, twierdzenie o prawdopodobieństwie całkowitym, często wykorzystywane w tej pracy, oraz kod do symulacji procesu BMAP.

Badania, których wyniki są tutaj przedstawione, były częściowo finansowane w ramach grantów KBN o numerach: 3 T11C 014 26 i N517 025 31/2997.

1. METODA POTENCJAŁU

W rozdziale tym zostanie pokazane, jak otrzymywać w zwartej postaci rozwiązania dużych układów równań liniowych o pewnej szczególnej budowie. Układy takiego typu będą się często pojawiać w dalszej części książki. Rozwiązania tych układów zostaną przedstawione przy pomocy specjalnego ciągu, zwanego potencjałem.

1.1. Przypadek wielowymiarowy

Teoria potencjału powstała z myślą o badaniu całkowitoliczbowych błędzeń przypadkowych. O ile w swojej głównej części jest ona słabo związana z tematyką tej monografii, o tyle niektóre z rezultatów otrzymanych przy jej pomocy okazują się niezwykle pożyteczne. Wynika to z faktu, że długość kolejki w systemie, zanotowana zaraz po ukończeniu obsługi, zachowuje się podobnie jak całkowitoliczbowe błędzenie przypadkowe ciągłe z dołu¹. W związku z tym przy analizie systemów kolejkowych często mamy do czynienia z układami równań w podobnej postaci jak ta dla błędzeń przypadkowych ciągłych z dołu. Przy pomocy teorii potencjału możemy przedstawiać w zwartej postaci rozwiązania tych układów.

Nie będziemy tu się zagłębiać w istotę teorii potencjału, podane zostaną jedynie najważniejsze wyniki potrzebne przy wyznaczaniu charakterystyk kolejkowania.

Rozpoczniemy od następującego twierdzenia.

Twierdzenie 1.1. *Dany jest ciąg $\psi_1, \psi_2, \psi_3, \dots$ wektorów kolumnowych o rozmiarze m oraz ciąg A_0, A_1, A_2, \dots macierzy o wymiarach $m \times m$ taki, że $\det(A_0) \neq 0$. Wtedy każde rozwiązanie układu równań*

$$\sum_{k=-1}^{n-1} A_{k+1}x_{n-k} - x_n = \psi_n, \quad n \geq 1, \quad (1.1)$$

¹Tzn. takie, w którym w kolejnych krokach możliwy jest tylko skok o dowolną liczbę naturalną w górę lub o 1 w dół (zob. [47]).

gdzie x_1, x_2, \dots jest szukanym ciągiem wektorów kolumnowych o rozmiarze m , ma postać

$$x_n = R_n c + \sum_{k=1}^n R_{n-k} \psi_k, \quad n \geq 1, \quad (1.2)$$

gdzie c jest pewnym, niezależnym od n , wektorem kolumnowym, zaś ciąg R_k jest zdefiniowany rekurencyjnie w następujący sposób:

$$R_0 = \mathbf{0}, \quad R_1 = A_0^{-1}, \quad R_{k+1} = R_1 \left(R_k - \sum_{i=0}^k A_{i+1} R_{k-i} \right), \quad k \geq 1, \quad (1.3)$$

gdzie $\mathbf{0}$ oznacza zerową macierz $m \times m$.

Zanim przeprowadzony zostanie dowód tego twierdzenia, zapamiętajmy, że ciąg R_k macierzy $m \times m$ nazywamy *potencjałem* dla ciągu A_k . Zwróćmy też uwagę na nietypowe indeksowanie, zaczynające się od $k = -1$, we wzorze (1.1). Oczywiście można przenumerać indeksy, zaczynając od $k = 0$ lub $k = 1$, jednak zapis od -1 ma uzasadnienie w teorii potencjału (skok błędzenia o 1 w dół), dlatego zostawimy go tu bez zmian.

Dowód Twierdzenia 1.1. Zdefiniujmy najpierw operator K działający na wektory kolumnowe rozmiaru m lub macierze kwadratowe $m \times m$ według następującego wzoru:

$$K\{a_n\} = \sum_{k=-1}^{n-1} A_{k+1} a_{n-k} - a_n, \quad (1.4)$$

Łatwo zauważyć, że układ (1.1) może zostać zapisany równoważnie przy pomocy K jako

$$K\{x_n\} = \psi_n.$$

Tak więc, aby udowodnić twierdzenie, musimy pokazać, że

$$K\left\{R_n c + \sum_{k=1}^n R_{n-k} \psi_k\right\} = \psi_n, \quad n \geq 1. \quad (1.5)$$

By tego dokonać, zauważmy najpierw, że z definicji (1.3) wynika, że

$$K\{R_n\} = \mathbf{0}$$

dla każdego $n \geq 1$. Dalej, biorąc po uwagę, że $R_0 = \mathbf{0}$, otrzymujemy

$$\begin{aligned}
& K\left\{\sum_{k=1}^n R_{n-k}\psi_k\right\} \\
&= \sum_{k=-1}^{n-1} A_{k+1} \sum_{m=1}^{n-k} R_{n-k-m}\psi_m - \sum_{k=1}^n R_{n-k}\psi_k \\
&= A_0 \sum_{k=1}^{n+1} R_{n+1-k}\psi_k + \sum_{k=0}^{n-1} A_{k+1} \sum_{m=1}^{n-k} R_{n-k-m}\psi_m - \sum_{k=1}^n R_{n-k}\psi_k \\
&= A_0 \sum_{k=0}^{n-1} R_{k+1}\psi_{n-k} + \sum_{m=1}^n \sum_{k=0}^{n-m} A_{k+1} R_{n-k-m}\psi_m - \sum_{k=0}^{n-1} R_k\psi_{n-k} \\
&= \psi_n + A_0 \sum_{k=1}^{n-1} R_{k+1}\psi_{n-k} + \sum_{k=1}^{n-1} \sum_{m=0}^k A_{m+1} R_{k-m}\psi_{n-k} - \sum_{k=1}^{n-1} R_k\psi_{n-k} \\
&= \psi_n + \sum_{k=1}^{n-1} K\{R_k\}\psi_{n-k} = \psi_n.
\end{aligned}$$

W ten sposób udowodniliśmy, że ciąg x_n w postaci (1.2) spełnia układ (1.1). Musimy jeszcze pokazać, że każde rozwiązanie (1.1) ma postać (1.2). W tym celu zauważmy, że (1.1) można zapisać w postaci

$$x_{n+1} = A_0^{-1}(\psi_n + x_n - \sum_{k=0}^{n-1} A_{k+1}x_{n-k}), \quad n \geq 1.$$

A zatem jakikolwiek ciąg x_n spełniający (1.1) jest jednoznacznie określony przez swój pierwszy element x_1 . Z (1.2) wynika natomiast, że $x_1 = R_1c$. Fakt ten powoduje, że możliwe jest otrzymanie dowolnego $x_1 \in R^m$. Wystarczy odpowiednio dobrać wektor c , mianowicie $c = R_1^{-1}x_1$. To kończy dowód Twierdzenia 1.1. ■

Oprócz Twierdzenia 1.1, w dalszej części książki wykorzystywane będą różne wnioski z niego wynikające. Sformułujemy je dla wygody od razu.

Po pierwsze, Twierdzenie 1.1 zostało pokazane, w wypadku gdy $\psi_1, \psi_2, \psi_3, \dots$ jest ciągiem wektorów kolumnowych o rozmiarze m . Łatwo zauważyć, że twierdzenie jest prawdziwe również dla ciągu $\Psi_1, \Psi_2, \Psi_3, \dots$ macierzy o wymiarach $m \times m$. Dowód można przepisać bez zmian.

Twierdzenie 1.2. *Niech dane będą dwa ciągi macierzy o wymiarach $m \times m$: $\Psi_1, \Psi_2, \Psi_3, \dots$ oraz A_0, A_1, A_2, \dots , przy czym $\det(A_0) \neq 0$. Wtedy każde rozwiązanie*

układu równań

$$\sum_{k=-1}^{n-1} A_{k+1} Y_{n-k} - Y_n = \Psi_n, \quad n \geq 1 \quad (1.6)$$

ma postać

$$Y_n = R_n C + \sum_{k=1}^n R_{n-k} \Psi_k, \quad n \geq 1, \quad (1.7)$$

gdzie C jest pewną, niezależną od n , macierzą o wymiarach $m \times m$, zaś ciąg R_k jest zdefiniowany w (1.3).

Po drugie, Twierdzenie 1.1 zostało sformułowane dla ciągu niewiadomych wektorów numerowanych od 1 (tzn. od x_1). Czasami będziemy rozwiązywać również układy numerowane od x_0 . Wykorzystując Twierdzenie 1.1, przy pomocy operacji na indeksach możemy otrzymać następujący wniosek.

Twierdzenie 1.3. Niech będzie dany ciąg $\psi_0, \psi_1, \psi_2, \dots$ wektorów kolumnowych o rozmiarze m oraz ciąg A_0, A_1, A_2, \dots macierzy o wymiarach $m \times m$ taki, że $\det(A_0) \neq 0$. Wtedy każde rozwiązanie układu równań

$$\sum_{k=-1}^n A_{k+1} x_{n-k} - x_n = \psi_n, \quad n \geq 0, \quad (1.8)$$

ma postać

$$x_n = R_{n+1} c + \sum_{k=0}^n R_{n-k} \psi_k, \quad n \geq 0, \quad (1.9)$$

gdzie c jest pewnym, niezależnym od n , wektorem kolumnowym, zaś ciąg R_k jest zdefiniowany w (1.3).

1.2. Przypadek jednowymiarowy

Jeżeli w Twierdzeniu 1.1 zredukujemy wymiary macierzy i wektorów do $m = 1$, to otrzymamy następujący rezultat.

Twierdzenie 1.4. Niech dane będą dwa ciągi liczbowe: $\psi_1, \psi_2, \psi_3, \dots$ oraz $a_0, a_1, a_2, \dots, a_0 \neq 0$. Wtedy każde rozwiązanie układu równań

$$\sum_{k=-1}^{n-1} a_{k+1} x_{n-k} - x_n = \psi_n, \quad n \geq 1 \quad (1.10)$$

ma postać

$$x_n = cR_n + \sum_{k=1}^n R_{n-k}\psi_k, \quad n \geq 1, \quad (1.11)$$

gdzie c jest pewną stałą niezależną od n , zaś ciąg R_k ma postać:

$$R_0 = 0, \quad R_1 = \frac{1}{a_0}, \quad R_{k+1} = R_1(R_k - \sum_{i=0}^k a_{i+1}R_{k-i}), \quad k \geq 1. \quad (1.12)$$

Redukując wymiar do $m = 1$ w Twierdzeniu 1.3, otrzymujemy następujący wniosek.

Twierdzenie 1.5. Niech dane będą dwa ciągi liczbowe: $\psi_0, \psi_1, \psi_2, \dots$ oraz a_0, a_1, a_2, \dots , przy czym $a_0 \neq 0$. Wtedy każde rozwiązanie układu równań

$$\sum_{k=-1}^n a_{k+1}x_{n-k} - x_n = \psi_n, \quad n \geq 0 \quad (1.13)$$

ma postać

$$x_n = cR_{n+1} + \sum_{k=0}^n R_{n-k}\psi_k, \quad n \geq 0, \quad (1.14)$$

gdzie c jest pewną stałą niezależną od n , zaś ciąg R_k jest zdefiniowany w (1.12).

1.3. Nota bibliograficzna do Rozdziału 1

Teoria potencjału została zaproponowana w pracy [122] jako narzędzie do badania całkowitoliczbowego błędzenia przypadkowego ciągłego z dołu. Przez błędzenie takie rozumiemy ciąg zmiennych losowych χ_n postaci

$$\chi_n = \sum_{k=1}^n \xi_k, \quad n \geq 1,$$

gdzie zmienne losowe ξ_k przyjmują tylko wartości całkowite, są niezależne, o jednokowym rozkładzie i spełniają warunki

$$\sum_{k=-1}^{\infty} p_k = 1, \quad p_{-1} \neq 0,$$

gdzie $p_k = \mathbf{P}(\xi_i = k)$. Jak widać, w błędzeniu tym możliwy jest skok o jeden w dół lub o dowolną liczbę naturalną w górę.

Jak pokazano w [122], wiele charakterystyk takiego błędzenia można przedstawić przy pomocy potencjału, tzn. ciągu R_k zdefiniowanego jak w (1.12) dla $p_k = a_{k+1}$.

Ponieważ długość kolejki zanotowana zaraz po ukończeniu obsługi zachowuje się podobnie jak całkowitoliczbowe błędzenie przypadkowe ciągle z dołu, metoda potencjału nadaje się doskonale do wyznaczania charakterystyk systemów kolejkowych.

Wersja jednowymiarowa najważniejszego dla nas wyniku teorii potencjału (Twierdzenie 1.4) została przedstawiona i udowodniona w [122], zaś wersja wielowymiarowa (Twierdzenie 1.1) w pracy [67]. Przykłady wykorzystania metody potencjału do obliczania charakterystyk systemów kolejkowych innych niż omawiane w tej pracy można znaleźć w [46, 48, 50].

2. KOLEJKOWANIE PROSTYCH I ZŁOŻONYCH STRUMIENI POISSONA

W tym rozdziale przedstawione będą charakterystyki kolejkowania ruchu, przy założeniu że ma on charakter zbliżony do prostego lub złożonego strumienia Poissona. Zbadane zostaną charakterystyki stanu ustalonego i nieustalonego, między innymi rozkład długości kolejki, rozkład opóźnienia, współczynnik strat pakietów, okres wypełnienia bufora.

2.1. Model systemu kolejkowego

W dalszej części zajmować się będziemy modelem kolejkowym nazywanym pojedynczym stanowiskiem obsługi ze skończoną poczekalnią (*ang.* single-server queue with finite buffer) [74]. W modelu tym klienci napływają do stanowiska w sposób określony przez pewien, mniej lub bardziej skomplikowany, proces stochastyczny, następnie są na tym stanowisku obsługiwani, co trwa przez pewien losowy czas. Po zakończeniu obsługi klient opuszcza stanowisko. Klienci, którzy w chwili swojego przybycia znajdują zajęte stanowisko, formują kolejkę w poczekalni, przy czym długość kolejki może być co najwyżej równa (skończonej) wielkości poczekalni. Klient, który w chwili przybycia znajduje pełną poczekalnię, odchodzi bez obsługi i nigdy nie wraca.

Będziemy zakładać, że czas obsługi klientów ma rozkład określony dystrybuantą $F(t)$, która może mieć dowolną postać. W szczególności rozkład czasu obsługi może być deterministyczny, jednostajny, wykładniczy lub jeszcze inny. W wypadkach, w których dystrybuanta $F(t)$ ma gęstość, będzie ona oznaczana $f_g(t)$, tzn.

$$F(t) = \int_0^t f_g(u) du.$$

Często będziemy używać transformaty Laplace'a-Stieltjesa rozkładu czasu obsługi, tzn.

$$f(s) = \int_0^{\infty} e^{-st} dF(t), \quad \operatorname{Re}(s) > 0.$$

Będziemy zakładać, że rozmiar poczekalni wynosi $b - 1$, co oznacza, że w systemie może przebywać co najwyżej b klientów, tzn. $b - 1$ w poczekalni i 1 aktualnie obsługiwany. Klienci z poczekalni podchodzą do stanowiska obsługi według dyscypliny FIFO¹ (*ang.* First In, First Out), co oznacza, że obsługiwani są najpierw klienci najdłużej oczekujący w kolejce.

W notacji systemów kolejkowych wprowadzonej przez Kendalla [115] system taki jest oznaczany jako $\cdot/G/1/b$, gdzie G oznacza, że rozkład czasu obsługi jest dowolnego typu (*ang.* general), liczba 1, że jest jedno stanowisko obsługi, zaś b to rozmiar systemu (wielkość poczekalni plus 1). Zamiast \cdot podaje się typ procesu wejściowego, np. M dla procesu Poissona, M^X dla złożonego procesu Poissona, $MMPP$ dla procesu Poissona z modulacją łańcuchem Markowa. Tak więc pełny zapis wygląda np. tak: $M/G/1/b$, $M^X/G/1/b$, $MMPP/G/1/b$ itp.

W analizie tego modelu zawsze przyjmować będziemy, że jeżeli system nie jest pusty, to w chwili $t = 0$ rozpoczyna się obsługa klienta.

Oczywiście znaczenie słów „klient”, „obsługa” czy „poczekalnia” jest umowne. Na wielu różnych polach, na których stosuje się teorię kolejek, słowa te zmieniają swój sens. Na przykład, zamiast klientów mamy do czynienia ze zdarzeniami (*ang.* events), zadaniami (*ang.* jobs) czy wiadomościami (*ang.* messages), różnie też może wyglądać proces obsługi.

Jeżeli zajmujemy się kolejkowaniem ruchu w sieciach komputerowych, to słowo klient oznacza zwykle pakiet IP lub komórkę ATM, obsługa to po prostu przesłanie pakietu lub komórki, zaś poczekalnia to bufor w ruterze internetowym, przełączniku ATM lub innym urządzeniu sieciowym, w którym kolejowane są pakiety lub komórki przed dalszą transmisją. W dalszej części książki zamiast słowa „klient” będziemy używali albo słowa „pakiet”, albo „zdarzenie” w strumieniu wejściowym, zaś zamiast słowa „poczekalnia” – „bufor”.

2.2. Kolejowanie procesu Poissona

W procesie Poissona (np. [162]) pakiety pojawiają się po odstępach czasu, które mają jednakowy rozkład wykładniczy z parametrem λ , tzn. rozkład o gęstości prawdopodobieństwa

$$g(x) = \lambda e^{-\lambda x}, \quad x > 0. \quad (2.1)$$

Parametr λ nazywany jest też **intensywnością** procesu Poissona, ponieważ, jak łatwo pokazać, jest równy średniej liczbie pakietów pojawiających się w jednostce czasu.

¹W literaturze można też spotkać nazwę FCFS od *ang.* First Come First Served.

W procesie Poissona odstępy czasu pomiędzy kolejnymi pakietami nie zależą od siebie i ich funkcja autokorelacji jest równa tożsamościowo 0, co jest niewątpliwie wadą tego modelu. Jego najważniejszą zaletą jest zaś **własność braku pamięci**, która jest intensywnie wykorzystywana w analizie procesu Poissona i jego charakterystyk kolejkowania. Z grubsza rzecz ujmując, własność ta oznacza, że czas do pojawienia się kolejnego pakietu po dowolnej chwili t ma dokładnie ten sam rozkład wykładniczy (2.1).

Własność braku pamięci procesu Poissona jest związana z brakiem pamięci rozkładu wykładniczego², który jest jedynym rozkładem ciągłym o tej własności.

Oprócz własności braku pamięci, w dalszej części będziemy korzystać z następujących własności procesu Poissona.

- Jeżeli $N(t)$ oznacza liczbę pakietów pojawiających się w dowolnym przedziale czasu o długości t , to **rozkład prawdopodobieństwa** $N(t)$ ma następującą postać

$$m_k(t) = \mathbf{P}(N(t) = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \quad k = 0, 1, 2, \dots \quad (2.2)$$

- **Funkcja tworząca** dla rozkładu prawdopodobieństwa $N(t)$ ma postać

$$\sum_{k=0}^{\infty} z^k \mathbf{P}(N(t) = k) = e^{-\lambda t(1-z)}.$$

- **Czas pojawienia się n -tego pakietu**, licząc od chwili 0, ma następującą gęstość rozkładu prawdopodobieństwa

$$s_n(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!}, \quad n = 1, 2, 3, \dots \quad (2.3)$$

Czas ten będziemy oznaczać przez t_n . Zgodnie z własnością braku pamięci, taki sam rozkład (2.3) ma też czas pojawienia się n -tego pakietu, licząc od dowolnej chwili.

Dowody tych własności są elementarne i można je znaleźć w podręcznikach rachunku prawdopodobieństwa.

Klasyczny proces Poissona może zostać użyty jako model ruchu jedynie w sytuacji, gdy autokorelacja odstępów czasów pomiędzy pakietami jest bliska zeru. W praktyce taka sytuacja jest rzadka. Procesem Poissona warto jednak się zająć chociażby

²Dokładniejsze omówienie tego faktu (z dowodem) można znaleźć w [32].

dlatego, że jest to najprostszy proces o markowskiej strukturze i na jego przykładzie można prześledzić podstawowe techniki wyznaczania charakterystyk kolejowania. Techniki te przenoszone są potem, w rozbudowanej formie, na znacznie bardziej skomplikowane procesy markowskie.

2.2.1. Długość kolejki

Wyznaczanie charakterystyk kolejowania procesu Poissona rozpoczniemy od rozkładu prawdopodobieństwa długości kolejki w systemie. Zaprezentowana tu metodologia pozwala na otrzymanie w zwartej postaci bardzo ogólnych wyników, zarówno dla stanu ustalonego, jak i nieustalonego.

Przez $X(t)$ będziemy oznaczali długość kolejki³ w systemie w chwili t . Zajmiemy się najpierw transformatą Laplace'a rozkładu $X(t)$, a mianowicie

$$\phi_n(s, l) = \int_0^{\infty} e^{-st} \Phi_n(t, l) dt, \quad (2.4)$$

gdzie

$$\Phi_n(t, l) = \mathbf{P}(X(t) = l | X(0) = n), \quad l = 0, 1, \dots, b. \quad (2.5)$$

Oczywiście rozkład długości kolejki w chwili t zależy od początkowej długości kolejki, tzn. od $X(0)$. Ta początkowa długość będzie tu oznaczana przez n i może oczywiście przyjmować wartości całkowite w przedziale $[0, b]$.

W dalszej części często będziemy się posługiwać ciągami $a_k(s)$, $d_k(s)$, $k = 0, 1, 2, \dots$, zdefiniowanymi w następujący sposób:

$$a_k(s) = \int_0^{\infty} \frac{e^{-(\lambda+s)t} (\lambda t)^k}{k!} dF(t), \quad (2.6)$$

$$d_k(s) = \int_0^{\infty} \frac{e^{-(\lambda+s)t} (\lambda t)^k}{k!} (1 - F(t)) dt. \quad (2.7)$$

Udowodnimy teraz następujące twierdzenie.

Twierdzenie 2.1. *Transformata Laplace'a rozkładu długości kolejki w systemie $M/G/1/b$ ma postać*

$$\phi_n(s, l) = c_{b-n}(s) \frac{\lambda h_{b-1}(s, l) - (s + \lambda) h_b(s, l) + \delta_{0l}}{(s + \lambda) c_b(s) - \lambda c_{b-1}(s)} + h_{b-n}(s, l), \quad (2.8)$$

³Przyjmujemy umowę, że do długości kolejki wliczamy również klienta (pakiet) aktualnie obsługiwanego.

gdzie

$$c_k(s) = R_{k+1}(s)a_0(s) + \sum_{i=0}^k R_{k-i}(s)b_i(s), \quad (2.9)$$

$$b_k(s) = a_{k+1}(s) + \frac{1}{f(s)} \sum_{i=0}^k a_i(s) - 1, \quad (2.10)$$

$$h_k(s, l) = \sum_{i=0}^k R_{k-i}(s) \left[r_b(s, l) \left(1 - \frac{1}{f(s)} \sum_{j=0}^i a_j(s) \right) - r_{b-i}(s, l) \right], \quad (2.11)$$

$$r_k(s, l) = \begin{cases} 0 & \text{dla } l < k, \\ d_{l-k}(s) & \text{dla } k \leq l < b, \\ (1 - f(s))/s - \sum_{i=0}^{b-n-1} d_i(s) & \text{dla } l = b, \end{cases} \quad (2.12)$$

ciąg $R_k(s)$ jest potencjałem dla ciągu $a_k(s)$ (zob. wzory (1.12) i (2.6)), zaś δ_{ij} oznacza symbol Kroneckera, tzn.

$$\delta_{ij} = \begin{cases} 1 & \text{dla } i = j, \\ 0 & \text{dla } i \neq j. \end{cases}$$

Dowód Twierdzenia 2.1. Dla uproszczenia zapisu w dowodzie będziemy pomijać zależność funkcji $\phi_n(s, l)$, $\Phi_n(s, l)$, $h_k(s, l)$, $r_k(s, l)$ od l , tzn. będziemy pisać po prostu $\phi_n(s)$, $\Phi_n(s)$, $h_k(s)$, $r_k(s)$, pamiętając o tej zależności.

Żałóżmy najpierw, że początkowo system nie jest pusty, tzn. $X(0) = n \in [1, b]$. Stosując wzór na prawdopodobieństwo całkowite⁴ względem pierwszej chwili ukończenia obsługi, otrzymujemy następujący układ równań całkowych:

$$\begin{aligned} \Phi_n(t) &= \sum_{k=0}^{b-n-1} \int_0^t \Phi_{n+k-1}(t-u) \frac{e^{-\lambda u} (\lambda u)^k}{k!} dF(u) \\ &+ \sum_{k=b-n}^{\infty} \int_0^t \Phi_{b-1}(t-u) \frac{e^{-\lambda u} (\lambda u)^k}{k!} dF(u) \\ &+ \rho_n(t), \quad 0 < n \leq b, \end{aligned} \quad (2.13)$$

gdzie

$$\rho_n(t) = (1 - F(t)) \cdot \begin{cases} 0 & \text{dla } l < n, \\ \frac{e^{-\lambda t} (\lambda t)^{l-n}}{(l-n)!} & \text{dla } n \leq l < b, \\ \sum_{k=b-n}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} & \text{dla } l = b. \end{cases}$$

⁴Używana wielokrotnie w tej książce ciągła wersja popularnego twierdzenia o prawdopodobieństwie całkowitym jest przedstawiona w podrozdziale 5.4.

Pierwsza suma w (2.13) odpowiada sytuacji, w której pierwsza chwila ukończenia obsługi u wypadła przed czasem t i do czasu u nie wystąpiło przepełnienie bufora. W tym wypadku liczba pakietów przybyłych do czasu u nie może przekroczyć wartości $b - n - 1$.

Druga suma w (2.13) odpowiada sytuacji, w której pierwsza chwila ukończenia obsługi u wypadła przed czasem t i do czasu u wystąpiło przepełnienie bufora. Oznacza to, że do chwili u przybyło co najmniej $b - n$ pakietów.

Wreszcie składnik $\rho_n(t)$ w (2.13) odpowiada sytuacji, w której pierwsza chwila ukończenia obsługi wypadła po czasie t . Prawdopodobieństwo takiego zdarzenia wynosi $1 - F(t)$ i w takiej sytuacji kolejka ma długość $l < b$, gdy do systemu przybyło $l - n$ pakietów, zaś kolejkę długości b (przepełnienie) obserwujemy, gdy do systemu przybyło $b - n$ lub więcej pakietów.

Założmy teraz, że początkowo system jest pusty, tzn. $X(0) = 0$. Wykorzystując wzór na prawdopodobieństwo całkowite, tym razem względem chwili przybycia pierwszego pakietu, możemy napisać następujące równanie całkowe:

$$\Phi_0(t) = \int_0^t \Phi_1(t-u)\lambda e^{-\lambda u} du + \delta_{0l} e^{-\lambda t}. \quad (2.14)$$

Pierwszy składnik w (2.14) odpowiada sytuacji, gdy przed chwilą t przybywa pierwszy pakiet. Drugi składnik w (2.14) odpowiada sytuacji, gdy do chwili t nie przybywa żaden pakiet. Prawdopodobieństwo takiego zdarzenia wynosi $e^{-\lambda t}$ i długość kolejki jest w takiej sytuacji równa 0.

Naszym celem będzie teraz rozwiązanie układu równań (2.13) przy pomocy metody potencjału, a dokładniej Twierdzenia 1.5. Najpierw jednak musimy sprowadzić (2.13) do postaci (1.13). W tym celu obie strony (2.14) i (2.13) przekształcamy transformatą Laplace'a. Wykorzystując fakt, że transformata spłotu funkcji jest równa iloczynowi transformat (zob. [161], str. 155), otrzymujemy

$$\phi_n(s) = \sum_{k=0}^{b-n-1} a_k(s)\phi_{n+k-1}(s) + \sum_{k=b-n}^{\infty} a_k(s)\phi_{b-1}(s) + r_n(s), \quad 0 < n \leq b \quad (2.15)$$

oraz

$$\phi_0(s) = \frac{\lambda}{s + \lambda} \phi_1(s) + \frac{\delta_{0l}}{s + \lambda}. \quad (2.16)$$

Następnie wykonujemy podstawienie

$$\varphi_n(s) = \phi_{b-n}(s).$$

W rezultacie z (2.15) i (2.16) dostajemy:

$$\sum_{k=-1}^n a_{k+1}(s)\varphi_{n-k}(s) - \varphi_n(s) = \psi_n(s), \quad 0 \leq n < b, \quad (2.17)$$

$$\varphi_b(s) = \frac{\lambda}{s + \lambda} \varphi_{b-1}(s) + \frac{\delta_{0l}}{s + \lambda}, \quad (2.18)$$

gdzie

$$\psi_n(s) = a_{n+1}(s)\varphi_0(s) - \sum_{k=n+1}^{\infty} a_k(s)\varphi_1(s) - r_{b-n}(s).$$

Zauważmy, że układ (2.17) ma teraz dokładnie taką samą postać jak (1.13). A zatem, gdy zastosuje się Twierdzenie 1.5, każde jego rozwiązanie ma postać

$$\varphi_n(s) = c(s)R_{n+1}(s) + \sum_{k=0}^n R_{n-k}(s)\psi_k(s), \quad n \geq 0, \quad (2.19)$$

gdzie $c(s)$ nie zależy od n .

Aby zakończyć dowód twierdzenia, musimy teraz wyznaczyć $c(s)$ oraz $\varphi_0(s)$, $\varphi_1(s)$ występujące w $\psi_n(s)$. Podstawiając $n = 0$ do (2.19), dostajemy

$$c(s) = \frac{\varphi_0(s)}{R_1(s)} = a_0(s)\varphi_0(s), \quad (2.20)$$

zaś podstawiając $n = 0$ do (2.17) oraz wykorzystując fakt, że

$$\sum_{k=0}^{\infty} a_k(s) = f(s),$$

otrzymujemy

$$\varphi_1(s) = \frac{\varphi_0(s) - r_b(s)}{f(s)}. \quad (2.21)$$

Następnie, podstawiając (2.20) oraz (2.21) do wzoru (2.19), otrzymujemy

$$\begin{aligned} \varphi_n(s) &= \varphi_0(s) \left[R_{n+1}(s)a_0(s) + \sum_{k=0}^n R_{n-k}(s) \left(a_{k+1}(s) - \frac{1}{f(s)} \sum_{i=k+1}^{\infty} a_i(s) \right) \right] \\ &\quad + \frac{1}{f(s)} \sum_{k=0}^n R_{n-k}(s) \left(\sum_{i=k+1}^{\infty} a_i(s)r_b(s) - r_{b-k}(s) \right) \\ &= \varphi_0(s)c_n(s) + h_n(s). \end{aligned} \quad (2.22)$$

Teraz możemy łatwo obliczyć $\varphi_0(s)$. Wystarczy podstawić $n = b$ oraz $n = b - 1$ do (2.22) i wykorzystać warunek (2.18). Otrzymujemy wtedy, że:

$$\varphi_0(s) = \phi_b(s) = \frac{\lambda h_{b-1}(s, l) - (s + \lambda)h_b(s, l) + \delta_{0l}}{(s + \lambda)c_b(s) - \lambda c_{b-1}(s)}. \quad (2.23)$$

Wstawiając (2.23) do (2.22), dostajemy tezę twierdzenia. ■

Twierdzenie 2.1 może zostać wykorzystane w praktyce na wiele sposobów.

Po pierwsze, można przy jego pomocy otrzymać rozkład długości kolejki w stanie ustalonym, tzn.

$$\Phi(l) = \Phi_n(\infty, l) = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) = l | X(0) = n).$$

Wartości $\Phi(l)$ obliczymy łatwo, korzystając z własności granicznych transformaty Laplace'a (zob. [161], str. 145):

$$\Phi(l) = \lim_{s \rightarrow 0^+} s\phi_n(s, l). \quad (2.24)$$

Ponieważ rozkład długości kolejki w stanie ustalonym nie zależy od początkowej długości kolejki, możemy przyjąć dowolne n we wzorze (2.24). Najwygodniej przyjmować $n = b$, ponieważ w tej sytuacji (2.8) redukuje się do najprostszej postaci, czyli (2.23).

Po drugie, Twierdzenie 2.1 można wykorzystać do wyznaczenia rozkładu długości kolejki w stanie nieustalonym, tzn. w dowolnej chwili t . W tym celu konieczne jest zastosowanie jednego z algorytmów numerycznego odwracania transformaty Laplace'a. Znanych jest kilka algorytmów tego rodzaju, np. [79, 1, 2]. Jednym z ciekawszych (szybszych) jest algorytm oparty na sumie Eulera [2]. Został on przedstawiony w podrozdziale 5.1.1 tej książki.

Po trzecie, podstawiając $l = b$ w Twierdzeniu 2.1, możemy wyznaczyć prawdopodobieństwo przepełnienia bufora. Tak jak poprzednio, charakterystykę tę możemy obliczać zarówno dla stanu nieustalonego (tzn. $\Phi_n(t, b)$), jak i ustalonego (tzn. $\Phi(b)$). Wartość $\Phi_n(t, b)$ informuje nas, jakie jest prawdopodobieństwo, że w chwili t bufor jest przepełniony, pod warunkiem że początkowa długość kolejki wynosiła n . Wartość $\Phi(b)$ informuje nas, jakie jest prawdopodobieństwo, że bufor jest przepełniony w stanie ustalonym.

Na zakończenie zauważmy, że ciągi $a_k(s)$, $d_k(s)$, na których oparte są wyniki Twierdzenia 2.1, nie sprawiają kłopotów obliczeniowych. Dla wielu typowych postaci dystrybuanty $F(t)$ ciągi te mogą zostać wyznaczone przez całkowanie symboliczne. Jeżeli, na przykład, czas obsługi jest stały i wynosi d , to

$$\begin{aligned} a_k(s) &= \frac{e^{-(\lambda+s)d} (\lambda d)^k}{k!}, \\ d_k(s) &= \frac{\lambda^k}{(s + \lambda)^{k+1} k!} (\Gamma(k + 1) - \Gamma(k + 1, sd + \lambda d)), \end{aligned}$$

gdzie $\Gamma(\cdot)$ oznacza funkcję gamma, zaś $\Gamma(\cdot, \cdot)$ – niekompletną funkcję gamma.

Jeżeli czas obsługi ma rozkład wykładniczy, tzn. $F(t) = 1 - e^{-\mu t}$, to

$$a_k(s) = \frac{\mu \lambda^k}{(s + \mu + \lambda)^{k+1}},$$

$$d_k(s) = \frac{\lambda^k}{(s + \lambda)^{k+1}} - \frac{\lambda^k}{(s + \mu + \lambda)^{k+1}}.$$

Wypadki, w których dystrybuanta F ma bardziej skomplikowaną postać i nie jest możliwe analityczne wyznaczenie $a_k(s)$, $d_k(s)$, również nie stanowią poważnego problemu, gdyż wzory (2.6), (2.7) dobrze poddają się standardowym algorytmom całkowania numerycznego.

2.2.2. Współczynnik strat

Współczynnik strat, zwykle obliczany dla stanu ustalonego, jest jedną z najważniejszych charakterystyk kolejkowania ruchu w sieciach pakietowych. Mówi on nam, jaka część ogólnej liczby pakietów jest tracona na skutek przepełnienia bufora w dłuższych okresach czasu. Współczynnik strat (*ang.* loss ratio) będziemy oznaczać przez LR .

Współczynnik strat dla procesu Poissona wyznacza się prosto – jest on równy prawdopodobieństwu przepełnienia bufora w stanie ustalonym, tzn.

$$LR = \Phi(b). \quad (2.25)$$

A zatem można go obliczyć, korzystając ze wzorów (2.24) i (2.8). Wzór (2.25) wiąże się ze specjalnymi własnościami procesu Poissona i nie jest prawdziwy dla bardziej skomplikowanych strumieni wejściowych.

2.2.3. Opóźnienie

Następną charakterystyką, którą się zajmiemy, będzie opóźnienie spowodowane kolejkowaniem (ściśle mówiąc – rozkład tego opóźnienia). Przez opóźnienie $V(t)$ będziemy rozumieć czas, jaki by spędził w buforze pakiet przybyły w chwili t . Charakterystyka ta bywa nazywana też wirtualnym czasem oczekiwania⁵. Do $V(t)$ nie będziemy zaliczać samego czasu obsługi, jedynie czas przebywania w kolejce. Ponadto zakładamy tu, że opóźnienie dla pakietu utraconego (tzn. przybyłego do systemu w chwili, gdy bufor jest przepełniony) wynosi 0. Czasem stosuje się też alternatywne podejście, w którym takie pakiety nie są w ogóle brane pod uwagę – łatwo

⁵W literaturze anglojęzycznej można spotkać kilka nazw: virtual waiting time, workload, delay.

jest przeliczyć rozkład opóźnienia z jednej sytuacji na drugą, wykorzystując $\Phi_n(t, b)$. Mamy mianowicie

$$\mathbf{P}(V^*(t) > x | X(0) = n) = \frac{\mathbf{P}(V(t) > x | X(0) = n)}{1 - \Phi_n(t, b)},$$

gdzie $V^*(t)$ oznacza opóźnienie liczone jedynie dla pakietów, które nie zostały utracone (por. [193], str. 220).

Oczywiste jest, jak ważnym parametrem z praktycznego punktu widzenia jest opóźnienie. To właśnie opóźnienia wywołane kolejkowaniem są jednym z istotnych czynników wpływających na odbieraną przez użytkowników jakość usług w sieciach pakietowych.

Głównym wynikiem, który tu udowodnimy, będzie wzór dla podwójnej transformaty Laplace'a rozkładu opóźnienia:

$$w_n(s, \sigma) = \int_0^\infty e^{-\sigma t} dt \int_0^\infty e^{-sx} \tilde{w}_n(x, t) dx,$$

gdzie

$$\tilde{w}_n(x, t) = \mathbf{P}(V(t) > x | X(0) = n). \quad (2.26)$$

Twierdzenie 2.2. *Transformata Laplace'a rozkładu opóźnienia w systemie M/G/1/b ma postać*

$$w_n(s, \sigma) = c_{b-n}(\sigma) \frac{\lambda z_{b-1}(s, \sigma) - (\sigma + \lambda) z_b(s, \sigma)}{(\sigma + \lambda) c_b(\sigma) - \lambda c_{b-1}(\sigma)} + z_{b-n}(s, \sigma), \quad (2.27)$$

gdzie

$$z_k(s, \sigma) = \sum_{i=0}^k R_{k-i}(\sigma) y_i(s, \sigma), \quad (2.28)$$

$$y_k(s, \sigma) = q_b(s, \sigma) \left(1 - \frac{1}{f(\sigma)} \sum_{i=0}^k a_i(\sigma) \right) - q_{b-k}(s, \sigma), \quad (2.29)$$

$$q_n(s, \sigma) = \frac{1}{s} \sum_{k=0}^{b-n-1} [d_k(\sigma) - e_k(s, \sigma) (f(s))^{n+k-1}], \quad (2.30)$$

$$e_k(s, \sigma) = \int_0^\infty \frac{e^{-(\lambda+\sigma)t} (\lambda t)^k}{k!} dt \int_0^\infty e^{-sx} d_x F(x+t), \quad (2.31)$$

ciąg $R_k(\sigma)$ jest potencjałem dla ciągu $a_k(\sigma)$ (wzory (1.12) i (2.6)), zaś ciagi $c_k(\sigma)$, $d_k(\sigma)$ są określone w (2.9) i (2.7).

Dowód Twierdzenia 2.2. Zakładamy najpierw, że w chwili $t = 0$ system nie jest pusty, tzn. $X(0) = n > 0$. Stosując wzór na prawdopodobieństwo całkowite względem pierwszego czasu ukończenia obsługi, otrzymujemy dla $0 < n \leq b$:

$$\begin{aligned} \tilde{w}_n(x, t) &= \sum_{k=0}^{b-n-1} \int_0^t \tilde{w}_{n+k-1}(x, t-u) \frac{e^{-\lambda u} (\lambda u)^k}{k!} dF(u) \\ &+ \sum_{k=b-n}^{\infty} \int_0^t \tilde{w}_{b-1}(x, t-u) \frac{e^{-\lambda u} (\lambda u)^k}{k!} dF(u) \\ &+ \sum_{k=0}^{b-n-1} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \int_t^{\infty} (1 - F^{(n+k-1)*}(x-u+t)) dF(u). \end{aligned} \quad (2.32)$$

Pierwsza suma w (2.32) odpowiada sytuacji, w której pierwszy czas ukończenia obsługi u wypadł przed t i do czasu u nie wystąpiło przepełnienie bufora; w tym wypadku liczba pakietów przybyłych do czasu u nie może przekroczyć $b - n - 1$. Druga suma w (2.32) odpowiada sytuacji, w której pierwszy czas ukończenia obsługi u wypadł przed t i do czasu u wystąpiło przepełnienie bufora. Trzecia suma w (2.32) odpowiada sytuacji, w której pierwszy czas ukończenia obsługi u wypadł po czasie t . Wówczas prawdopodobieństwo zdarzenia $V(t) > x$ jest równe

$$1 - F^{(n+k-1)*}(x-u+t),$$

gdzie $F^{(n+k-1)*}$ oznacza $n+k-1$ -krotny spłot dystrybuanty F , zaś k jest liczbą pakietów przybyłych w przedziale czasu $(0, t]$.

Zakładając, że początkowo system jest pusty, i stosując wzór na prawdopodobieństwo całkowite względem chwili przybycia pierwszego pakietu, otrzymujemy, że

$$\tilde{w}_0(x, t) = \int_0^t \tilde{w}_1(x, t-u) \lambda e^{-\lambda u} du. \quad (2.33)$$

Stosując transformatę Laplace'a do obu stron (2.32) i (2.33), otrzymujemy:

$$\begin{aligned} w_n(s, \sigma) &= \sum_{k=0}^{b-n-1} a_k(\sigma) w_{n+k-1}(s, \sigma) \\ &+ \sum_{k=b-n}^{\infty} a_k(\sigma) w_{b-1}(s, \sigma) + q_n(s, \sigma), \quad 0 < n \leq b \end{aligned} \quad (2.34)$$

i

$$w_0(s, \sigma) = \frac{\sigma}{\sigma + \lambda} w_1(s, \sigma). \quad (2.35)$$

Zauważmy teraz, że równania (2.34) i (2.35) mają taką samą postać jak (2.15) i (2.16). Rozwiązując je krok po kroku, dokładnie tak jak w dowodzie Twierdzenia 2.1, łatwo otrzymujemy tezę Twierdzenia 2.2. ■

Dzięki temu, że w definicji (2.26) występuje ogon rozkładu $V(t)$, wzór (2.27) można łatwo wykorzystać do obliczania średniej wartości opóźnienia – wystarczy podstawić $s = 0$. Stąd średnia wartość opóźnienia w stanie ustalonym jest równa

$$\lim_{\sigma \rightarrow 0^+} \sigma w_n(0, \sigma),$$

gdzie n jest dowolne⁶. Obliczenie średniej wartości opóźnienia w dowolnej chwili t wymaga odwrócenia (pojedynczej) transformaty $w_n(0, \sigma)$ ze względu na zmienną σ . Wreszcie, jeżeli interesuje nas dokładna postać rozkładu opóźnienia, a nie tylko jego wartość średnia, należy zastosować algorytm do odwracania podwójnej transformaty Laplace'a. Algorytm taki został przedstawiony w podrozdziale 5.1.2.

2.2.4. Okres przepelnienia bufora

Jak już zostało wcześniej powiedziane, jednym z ważniejszych zjawisk wpływających na wydajność kolejowania ruchu w sieciach pakietowych są straty, tzn. pakiety, które zostały utracone w czasie przeciążenia sieci na skutek przepelnienia buforów. Podstawowymi parametrami, które opisują to zjawisko, są prawdopodobieństwo przepelnienia bufora $\Phi(b)$ oraz współczynnik strat – parametry te zostały omówione w podrozdziale 2.2.1.

Jak się jednak okazuje, parametry te niosą bardzo ogólną informację dotyczącą procesu strat i informacja ta jest czasami niewystarczająca.

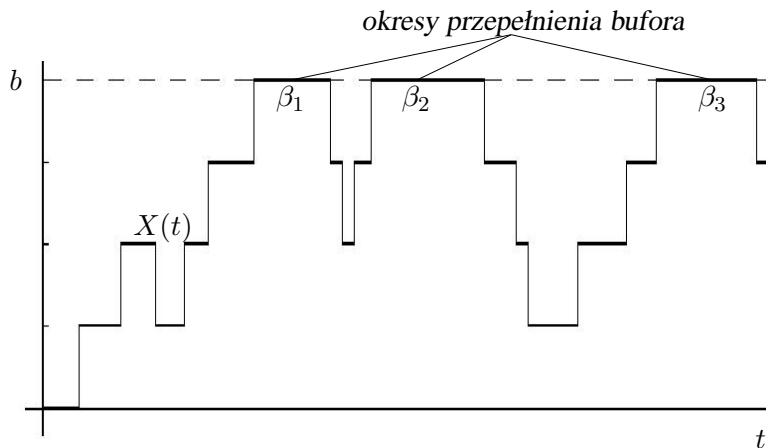
Przykładowo, zastosowanie techniki FEC⁷ pozwala znacząco zwiększyć tolerowany współczynnik strat pakietów. Technika FEC polega na tym, że do każdego bloku k wysyłanych pakietów dodaje się h pakietów nadmiarowych, tak aby utrata dowolnych h pakietów z bloku $k + h$ nie powodowała konieczności retransmisji (tzn. informacja może zostać odtworzona). Jak widać z tego opisu, przy zastosowaniu FEC nie tyle istotny staje się współczynnik strat, co statystyczna struktura strat. Innymi słowy, ważne jest, czy straty mają tendencję do grupowania się, występowania w dużych blokach, czy też są raczej równomiernie rozłożone w czasie.

⁶Można też wyznaczyć momenty wyższych rzędów w stanie ustalonym. Więcej informacji na temat obliczania momentów rozkładów przy pomocy ich transformat można znaleźć w [44].

⁷Forward Error Correction (zobacz np. w [110]). Przykłady zastosowania techniki FEC w sieciach pakietowych można znaleźć w [123, 177, 157, 36, 164].

Podobny efekt obserwujemy przy transmisji VBR (Variable Bit Rate) strumieni video, gdzie na jakość przekazu odbieraną przez użytkownika większy wpływ mają bloki utraconych pakietów niż nawet duże, ale równomiernie rozłożone w czasie straty.

W związku z tym do opisu procesu strat stosuje się uzupełniające charakterystyki, pozwalające ocenić statystyczną strukturę tych strat.



Rys. 2.1. Przykładowa ewolucja długości kolejki w systemie z buforem o rozmiarze b
Fig. 2.1. Sample evolution of the queue size process in a system with a buffer of size b

Pierwszą tego typu charakterystyką jest rozkład długości okresu przepełnienia bufora. Okresy przepełnienia bufora dla przykładowej ewolucji procesu długości kolejki pokazuje rys. 2.1. Ponieważ w okresie przepełnienia nadchodzące pakiety są kolejno tracone, to rozkład długości okresu przepełnienia wpływa na rozkład liczby pakietów, które zostały utracone jeden po drugim.

W tym podrozdziale zajmiemy się właśnie rozkładem długości okresu przepełnienia bufora. Rozpocniemy od pierwszego okresu przepełnienia bufora, tzn. β_1 na rys. 2.1. Formalnie pierwszy okres przepełnienia definiuje się jako

$$\beta_1 = \zeta - \tau,$$

gdzie τ oznacza pierwszy czas przepełnienia bufora, tzn.

$$\tau = \inf\{t > 0 : X(t) = b\}, \quad (2.36)$$

zaś ζ oznacza pierwszy po czasie τ czas ukończenia obsługi.

Z definicji tej wynika natychmiast, że okres przepełnienia bufora jest równoznaczny z czasem obsługi pozostałym po osiągnięciu przez długość kolejki poziomu b .

Oznaczmy przez $H_n(t)$ dystrybuantę rozkładu pierwszego okresu przepelnienia bufora, tzn.

$$H_n(t) = \mathbf{P}(\beta_1 < t | X(0) = n),$$

gdzie $0 \leq n < b$.

Twierdzenie 2.3. *W systemie M/G/1/b rozkład pierwszego okresu przepelnienia bufora ma postać*

$$H_n(t) = 1 - \frac{\sum_{k=1}^{b-1} (R_{b-k} - R_{b-1-k}) v_k(t)}{\sum_{k=0}^{b-1} (R_{b-k} - R_{b-1-k}) a_k} (R_{b-n-1} + \delta_{0, b-n-1}) + \sum_{k=1}^{b-n} v_k(t) R_{b-n-k}, \quad (2.37)$$

gdzie

$$v_k(t) = \lambda \int_0^\infty (1 - F(u+t)) \frac{e^{-\lambda u} (\lambda u)^{k-1}}{(k-1)!} du, \quad k = 1, 2, \dots,$$

ciąg $a_k = a_k(0)$ jest określony w (2.6) oraz ciąg R_k jest potencjałem dla ciągu a_k .

Dowód Twierdzenia 2.3. Jeżeli system początkowo nie jest pusty, to stosując wzór na prawdopodobieństwo całkowite względem pierwszego czasu ukończenia obsługi u , otrzymujemy układ równań dla $0 < n < b$:

$$\begin{aligned} & \mathbf{P}(\beta_1 > t | X(0) = n) \\ &= \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\beta_1 > t | X(0) = n+k-1) \frac{e^{-\lambda u} (\lambda u)^k}{k!} dF(u) \\ & \quad + \int_0^\infty dF(u) \int_0^u \lambda \frac{e^{-\lambda v} (\lambda v)^{b-n-1}}{(b-n-1)!} I(u-v > t) dv, \end{aligned} \quad (2.38)$$

gdzie

$$I(x > y) = \begin{cases} 1 & \text{dla } x > y, \\ 0 & \text{dla } x \leq y. \end{cases}$$

Pierwszy składnik (2.38) odpowiada sytuacji, w której do chwili pierwszego ukończenia obsługi u nie wystąpiło przepelnienie, tzn. w czasie $(0, u]$ przybyło najwyżej $b-n-1$ pakietów. Drugi składnik (2.38) odpowiada sytuacji, w której $(b-n-1)$ -szy pakiet pojawia się w chwili v , gdzie $v < u$, i wtedy

$$\mathbf{P}(\beta_1 > t) = I(u-v > t).$$

Gęstość rozkładu czasu pojawienia się $(b - n - 1)$ -szego pakietu ma postać

$$\lambda \frac{e^{-\lambda v} (\lambda v)^{b-n-1}}{(b-n-1)!}$$

(zobacz (2.3)).

Dla systemu, który jest pusty w chwili $t = 0$, możemy napisać następujące równanie:

$$\mathbf{P}(\beta_1 > t | X(0) = 0) = \mathbf{P}(\beta_1 > t | X(0) = 1). \quad (2.39)$$

Innymi słowy, rozkład β_1 jest taki sam dla początkowo pustego systemu, jak dla systemu z jednym pakietem. Wynika to z faktu, że system, który jest początkowo pusty, musi przed przepełnieniem bufora choć raz zawierać dokładnie jeden pakiet.

W dalszych obliczeniach użyjemy skróconego zapisu:

$$\chi_n(t) = \mathbf{P}(\beta_1 > t | X(0) = n).$$

Stosując ten zapis do (2.38) i (2.39) oraz wykonując całkowanie przez części drugiego składnika (2.38), otrzymujemy

$$\chi_n(t) = \sum_{k=0}^{b-n-1} \chi_{n+k-1}(t) a_k + v_{b-n}(t), \quad 0 < n < b \quad (2.40)$$

i

$$\chi_0(t) = \chi_1(t). \quad (2.41)$$

Wykonując podstawienie

$$x_n(t) = \chi_{b-n}(t),$$

z (2.40) i (2.41) dostajemy:

$$\sum_{k=-1}^{n-1} a_{k+1} x_{n-k}(t) - x_n(t) = \psi_n(t), \quad 0 < n < b, \quad (2.42)$$

$$x_b(t) = x_{b-1}(t) \quad (2.43)$$

dla

$$\psi_n(t) = x_1(t) a_n - v_n(t).$$

Zauważmy, że (2.42) ma dokładnie taką samą postać jak (1.10). Zatem możemy zastosować Twierdzenie 1.4. Otrzymujemy rozwiązanie postaci

$$x_n(t) = c(t) R_n + \sum_{k=1}^n \psi_k(t) R_{n-k}, \quad n \geq 1. \quad (2.44)$$

Wstawiając $n = 1$ do (2.44), obliczamy $c(t)$

$$c(t) = \frac{x_1(t)}{R_1}.$$

Dalej, podstawiając $n = b$ oraz $n = b - 1$ do (2.44) oraz stosując (2.43), obliczamy $x_1(t)$:

$$x_1(t) = \frac{\sum_{k=1}^{b-1} (R_{b-k} - R_{b-1-k})v_k(t)}{\sum_{k=0}^{b-1} (R_{b-k} - R_{b-1-k})a_k}. \quad (2.45)$$

Podstawienie odwrotne χ_n za x_{b-n} w (2.44) daje

$$\chi_n(t) = x_1(t) \sum_{k=0}^{b-n} a_k R_{b-n-k} - \sum_{k=1}^{b-n} v_k(t) R_{b-n-k}. \quad (2.46)$$

Ponieważ, zgodnie z (1.12)

$$\sum_{k=0}^{b-n} a_k R_{b-n-k} = R_{b-n-1} + \delta_{0,b-n-1},$$

z (2.46) dostajemy

$$\begin{aligned} \chi_n(t) &= \frac{\sum_{k=1}^{b-1} (R_{b-k} - R_{b-1-k})v_k(t)}{\sum_{k=0}^{b-1} (R_{b-k} - R_{b-1-k})a_k} (R_{b-n-1} + \delta_{0,b-n-1}) \\ &\quad - \sum_{k=1}^{b-n} v_k(t) R_{b-n-k}, \end{aligned}$$

co kończy dowód Twierdzenia 2.3. ■

Obliczymy teraz przy pomocy Twierdzenia 2.3 rozkład pierwszego okresu przepełnienia, przy założeniu że czas obsługi jest wykładniczy, tzn. $F(x) = 1 - e^{-\mu x}$. W tym celu wyznaczmy najpierw wszystkie ciągi występujące w (2.37). Po pierwsze, mamy:

$$a_k = \int_0^\infty \frac{e^{-\lambda u} (\lambda u)^k}{k!} \mu e^{-\mu u} du = \frac{\mu \lambda^k}{(\mu + \lambda)^{k+1}}, \quad k \geq 0,$$

$$v_k(t) = \lambda \int_0^\infty e^{-\mu(u+t)} \frac{e^{-\lambda u} (\lambda u)^{k-1}}{(k-1)!} du = e^{-\mu t} \left(\frac{\lambda}{\mu + \lambda} \right)^k, \quad k \geq 1.$$

Jak widać, oba te ciągi są ciągami geometrycznymi z tym samym ilorazem $\frac{\lambda}{\mu+\lambda}$. Po drugie, stosując definicję potencjału R_k , otrzymujemy

$$R_k = \sum_{i=0}^k \left(\frac{\lambda}{\mu}\right)^i, \quad k \geq 1,$$

oraz

$$R_k - R_{k-1} = \left(\frac{\lambda}{\mu}\right)^k + \delta_{0,k-1}, \quad k \geq 1.$$

Następnie możemy obliczyć $x_1(t)$:

$$\begin{aligned} x_1(t) &= \frac{\sum_{k=1}^{b-1} (R_{b-k} - R_{b-1-k}) v_k(t)}{\sum_{k=0}^{b-1} (R_{b-k} - R_{b-1-k}) a_k} \\ &= \frac{e^{-\mu t} \sum_{k=1}^{b-1} \left(\frac{\lambda}{\mu+\lambda}\right)^k \left(\left(\frac{\lambda}{\mu}\right)^{b-k} + \delta_{0,b-k-1}\right)}{\frac{\mu}{\mu+\lambda} \sum_{k=0}^{b-1} \left(\frac{\lambda}{\mu+\lambda}\right)^k \left(\left(\frac{\lambda}{\mu}\right)^{b-k} + \delta_{0,b-k-1}\right)} \\ &= e^{-\mu t}. \end{aligned}$$

Wreszcie, biorąc pod uwagę, że

$$\sum_{k=1}^{b-n} v_k(t) R_{b-n-k} = \begin{cases} e^{-\mu t} \sum_{k=1}^{b-n-1} \left(\frac{\lambda}{\mu}\right)^k & \text{dla } b-n > 1, \\ 0 & \text{dla } b-n = 1. \end{cases}$$

z (2.37) dostajemy

$$H_n(t) = 1 - e^{-\mu t}$$

dla każdego $0 \leq n < b$. A zatem, w wypadku gdy czas obsługi ma rozkład wykładniczy, okres przepełnienia bufora też ma rozkład wykładniczy z takim samym parametrem μ i rozkład ten nie zależy od początkowej długości kolejki n . Wynik taki jest oczywiście związany z własnością braku pamięci rozkładu wykładniczego.

Zauważmy, że Twierdzenie 2.3 podaje bezpośrednio wzór na dystrybuantę rozkładu, bez użycia transformat, dlatego łatwo stosuje się je do obliczeń numerycznych. Łatwo możemy też wyznaczyć gęstość rozkładu β_1 , tzn.

$$h_n(t) = \frac{dH_n(t)}{dt}.$$

Twierdzenie 2.4. *Jeżeli rozkład czasu obsługi ma gęstość $f_g(t)$, to gęstość rozkładu pierwszego okresu przepełnienia bufora w systemie $M/G/1/b$ ma postać*

$$h_n(t) = - \frac{\sum_{k=1}^{b-1} (R_{b-k} - R_{b-1-k}) w_k(t)}{\sum_{k=0}^{b-1} (R_{b-k} - R_{b-1-k}) a_k} (R_{b-n-1} + \delta_{0,b-n-1}) + \sum_{k=1}^{b-n} w_k(t) R_{b-n-k}, \quad (2.47)$$

gdzie

$$w_k(t) = -\lambda \int_0^\infty f_g(u+t) \frac{e^{-\lambda u} (\lambda u)^{k-1}}{(k-1)!} du, \quad k = 1, 2, \dots$$

Dowód tego twierdzenia jest bardzo prosty, wystarczy jedynie zróżniczkować wzór (2.37). ■

Do tej pory zajmowaliśmy się tylko rozkładem β_1 , tzn. rozkładem pierwszego okresu przepełnienia bufora. Rozkład ten zależy od początkowej długości kolejki, czyli od $X(0)$. Jak się okazuje, stosując już otrzymane rezultaty, nietrudno znaleźć rozkłady kolejnych okresów przepełnienia, tzn. β_k , $k \geq 2$. W tym celu zauważmy, że zaraz po pierwszym okresie przepełnienia długość kolejki w systemie wynosi zawsze $b-1$. Tak więc $b-1$ jest początkową długością kolejki dla drugiego, trzeciego, itd. okresu przepełnienia. W związku z tym wszystkie β_k , $k \geq 2$ mają taki sam rozkład i jest on określony dystrybuantą $H_{b-1}(t)$ lub gęstością $h_{b-1}(t)$. Zwróćmy jeszcze uwagę, że dla $n = b-1$ wzór (2.37) redukuje się do swojej najprostszej postaci, tzn. $1 - x_1(t)$, gdzie $x_1(t)$ obliczamy, korzystając z (2.45).

Rozważania te możemy podsumować w formie następującego twierdzenia.

Twierdzenie 2.5. *W systemie $M/G/1/b$ rozkład β_k dla $k \geq 2$ nie zależy od k i jego dystrybuanta ma postać*

$$\mathbf{P}(\beta_k < t) = H_{b-1}(t) = 1 - \frac{\sum_{i=1}^{b-1} (R_{b-i} - R_{b-1-i}) v_i(t)}{\sum_{i=0}^{b-1} (R_{b-i} - R_{b-1-i}) a_i}. \quad (2.48)$$

Jeżeli rozkład czasu obsługi ma gęstość $f_g(t)$, to gęstość rozkładu β_k , $k \geq 2$ ma postać

$$h_{b-1}(t) = - \frac{\sum_{i=1}^{b-1} (R_{b-i} - R_{b-1-i}) w_i(t)}{\sum_{i=0}^{b-1} (R_{b-i} - R_{b-1-i}) a_i}. \quad (2.49)$$

■

Skoro rozkłady wszystkich β_k dla $k \geq 2$ są takie same, więc również rozkład graniczny przy $k \rightarrow \infty$ istnieje i jest taki jak w (2.48). W związku z tym można powiedzieć, że (2.48) jest również rozkładem okresu przepełnienia bufora dla stanu ustalonego.

2.2.5. Liczba strat w okresie przepełnienia bufora

Badany w poprzednim podrozdziale rozkład okresu przepełnienia bufora miał na celu opisanie statystycznej struktury kolejnych strat pakietów, tzn. tendencji do występowania strat w grupach. Mając do dyspozycji wzory opisujące rozkład okresu przepełnienia, możemy łatwo napisać wzory opisujące rozkład liczby kolejnych strat pakietów w okresie przepełnienia bufora.

Przez γ_k będziemy oznaczali liczbę pakietów utraconych w k -tym okresie przepełnienia. Wykorzystując wzór (2.2) i wyniki poprzedniego rozdziału, otrzymujemy natychmiast następujące twierdzenie.

Twierdzenie 2.6. *W systemie M/G/1/b rozkład liczby pakietów utraconych w pierwszym okresie przepełnienia bufora ma postać*

$$q_n(l) = \mathbf{P}(\gamma_1 = l | X(0) = n) = \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^l}{l!} dH_n(t), \quad (2.50)$$

zaś rozkład liczby pakietów utraconych w k -tym, $k \geq 2$, okresie przepełnienia ma postać

$$q(l) = \mathbf{P}(\gamma_k = l) = \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^l}{l!} dH_{b-1}(t) \quad (2.51)$$

i nie zależy od k . Dystrybuanty $H_n(t)$ i $H_{b-1}(t)$ mają postać odpowiednio jak w (2.37) i (2.48).

■

Jeżeli ponadto rozkład czasu obsługi ma gęstość $f_g(t)$, to całki Stieltjesa występujące w (2.50) i (2.51) redukują się do zwykłych całek i mamy wtedy

$$q_n(l) = \mathbf{P}(\gamma_1 = l | X(0) = n) = \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^l}{l!} h_n(t) dt, \quad (2.52)$$

$$q(l) = \mathbf{P}(\gamma_k = l) = \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^l}{l!} h_{b-1}(t) dt, \quad k \geq 2, \quad (2.53)$$

gdzie gęstości $h_n(t)$ i $h_{b-1}(t)$ mają postać odpowiednio jak w (2.47) i (2.49).

Korzystając z powyższych wzorów możemy bez trudu wyznaczać rozkłady liczby pakietów traconych kolejno w okresie przepełnienia bufora. Pojawia się jednak ciekawe pytanie: czy rozkład ten nie jest funkcją współczynnika strat pakietów? Gdyby tak było, jego wyznaczanie nie byłoby celowe, wystarczyłoby obliczenie $\Phi(b)$. Innymi słowy, pytanie brzmi: czy w ogóle jest możliwe, że w dwóch systemach współczynnik strat jest taki sam, np. 1%, ale w jednym systemie straty pojawiają się w dużych grupach, a w drugim są raczej równomiernie rozproszone w czasie?

Odpowiedź na to pytanie jest twierdząca. Ponadto, statystyczna struktura strat może być różna pomimo wielu wspólnych parametrów, nie tylko współczynnika strat. Aby to udowodnić, przeanalizujemy teraz kilka par prostych systemów. Każda para została dobrana tak, by współczynnik strat był taki sam w obydwu wypadkach.

Przykład 2.1. W pierwszym przykładzie będziemy analizować System A, w którym czas obsługi jest stały i równy 1 oraz System B z wykładniczym czasem obsługi. Ponadto założymy taki sam rozmiar bufora w obu systemach. Dokładna parametryzacja systemów A i B przedstawia się następująco:

System A:

$$\lambda = 0.2, \quad b = 5, \quad F(t) = \begin{cases} 0 & \text{dla } x \leq 1, \\ 1 & \text{dla } x > 1. \end{cases}$$

System B:

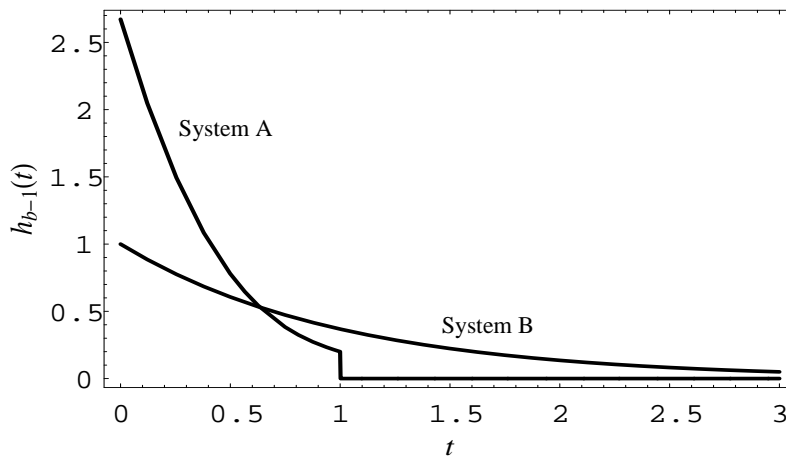
$$\lambda = 9.82711 \cdot 10^{-2}, \quad b = 5, \quad F(t) = 1 - e^{-t}.$$

Intensywność λ w Systemie B została tak dobrana, by w obu systemach współczynnik strat $\Phi(b)$ (obliczony wg (2.24)) był taki sam. Jego wspólna wartość wynosi

$$LR = \Phi(b) = 8.2643 \cdot 10^{-6}.$$

Gęstości rozkładów okresów przepełnienia dla systemów A i B są przedstawione na rys. 2.2, zaś prawdopodobieństwa utraty l kolejnych pakietów w okresie przepełnienia bufora zawiera tab. 2.1.

Jak widzimy na rys. 2.2, różnice między gęstościami okresów przepełnienia bufora są znaczne. Musi to oczywiście wpływać na statystyczną strukturę strat. Rzeczywiście, prawdopodobieństwo utraty trzech pakietów pod rząd w okresie przepełnienia jest w Systemie B sześciokrotnie większe i ta różnica pogłębia się, gdy l rośnie. Utrata pięciu pakietów pod rząd jest już pięćdziesięciokrotnie bardziej prawdopodobna w Systemie B niż w Systemie A.



Rys. 2.2. Gęstość rozkładu okresu przepełnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.1)

Fig. 2.2. Probability density of the buffer overflow period in systems A and B in steady state (Example 2.1)

l	$q(l)$ (System A)	$q(l)$ (System B)
1	$5.5024 \cdot 10^{-2}$	$8.1471 \cdot 10^{-2}$
2	$2.6790 \cdot 10^{-3}$	$7.2899 \cdot 10^{-3}$
3	$1.0987 \cdot 10^{-4}$	$6.5229 \cdot 10^{-4}$
4	$3.8484 \cdot 10^{-6}$	$5.8365 \cdot 10^{-5}$
5	$1.1684 \cdot 10^{-7}$	$5.2224 \cdot 10^{-6}$
6	$3.1173 \cdot 10^{-9}$	$4.6729 \cdot 10^{-7}$
7	$7.3984 \cdot 10^{-11}$	$4.1812 \cdot 10^{-8}$
8	$1.5787 \cdot 10^{-12}$	$3.7413 \cdot 10^{-9}$
9	$3.0565 \cdot 10^{-14}$	$3.3476 \cdot 10^{-10}$
10	$5.4120 \cdot 10^{-16}$	$2.9954 \cdot 10^{-11}$

Tab. 2.1. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym (Przykład 2.1)

Przykład 2.2. W poprzednim przykładzie rozmiar bufora był taki sam w obu systemach. W tym przykładzie wspólna będzie intensywność strumienia wejściowego, mianowicie:
System A:

$$\lambda = 0.5217976, \quad b = 10, \quad F(t) = \begin{cases} 0 & \text{dla } x \leq 1, \\ 1 & \text{dla } x > 1. \end{cases}$$

System B:

$$\lambda = 0.5217976, \quad b = 17, \quad F(t) = 1 - e^{-t}.$$

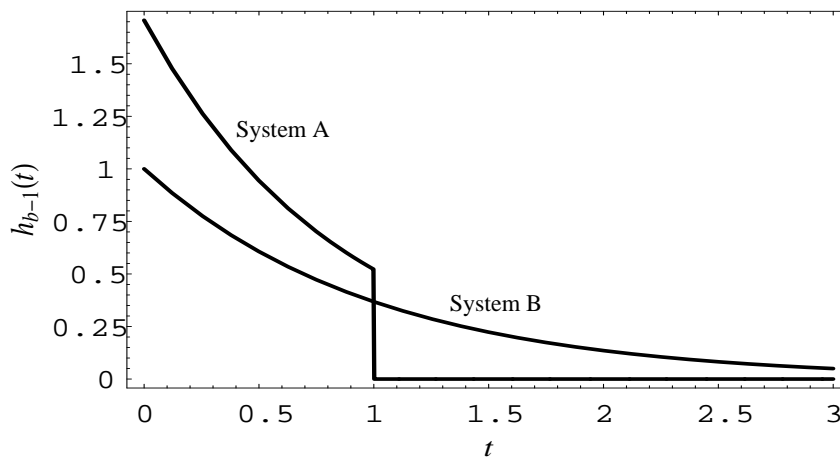
Wspólna wartość współczynnika strat wynosi teraz

$$LR = 7.5362 \cdot 10^{-6}.$$

Ponieważ średni czas obsługi w obu systemach jest taki sam, również obciążenie⁸ obydwu systemów jest takie samo i wynosi

$$\rho = 52.181\%.$$

Gęstości rozkładów okresów przepelnienia dla systemów A i B są przedstawione na rys. 2.3, zaś prawdopodobieństwa utraty l kolejnych pakietów w okresie przepelnienia bufora zawiera tab. 2.2.



Rys. 2.3. Gęstość rozkładu okresu przepelnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.2)

Fig. 2.3. Probability density of the buffer overflow period in systems A and B in steady state (Example 2.2)

Podobnie jak w poprzednim przykładzie, mamy tu do czynienia ze znaczną różnicą w strukturze strat kolejnych pakietów.

Przykład 2.3. W tym przykładzie pokażemy, że statystyczna struktura strat może być różna w dwóch systemach, pomimo że rozkład czasu obsługi jest identyczny. Przykładowa parametryzacja tych systemów może być następująca.

⁸Przez obciążenie w tej pracy rozumiemy zawsze obciążenie oferowane, tzn. iloczyn intensywności strumienia wejściowego i średniego czasu obsługi. W teorii ruchu iloczyn taki nazywany jest też *ruchem oferowanym*, w odróżnieniu od *ruchu załatwianego*, tzn. iloczynu intensywności strumienia wyjściowego i średniego czasu obsługi.

l	$q(l)$ (System A)	$q(l)$ (System B)
1	$0.1556 \cdot 10^{-1}$	$2.2531 \cdot 10^{-1}$
2	$2.2858 \cdot 10^{-2}$	$7.7256 \cdot 10^{-2}$
3	$2.6918 \cdot 10^{-3}$	$2.6490 \cdot 10^{-2}$
4	$2.6248 \cdot 10^{-4}$	$9.0829 \cdot 10^{-3}$
5	$2.1757 \cdot 10^{-5}$	$3.1144 \cdot 10^{-3}$
6	$1.5651 \cdot 10^{-6}$	$1.0679 \cdot 10^{-3}$
7	$9.9321 \cdot 10^{-8}$	$3.6615 \cdot 10^{-4}$
8	$5.6347 \cdot 10^{-9}$	$1.2555 \cdot 10^{-4}$
9	$2.8888 \cdot 10^{-10}$	$4.3048 \cdot 10^{-5}$
10	$1.3506 \cdot 10^{-11}$	$1.4760 \cdot 10^{-5}$

Tab. 2.2. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym (Przykład 2.2)

System A:

$$\lambda = 0.2, \quad b = 5, \quad F(t) = \begin{cases} 0 & \text{dla } x \leq 1, \\ 1 & \text{dla } x > 1. \end{cases}$$

System B:

$$\lambda = 0.9085966, \quad b = 50, \quad F(t) = \begin{cases} 0 & \text{dla } x \leq 1, \\ 1 & \text{dla } x > 1. \end{cases}$$

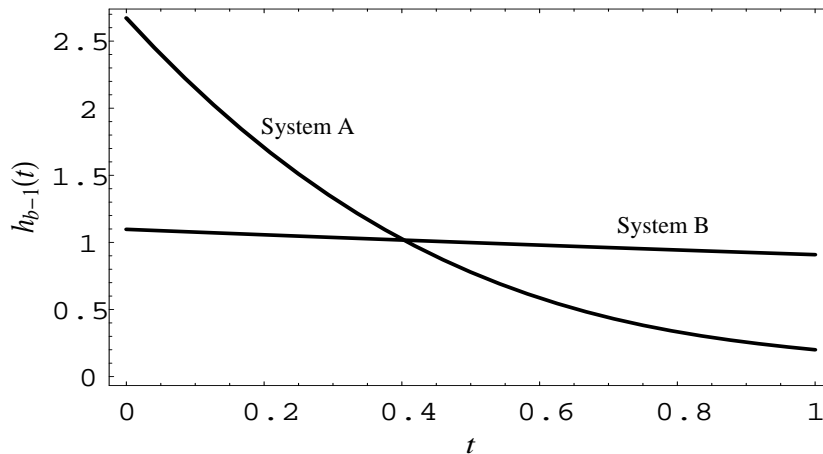
Tak więc w obydwu systemach mamy do czynienia ze stałym czasem obsługi równym 1. Wspólna wartość współczynnika strat wynosi teraz

$$LR = 8.2643 \cdot 10^{-6}.$$

Wykresy gęstości $h_{b-1}(t)$ w obydwu systemach przedstawia rys. 2.4, zaś prawdopodobieństwa $q(l)$ przedstawia tab. 2.3. Jak widzimy, różnice w strukturze strat pakietów są jeszcze większe niż w poprzednich wypadkach – prawdopodobieństwo utraty trzech pakietów pod rząd w okresie przepełnienia różni się w obu systemach o dwa rzędy wielkości.

Przykłady 2.1, 2.2 i 2.3 pokazują, że nawet w wypadku prostych systemów mogą występować duże różnice w statystycznej strukturze strat kolejnych pakietów i to pomimo jednakowych wartości współczynnika strat, obciążenia, intensywności strumienia wejściowego czy czasu obsługi. Można pokazać też (zob. [55]), że struktura strat może być różna w systemach z jednakowym buforem i jednakowym współczynnikiem strat.

Różnice te mogą być jeszcze bardziej istotne w wypadku bardziej złożonych systemów, które będą omawiane w następnych rozdziałach.



Rys. 2.4. Gęstość rozkładu okresu przepełnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.3)

Fig. 2.4. Probability density of the buffer overflow period in systems A and B in steady state (Example 2.3)

2.3. Kolejowanie złożonego procesu Poissona

Złożony proces Poissona tym różni się od (zwykłego) procesu Poissona, że zdarzenia pojawiają się nie pojedynczo, lecz w grupach. Rozmiary kolejnych grup zdarzeń będziemy oznaczać $\alpha_1, \alpha_2, \dots$. Tak jak poprzednio, czas pomiędzy pojawieniem się kolejnych grup zdarzeń jest wykładniczy i ma gęstość

$$g(x) = \lambda e^{-\lambda x}, \quad x > 0. \quad (2.54)$$

Wszystkie α_k są niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa opisanym ciągiem p_i , gdzie

$$p_i = \mathbf{P}(\alpha_k = i)$$

oznacza prawdopodobieństwo pojawienia się grupy o rozmiarze i .

Złożony proces Poissona również cechuje się brakiem pamięci. Ponadto, w dowodach będziemy korzystać z następujących jego własności.

l	$q(l)$ (System A)	$q(l)$ (System B)
1	$5.5024 \cdot 10^{-2}$	$2.4840 \cdot 10^{-1}$
2	$2.6790 \cdot 10^{-3}$	$6.7906 \cdot 10^{-2}$
3	$1.0987 \cdot 10^{-4}$	$1.4501 \cdot 10^{-2}$
4	$3.8484 \cdot 10^{-6}$	$2.5294 \cdot 10^{-3}$
5	$1.1684 \cdot 10^{-7}$	$3.7207 \cdot 10^{-4}$
6	$3.1173 \cdot 10^{-9}$	$4.7261 \cdot 10^{-5}$
7	$7.3984 \cdot 10^{-11}$	$5.2787 \cdot 10^{-6}$
8	$1.5787 \cdot 10^{-12}$	$5.2588 \cdot 10^{-7}$
9	$3.0565 \cdot 10^{-14}$	$4.7267 \cdot 10^{-8}$
10	$5.4120 \cdot 10^{-16}$	$3.8694 \cdot 10^{-9}$

Tab. 2.3. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepelnienia bufora w stanie ustalonym (Przykład 2.3)

- **Rozkład prawdopodobieństwa** $N(t)$, tzn. liczby zdarzeń pojawiających się w dowolnym przedziale czasu o długości t , ma postać

$$\begin{aligned}
 m_k(t) &= \mathbf{P}(N(t) = k) = \sum_{i=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^i}{i!} \mathbf{P}\left(\sum_{j=1}^i \alpha_j = k\right) \\
 &= \sum_{i=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^i}{i!} p_k^{i*}, \tag{2.55}
 \end{aligned}$$

gdzie p_k^{i*} oznacza i -krotny splot ciągu p_k , tzn.

$$p_k^{0*} = \delta_{k0}, \quad p_k^{(i+1)*} = \sum_{j=0}^k p_j p_{k-j}^{i*}, \quad i = 0, 1, 2, \dots$$

- **Funkcja tworząca** dla rozkładu prawdopodobieństwa $N(t)$ ma postać

$$\sum_{k=0}^{\infty} z^k \mathbf{P}(N(t) = k) = e^{-\lambda t(1-p(z))},$$

gdzie

$$p(z) = \sum_{i=0}^{\infty} z^i p_i.$$

- **Czas pojawienia się k -tego zdarzenia**, licząc od chwili 0, ma następującą gęstość rozkładu prawdopodobieństwa

$$s_k(t) = \sum_{i=1}^{\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^{i-1}}{(i-1)!} c_{k,i}, \quad (2.56)$$

gdzie

$$c_{k,i} = \mathbf{P}\left(\sum_{j=1}^{i-1} \alpha_j < k \leq \sum_{j=1}^i \alpha_j\right) = \sum_{m=0}^{k-1} (p_m^{(i-1)*} - p_m^{i*}).$$

Innymi słowy, $c_{k,i}$ oznacza prawdopodobieństwo, że k -te zdarzenie pojawiło się w i -tej grupie zdarzeń.

Złożony proces Poissona, ze względu na swoją bardziej skomplikowaną strukturę, ma dużo większe możliwości modelowania ruchu w sieciach pakietowych niż proces Poissona. Modelowanie to może się odbywać na dwóch poziomach:

1. Zdarzenia mogą reprezentować np. pakiety IP czy komórki ATM w całości, tzn. jedno zdarzenie = jeden pakiet. Wówczas grupowa struktura może zostać wykorzystana do modelowania spiętrzeń (*ang.* burstiness) w ruchu sieciowym. Przykładowo, ustawienie dużej wartości p_1 (np. $p_1 = 0.99$) powoduje, że przez większą część czasu mamy do czynienia ze zwykłym procesem Poissona, czasami jednak będzie się pojawiała większa grupa pakietów jednocześnie (burst).
2. Zdarzenia mogą reprezentować nie pakiety, lecz pojedyncze bajty. Wówczas cała grupa zdarzeń reprezentuje jeden pakiet o rozmiarze równym rozmiarowi grupy, np. $\alpha_i = 1500$ oznacza pojawienie się i -tego pakietu o rozmiarze 1500 bajtów. Podejście to umożliwia uwzględnienie rozkładu rozmiaru pakietów IP, co (jak udowodniono w [182]) jest konieczne przy bardzo precyzyjnym wyznaczaniu charakterystyk kolejowania.

Analiza kolejowania złożonego procesu Poissona nie odbiega bardzo istotnie od tej w wypadku prostego procesu Poissona. Pewne utrudnienia są związane z grupową strukturą strumienia wejściowego – można jednak je przewyciężyć, stosując techniki oparte na funkcjach tworzących. Zostanie to pokazane na przykładzie okresu przepełnienia bufora oraz czasu do przepełnienia bufora. Pozostałe charakterystyki, takie jak rozkład długości kolejki czy opóźnienia, można wyznaczyć w podobny sposób.

2.3.1. Okres przepelnienia bufora

Rozpocznijmy od wyznaczenia dystrybuanty rozkładu pierwszego okresu przepelnienia bufora, tzn.

$$H_n(t) = \mathbf{P}(\beta_1 < t | X(0) = n).$$

Twierdzenie 2.7. *W systemie $M^X/G/1/b$ rozkład pierwszego okresu przepelnienia bufora ma postać*

$$H_n(t) = 1 - (1 - H_{b-1}(t))(R_{b-n-1} + \delta_{0,b-n-1}) + \sum_{k=1}^{b-n} v_k(t)R_{b-n-k}, \quad 0 \leq n \leq b-2, \quad (2.57)$$

gdzie

$$H_{b-1}(t) = 1 - \left[\sum_{k=1}^b \left(\sum_{i=1}^k v_i(t)R_{k-i}p_{b-k} - v_k(t)R_{b-k} \right) - (1 - F(t)) \left(1 - \sum_{k=0}^{b-1} p_k \right) \right] / \left[\sum_{k=1}^b R_{k-1}p_{b-k} + p_{b-1} - R_{b-1} - \delta_{b,1} \right], \quad (2.58)$$

zaś ciągi R_k i $v_k(t)$ są określone przez następujące funkcje tworzące:

$$R(z) = \sum_{k=0}^{\infty} z^k R_k = \frac{z}{f(\lambda(1-p(z)) - z)}, \quad (2.59)$$

$$v(z, t) = \sum_{k=1}^{\infty} z^k v_k(t) = \frac{\lambda z(1-p(z))}{1-z} \int_0^{\infty} (1 - F(u+t)) e^{-\lambda u(1-p(z))} du. \quad (2.60)$$

Dowód Twierdzenia 2.7. Zakładamy najpierw, że $X(0) \in [1, b-1]$. Stosując wzór na prawdopodobieństwo całkowite, otrzymujemy

$$\begin{aligned} \mathbf{P}(\beta_1 > t | X(0) = n) &= \sum_{k=0}^{b-n-1} \int_0^{\infty} \mathbf{P}(\beta_1 > t | X(0) = n+k-1) m_k(u) dF(u) \\ &\quad + \int_0^{\infty} dF(u) \int_0^u I(u-v > t) s_{b-n}(v) dv, \end{aligned} \quad (2.61)$$

gdzie $m_k(u)$ i $s_k(v)$ są określone odpowiednio przez (2.55) i (2.56).

Pierwsza część (2.61) odpowiada przypadkowi, w którym do pierwszego czasu ukończenia obsługi u nie następuje przepełnienie bufora, tzn. liczba zdarzeń w przedziale $(0, u]$ jest mniejsza niż $b - n$. Druga część (2.61) odpowiada przypadkowi, w którym następuje przepełnienie bufora w chwili v , $v < u$. Wtedy $\beta_1 = u - v$ i $\mathbf{P}(\beta_1 > t) = I(u - v > t)$.

Założmy teraz, że $X(0) = 0$. Stosując wzór na prawdopodobieństwo całkowite ze względu na rozmiar pierwszej grupy zdarzeń, otrzymujemy

$$\begin{aligned} \mathbf{P}(\beta_1 > t | X(0) = 0) &= \sum_{k=0}^{\infty} \mathbf{P}(\beta_1 > t | X(0) = 0, \alpha_1 = k) p_k \\ &= \sum_{k=0}^{b-1} \mathbf{P}(\beta_1 > t | X(0) = k) p_k + (1 - F(t)) \sum_{k=b}^{\infty} p_k. \end{aligned} \quad (2.62)$$

Pierwsza część (2.62) odpowiada zdarzeniu $\alpha_1 < b$, zaś druga zdarzeniu $\alpha_1 \geq b$. W tym drugim wypadku zachodzi, oczywiście, $\mathbf{P}(\beta_1 > t) = 1 - F(t)$.

Wprowadźmy teraz oznaczenia:

$$\begin{aligned} \chi_n(t) &= \mathbf{P}(\beta_1 > t | X(0) = n) = 1 - H_n(t), \\ a_k &= \sum_{i=0}^{\infty} p_k^{i*} \int_0^{\infty} \frac{e^{-\lambda u} (\lambda u)^i}{i!} dF(u), \end{aligned} \quad (2.63)$$

$$v_k(t) = \lambda \sum_{i=1}^{\infty} c_{k,i} \int_0^{\infty} (1 - F(u+t)) \frac{e^{-\lambda u} (\lambda u)^{i-1}}{(i-1)!} du. \quad (2.64)$$

Wtedy (2.61) i (2.62) możemy zapisać jako

$$\chi_n(t) = \sum_{k=0}^{b-n-1} \chi_{n+k-1}(t) a_k + v_{b-n}(t), \quad 0 < n < b, \quad (2.65)$$

$$\chi_0(t) = \sum_{k=0}^{b-1} \chi_k(t) p_k + (1 - F(t)) \sum_{k=b}^{\infty} p_k, \quad (2.66)$$

przy czym $v_{b-n}(t)$ w (2.65) otrzymujemy, całkując przez części ostatni człon (2.61). Układ równań (2.65) ma taką samą postać jak (2.40), więc może zostać rozwiązany w ten sam sposób. Ostatecznie otrzymujemy

$$\begin{aligned} \chi_n(t) &= \chi_{b-1}(t) \sum_{k=0}^{b-n} a_k R_{b-n-k} - \sum_{k=1}^{b-n} v_k(t) R_{b-n-k} \\ &= \chi_{b-1}(t) (R_{b-n-1} + \delta_{0,b-n-1}) - \sum_{k=1}^{b-n} v_k(t) R_{b-n-k}, \end{aligned} \quad (2.67)$$

gdzie R_k jest potencjałem (1.12) dla ciągu a_k określonego w (2.63). Z uwagi na skomplikowaną postać a_k taka postać ciągu R_k jest niewygodna do celów numerycznych. Zamiast niej w tezie twierdzenia pojawił się zapis R_k wykorzystujący funkcję tworzącą. Istotnie, łatwo pokazać, że funkcja tworząca potencjału dla ciągu a_k z (2.63), tzn.

$$R(z) = \sum_{k=0}^{\infty} z^k R_k, \quad (2.68)$$

ma postać (2.59). Podobnie, funkcja tworząca dla ciągu $v_k(t)$ zdefiniowanego w (2.64) ma postać (2.60).

Bazując na (2.67), możemy łatwo skończyć dowód. Wzór dla $\chi_{b-1}(t)$ znajdziemy przy pomocy (2.67) i (2.66). Podstawiając $H_{b-1}(t) = 1 - \chi_{b-1}(t)$, dostajemy następnie (2.58). Ostatecznie, (2.57) wynika z (2.67). ■

W Twierdzeniu 2.7, inaczej niż w poprzednich podrozdziałach, ciągi R_k i $v_k(t)$ są określone przez swoje funkcje tworzące. W związku z tym, by można je było wykorzystać w praktyce, konieczne jest odwracanie funkcji tworzących. Jedną z metod służących do tego celu jest opisana w podrozdziale 5.1.3.

Różniczkując (2.57), (2.58) i (2.60), możemy łatwo otrzymać gęstości pierwszego okresu przepełnienia bufora.

Twierdzenie 2.8. *Jeżeli rozkład czasu obsługi ma gęstość $f_g(t)$, to gęstość rozkładu pierwszego okresu przepełnienia bufora w systemie $M^X/G/1/b$ ma postać*

$$h_n(t) = h_{b-1}(t)(R_{b-n-1} + \delta_{0,b-n-1}) + \sum_{k=1}^{b-n} w_k(t)R_{b-n-k}, \quad 0 \leq n \leq b-2, \quad (2.69)$$

gdzie

$$h_{b-1}(t) = -\frac{\sum_{k=1}^b \left(\sum_{i=1}^k w_i(t)R_{k-i}p_{b-k} - w_k(t)R_{b-k} \right) + f_g(t)(1 - \sum_{k=0}^{b-1} p_k)}{\sum_{k=1}^b R_{k-1}p_{b-k} + p_{b-1} - R_{b-1} - \delta_{b,1}}, \quad (2.70)$$

ciąg $w_k(t)$ ma funkcję tworzącą

$$w(z, t) = \sum_{k=1}^{\infty} z^k w_k(t) = -\frac{\lambda z(1-p(z))}{1-z} \int_0^{\infty} f_g(u+t) e^{-\lambda u(1-p(z))} du, \quad (2.71)$$

zaś R_k jest określone w (2.59).

W ten sposób możemy zakończyć analizę β_1 , tzn. pierwszego okresu przepełnienia bufora. Dla każdego kolejnego okresu przepełnienia bufora początkową długością kolejki jest $b - 1$, mamy zatem:

Twierdzenie 2.9. *W systemie $M^X/G/1/b$ rozkład β_k dla $k \geq 2$ nie zależy od k i jego dystrybuanta ma postać (2.58). Jeżeli rozkład czasu obsługi ma gęstość $f_g(t)$, to gęstość rozkładu β_k , $k \geq 2$ ma postać (2.70). Ponadto (2.58), jako rozkład graniczny dla $k \rightarrow \infty$, jest również rozkładem okresu przepełnienia bufora dla stanu ustalonego.*

Ponieważ (2.58) i (2.70) mają dość skomplikowaną postać i ponadto wymagają odwracania funkcji tworzących, pojawia się pytanie, czy nie można znaleźć prostszych wzorów przybliżonych, które można by było stosować w pewnych sytuacjach.

Okazuje się, że można znaleźć wzory aproksymujące (2.58) i (2.70), w wypadku gdy rozmiar bufora jest duży. Rozpocznijmy od wyznaczenia

$$H(t) = \lim_{b \rightarrow \infty} H_{b-1}(t).$$

Postać wzoru na $H(t)$ zależęć będzie od obciążenia (oferowanego) systemu

$$\rho = \lambda \alpha S, \quad (2.72)$$

gdzie α oznacza średni rozmiar grupy w złożonym procesie Poissona, zaś S to średni czas obsługi. Wykorzystując własności transformat, obciążenie możemy też obliczyć jako

$$\rho = -\lambda p'(1) f'(0). \quad (2.73)$$

Twierdzenie 2.10. *Rozkład okresu przepełnienia bufora systemie $M^X/G/1/b$ dla dużych b ma następującą postać.*

1. Jeżeli $\rho = 1$, to

$$H(t) = 1 - \lambda p'(1) \int_0^\infty (1 - F(u + t)) du. \quad (2.74)$$

2. Jeżeli $\rho \neq 1$, to

$$H(t) = 1 - \lambda \frac{1 - p(z_0)}{1 - z_0} \int_0^\infty (1 - F(u + t)) e^{-\lambda u(1 - p(z_0))} du, \quad (2.75)$$

gdzie z_0 jest dodatnim i różnym od 1 rozwiązaniem równania

$$f(\lambda(1 - p(z))) = z.$$

Zanim udowodnimy to twierdzenie, zauważmy najpierw, że jeżeli $\rho > 1$, to z_0 musi istnieć i $0 < z_0 < 1$. Wynika to z dwóch faktów. Po pierwsze, pochodna $f(\lambda(1 - p(z)))$ w $z = 1$ jest większa od 1, gdyż

$$\left. \frac{df(\lambda(1 - p(z)))}{dz} \right|_{z=1} = -\lambda p'(1) f'(0) = \rho > 1.$$

Po drugie, mamy

$$f(\lambda(1 - p(0))) > 0.$$

A zatem wykres funkcji

$$y = f(\lambda(1 - p(z)))$$

leży poniżej $y = z$ dla $z = 1-$ i powyżej $y = z$ dla $z = 0+$. Tak więc gdzieś pomiędzy 0 i 1 znajduje się ich punkt przecięcia, czyli z_0 .

Jeżeli $\rho < 1$, to pierwiastek z_0 może nie istnieć i (2.75) wówczas nie zachodzi. Tak może się zdarzyć na przykład wtedy, gdy rozkład czasu obsługi ma ciężki ogon. Jednak, jak pokazuje praktyka, w większości wypadków z_0 istnieje i wtedy dla $\rho < 1$ mamy $z_0 > 1$.

Dowód Twierdzenia 2.10. Załóżmy najpierw, że $\rho > 1$, i rozważmy zamiast systemu $M^X/G/1/b$ system $G/M^X/1/\infty$, tzn. z rozkładem czasu pomiędzy pakietami $F(t)$, obsługą według złożonego procesu Poissona oraz bez ograniczenia na długość kolejki (nieskończony bufor). Oczywiście, w systemie $G/M^X/1/\infty$ będzie wtedy $\rho < 1$, co zapobiega nieograniczonemu wzrostowi długości kolejki.

Zauważmy, że rozkład β_1 dla $X(0) = b - 1$ w systemie $M^X/G/1/b$ z dużym b jest taki sam jak rozkład pierwszego okresu bezczynności⁹ w systemie $G/M^X/1/\infty$ przy warunku $X(0) = 1$.

Tak więc w tej części dowodu zajmijmy się wyznaczeniem rozkładu okresu bezczynności w systemie $G/M^X/1/\infty$. Jego długość będziemy oznaczać przez ξ .

Wykorzystując wzór na prawdopodobieństwo całkowite ze względu na czas pojawienia się pierwszego pakietu, otrzymujemy następujący układ równań:

$$\begin{aligned} \mathbf{P}(\xi > t | X(0) = n) &= \sum_{k=0}^{n-1} \int_0^\infty \mathbf{P}(\xi > t | X(0) = n - k + 1) \int_0^\infty m_k(u) dF(u) \\ &\quad + \int_0^\infty dF(u) \int_0^u s_n(v) I(u - v > t) dv, \quad n \geq 1. \end{aligned} \quad (2.76)$$

⁹Ang. idle time – czas od osiągnięcia przez długość kolejki poziomu 0 (pusty system) do wpływu pierwszego pakietu.

Oznaczając

$$\begin{aligned}\chi_n(t) &= \mathbf{P}(\xi > t | X(0) = n), \\ \psi_n(t) &= a_n \chi_1(t) - v_n(t),\end{aligned}$$

z (2.76) otrzymujemy, że

$$\sum_{k=-1}^{n-1} a_{k+1} \chi_{n-k}(t) - \phi_n(t) = \psi_n(t). \quad (2.77)$$

Na mocy Twierdzenia 1.4 mamy:

$$\chi_n(t) = cR_n + \sum_{k=1}^n \psi_k(t)R_{n-k}, \quad n \geq 1. \quad (2.78)$$

Obliczając $c = \chi_1(t)a_0$ oraz podstawiając

$$\sum_{k=0}^n a_k R_{n-k} = R_{n-1} + \delta_{0,n-1},$$

dostajemy następnie

$$\chi_n(t) = \chi_1(t)(R_{n-1} + \delta_{0,n-1}) - \sum_{k=1}^n v_k(t)R_{n-k}. \quad (2.79)$$

Pozostało nam teraz wyznaczenie $\chi_1(t) = 1 - H(t)$.

W poprzednich dowodach wykorzystywaliśmy do tego celu warunki brzegowe dla $n = 0$. Tutaj nie mamy analogicznego warunku, musimy więc zastosować nieco bardziej wyrafinowane techniki.

Rozważmy następującą funkcję:

$$B(z, t) = \sum_{n=2}^{\infty} z^n \chi_n(t).$$

Jak łatwo sprawdzić, jest ona dobrze określona co najmniej dla $|z| < 1$. Dzięki (2.59) mamy:

$$B(z, t) = \frac{z(\chi_1(t)z - \sum_{i=1}^{\infty} z^i v_i(t))}{f(\lambda(1-p(z))) - z}. \quad (2.80)$$

Tak jak pokazaliśmy przed dowodem twierdzenia, z założenia

$$-\lambda p'(1) f'(0) > 1$$

wiemy, że mianownik we wzorze (2.80) ma jeden pierwiastek w przedziale $(0, 1)$. Oznaczmy ten pierwiastek przez z_0 . Teraz, podstawiając z_0 do licznika w (2.80), dostajemy

$$\begin{aligned}\chi_1(t) &= \sum_{i=0}^{\infty} z_0^i v_{i+1}(t) \\ &= \lambda \frac{1-p(z_0)}{1-z_0} \int_0^{\infty} (1-F(u+t)) e^{-\lambda u(1-p(z_0))} du,\end{aligned}$$

co kończy dowód w przypadku $\rho > 1$.

Wracamy teraz do systemu $M^X/G/1/b$. Dla $\rho < 1$ ciąg R_k ma prostą postać asymptotyczną dla dużych k , a mianowicie:

$$R_k = \frac{1}{1-\rho} - \frac{a}{z_0^k} + o(z_0^{-k}),$$

gdzie $a = -1/[\lambda p'(z_0) f'(\lambda(1-p(z_0))) + 1]$ (zob. [58]).

Wykorzystując ten fakt, wzór (2.75) otrzymamy, obliczając granicę przy $b \rightarrow \infty$ w (2.58). W końcu, (2.74) dla $\rho = 1$ można łatwo otrzymać, przechodząc do granicy $z_0 \rightarrow 1$ w (2.75). ■

Twierdzenie 2.10 ma duże znaczenie praktyczne z dwóch powodów. Po pierwsze, rozkład graniczny ma o wiele prostszą formę niż jego dokładny odpowiednik z Twierdzenia 2.7. Po drugie, eksperymenty numeryczne pokazują, że zwykle $H_{b-1}(t)$ bardzo szybko zbiega do $H(t)$, w związku z tym nawet dla małych b , rzędu kilkunastu, można stosować wzory (2.75) i (2.74).

Różniczkując (2.75) i (2.74), możemy otrzymać wzory na gęstości okresu przepełnienia dla dużych buforów.

Twierdzenie 2.11. *Jeżeli czas obsługi ma gęstość $f_g(t)$, to gęstość rozkładu okresu przepłynięcia bufora w systemie $M^X/G/1/b$ dla dużych b ma następującą postać.*

1. Jeżeli $\rho = 1$, to

$$h(t) = \lambda p'(1) \int_0^{\infty} f_g(u+t) du. \quad (2.81)$$

2. Jeżeli $\rho \neq 1$, to

$$h(t) = \lambda \frac{1-p(z_0)}{1-z_0} \int_0^{\infty} f_g(u+t) e^{-\lambda u(1-p(z_0))} du, \quad (2.82)$$

gdzie z_0 jest określone tak jak w Twierdzeniu 2.10.

W praktyce często mamy do czynienia z sytuacją, w której czas obsługi nie jest zmienną losową i przyjmuje stałą wartość d .

Wówczas powyższe wzory redukują się do prostszej postaci, w szczególności:

$$\begin{aligned} R(z) &= \frac{z}{e^{-\lambda d(1-p(z))} - z}, \\ v(z, t) &= I(t < d) \frac{z}{z-1} (e^{-\lambda(d-t)(1-p(z))} - 1), \\ H(t) &= 1 - I(t < d) \frac{e^{-\lambda(d-t)(1-p(z_0))} - 1}{z_0 - 1} \quad \text{dla } \rho \neq 1, \\ H(t) &= 1 - I(t < d) \lambda p'(1)(d-t) \quad \text{dla } \rho = 1. \end{aligned}$$

2.3.2. Czas do przepełnienia bufora

Czas do przepełnienia bufora jest zdefiniowany jako

$$\tau_n = \inf\{t > 0 : X(t) = b | X(0) = n\}$$

i zależy on od początkowej długości kolejki.

Rozkład czasu do przepełnienia bufora wyznaczymy w postaci transformaty

$$l_n(s) = \int_0^\infty e^{-st} \mathbf{P}(\tau_n > t) dt.$$

Twierdzenie 2.12. *Transformata rozkładu czasu do przepełnienia bufora w systemie $M^X/G/1/b$ ma postać*

$$l_n(s) = \frac{v_b(s)}{w_b(s)} \sum_{k=0}^{b-n} a_k(s) R_{b-n-k}(s) - \sum_{k=1}^{b-n} \tilde{d}_k(s) R_{b-n-k}(s), \quad (2.83)$$

gdzie

$$a_k(s) = \int_0^\infty e^{-st} m_k(t) dF(t), \quad (2.84)$$

$$\tilde{d}_k(s) = \sum_{i=0}^{k-1} \int_0^\infty e^{-st} m_i(t) (1 - F(t)) dt, \quad (2.85)$$

$$v_b(s) = \frac{\lambda}{s + \lambda} \sum_{k=1}^b p_{b-k} \sum_{i=1}^k \tilde{d}_i(s) R_{k-i}(s) - \sum_{k=1}^b \tilde{d}_k(s) R_{b-k}(s) - \frac{1}{s + \lambda}, \quad (2.86)$$

$$w_b(s) = \frac{\lambda}{s + \lambda} \sum_{k=0}^b p_{b-k} \sum_{i=0}^k a_i(s) R_{k-i}(s) - \sum_{k=0}^b a_k(s) R_{b-k}(s), \quad (2.87)$$

$m_k(t)$ jest określone wzorem (2.55) oraz $R_k(s)$ jest potencjałem dla ciągu $a_k(s)$.

Dowód Twierdzenia 2.12. Ze wzoru na prawdopodobieństwo całkowite dla pierwszej chwili ukończenia obsługi dostajemy dla $X(0) > 0$, że

$$\begin{aligned} \mathbf{P}(\tau_n > t) &= \sum_{k=0}^{b-n-1} \int_0^t \mathbf{P}(\tau_{n+k-1} > t-u) m_k(u) dF(u) \\ &+ (1 - F(t)) \sum_{k=0}^{b-n-1} m_k(t), \quad 0 < n < b. \end{aligned} \quad (2.88)$$

Dla $X(0) = 0$ mamy z kolei

$$\mathbf{P}(\tau_0 > t) = \sum_{k=0}^{b-1} p_k \int_0^t \mathbf{P}(\tau_k > t-u) \lambda e^{-\lambda u} du + e^{-\lambda t}. \quad (2.89)$$

Zastosowanie transformaty Laplace'a do (2.88) i (2.89) daje:

$$l_n(s) = \sum_{k=0}^{b-n-1} l_{n+k-1}(s) a_k(s) + \tilde{d}_{b-n}(s), \quad 0 < n < b, \quad (2.90)$$

$$l_0(s) = \frac{\lambda}{s + \lambda} \sum_{k=0}^{b-1} p_k l_k(s) + \frac{1}{s + \lambda}, \quad (2.91)$$

Podstawiając

$$l_n(s) = u_{b-n}(s),$$

otrzymujemy

$$\sum_{k=-1}^{n-1} u_{n-k}(s) a_{k+1}(s) - u_n(s) = \psi_n(s), \quad 0 < n < b, \quad (2.92)$$

$$u_b(s) = \frac{\lambda}{s + \lambda} \sum_{k=0}^{b-1} p_k u_{b-k}(s) + \frac{1}{s + \lambda}, \quad (2.93)$$

gdzie

$$\psi_n(s) = u_1(s)a_n(s) - \tilde{d}_n(s).$$

Z Twierdzenia 1.4 wynika, że rozwiązanie układu (2.92) ma postać

$$u_n(s) = c(s)R_n(s) + \sum_{k=1}^n \psi_k(s)R_{n-k}(s). \quad (2.94)$$

Wstawiając $n = 1$ do (2.94), mamy

$$c(s) = u_1(s)/R_1(s),$$

natomiast ze wzorów (2.93) i (2.94) otrzymujemy

$$u_1(s) = \frac{v_b(s)}{w_b(s)}.$$

Wreszcie, zapisując (2.94) w postaci

$$u_n(s) = u_1(s) \sum_{k=0}^n a_k(s)R_{n-k}(s) - \sum_{k=1}^n \tilde{d}_k(s)R_{n-k}(s),$$

kończymy dowód twierdzenia. ■

Aby obliczyć $a_k(s)$ i $\tilde{d}_k(s)$, możemy ponownie postawić się funkcjami tworzącymi. Jak łatwo sprawdzić, mamy:

$$a(z, s) = \sum_{k=0}^{\infty} z^k a_k(s) = f(s + \lambda(1 - p(z))),$$

$$\tilde{d}(z, s) = \sum_{k=1}^{\infty} z^k \tilde{d}_k(s) = \frac{z[1 - f(s + \lambda(1 - p(z)))]}{(1 - z)(s + \lambda(1 - p(z)))}.$$

Wystarczy teraz zastosować algorytm odwracania funkcji tworzących zaprezentowany w podrozdziale 5.1.3.

2.4. Nota bibliograficzna do Rozdziału 2

Charakterystyki kolejkowania prostych strumieni Poissona są wyznaczane w większości monografii poświęconych teorii kolejek, np. [117, 76, 74, 193, 90, 78], jednak najczęściej są to jedynie podstawowe charakterystyki (tzn. bez okresu przepełnienia, procesu strat pakietów czy czasu do przepełnienia bufora) i tylko dla stanu ustalonego. Zwykle stosowana jest klasyczna, oparta na złożonych łańcuchach Markowa metodologia, która w wypadku systemów ze skończonym buforem nie pozwala na zapisanie jawnych wzorów dla poszukiwanych charakterystyk.

Wyniki dla stanu nieustalonego są prezentowane rzadko, np. w [192, 75] można znaleźć wzory dla rozkładu długości kolejki w stanie nieustalonym, jednak tylko w wypadku nieograniczonego bufora.

W niektórych pracach, oprócz charakterystyk dla prostych strumieni Poissona, wyznaczane są też podstawowe charakterystyki w stanie ustalonym dla złożonych strumieni Poissona, np. [192, 193, 76, 74].

Do tej pory prowadzono niewiele badań nad okresem przepełnienia bufora. Kilka wcześniejszych wyników zawierają prace [88, 97, 31]. I tak, w [88] znajdziemy wzór dla średniej długości okresu przepełnienia w systemie z ogólnymi typami czasu pomiędzy wpływem pakietów i obsługi, w [97] zaproponowano proste przybliżenie dla okresu przepełnienia bufora w sytuacji, gdy czas obsługi jest stały, wreszcie w [31] wyprowadzone zostały wzory graniczne (duży rozmiar bufora) dla prostego strumienia Poissona.

Przedstawione w tym rozdziale wyniki pochodzą z następujących prac. Rozkład długości kolejki (podrozdział 2.2.1) został zbadany w [51], zaś rozkład opóźnienia (podrozdział 2.2.3) w [71]. Rozkład okresu przepełnienia bufora dla prostego strumienia Poissona (podrozdział 2.2.4) był analizowany w [63, 52]. W szczególności, w [63] pokazane zostały wzory dla stałego czasu obsługi, zaś w [52] wzory dla rozkładu obsługi w ogólnej postaci. Rozkład liczby kolejnych strat pakietów (podrozdział 2.2.5) został zbadany w [54, 55]. W szczególności, w [55] zostało pokazane, że rozkład liczby kolejnych strat pakietów w okresie przepełnienia bufora nie jest funkcją współczynnika strat pakietów (Przykłady 2.1, 2.2, 2.3).

Wzory opisujące okres przepełnienia bufora dla złożonego strumienia Poissona (podrozdział 2.3.1) zostały wyprowadzone w [53, 58, 59]. W szczególności, w [59] pokazane są wyniki dla stałego czasu obsługi, w [58] wyniki dla ogólnej postaci rozkładu czasu obsługi i twierdzenie graniczne dla dużych rozmiarów buforów, zaś [53] zawiera dalsze rezultaty graniczne.

Wreszcie wzór dla czasu do przepełnienia bufora (podrozdział 2.3.2) został udowodniony dla prostego strumienia Poissona i stałego czasu obsługi w [56] oraz dla złożonego strumienia Poissona w [72].

3. KOLEJKOWANIE STRUMIENI MMPP

W tym rozdziale przedstawione będą charakterystyki kolejkowania procesów typu MMPP. MMPP jako model ruchu łączy w sobie dwie ważne cechy – względną prostotę opisu i duże możliwości naśladowania statystycznych własności ruchu, w tym samopodobieństwa. Wyznaczone zostaną wszystkie istotne charakterystyki kolejkowania MMPP dla stanu ustalonego i nieustalonego, m. in. długość kolejki, opóźnienie, liczba strat pakietów i okres przepięnienia bufora.

3.1. Definicja i własności MMPP

Proces MMPP (*ang.* Markov-modulated Poisson process¹) różni się tym od zwykłego procesu Poissona, że w wypadku MMPP intensywność strumienia jest zmienna, a nie stała. Ściślej mówiąc, intensywność ta zmienia się zgodnie ze zmianami stanu łańcucha Markowa z ciągłym czasem $J(t)$ (łańcuch ten nazywać będziemy łańcuchem modulującym). A zatem, jeżeli łańcuch modulujący znajduje się w stanie $i \in E$, gdzie $E = \{1, \dots, m\}$ jest przestrzenią stanów tego łańcucha, to chwilowo MMPP zachowuje się dokładnie tak, jak proces Poissona z parametrem λ_i , gdzie $\lambda_i \in \{\lambda_1, \dots, \lambda_m\}$.

W związku z tym, MMPP parametryzuje się, podając:

- macierz intensywności łańcucha modulującego Q ,
- wektor $(\lambda_1, \dots, \lambda_m)$.

W zapisie wzorów związanych z MMPP często wygodnie jest się posługiwać macierzą diagonalną

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m),$$

w której na przekątnej znajdują się intensywności $\lambda_1, \dots, \lambda_m$, zaś poza przekątną – zera.

¹Tzn. modulowany markowsko proces Poissona. Dla wygody będziemy używać po prostu skrótu MMPP.

Okazuje się, że przedstawiona powyżej konstrukcja ma bardzo duże możliwości naśladowania zjawisk statystycznych obserwowanych w ruchu w sieciach pakietowych. Odpowiednio parametryzując MMPP, możemy modelować dokładnie rozkład liczby pakietów wpływających do kolejki w jednostce czasu. Nie chodzi tu tylko o zgodność średniej i wariancji, ale również o dokładne dopasowanie kształtu funkcji rozkładu. Ponadto, wykorzystując MMPP, możemy ukształtować precyzyjnie funkcję autokorelacji intensywności ruchu zgodnie z zapisanymi śladami ruchu w dowolnie długim (ale skończonym) horyzoncie czasowym [181]. Ponieważ samopodobieństwo wiąże się z wolno malejącą funkcją autokorelacji, to właśnie ta własność umożliwia naśladowanie samopodobieństwa w ruchu sieciowym. Używając MMPP nigdy nie otrzymujemy prawdziwego samopodobieństwa, jednak możemy doprowadzić do tego, że proces zachowuje się samopodobnie przy kilku (dowolnej skończonej liczbie) przeskalowaniach osi czasu. Do zastosowań praktycznych taka imitacja w zupełności wystarcza, gdyż nasze obserwacje ruchu w sieci pakietowej mogą dotyczyć również tylko skończonej liczby skal czasu.

Biorąc pod uwagę te własności MMPP, nie dziwi fakt, że proces ten stał się niezwykle popularny w modelowaniu ruchu w sieciach pakietowych (przykładowa literatura jest podana na końcu tego rozdziału).

Najważniejszą własnością MMPP, z której będziemy wielokrotnie korzystać, jest **markowska struktura** tego procesu. Własność ta oznacza, że znając stan łańcucha modulującego w pewnej chwili T , możemy określić zachowanie probabilistyczne tego procesu po chwili T jedynie na podstawie $J(T)$, bez znajomości historii procesu przed T .

Przejrzymy teraz pozostałe charakterystyki i własności MMPP.

Podstawową charakterystyką MMPP jest, oczywiście, **średnia intensywność**. Do jej wyznaczenia potrzebny jest **rozkład stacjonarny łańcucha modulującego**, tzn. wektor π o własnościach

$$\pi Q = (0, \dots, 0),$$

$$\pi \cdot \mathbf{1} = 1,$$

gdzie

$$\mathbf{1} = (1, \dots, 1)^T$$

to złożony z jedynek wektor kolumnowy. Znając wektor π , średnią intensywność MMPP obliczamy jako

$$\lambda = \pi \cdot (\lambda_1, \dots, \lambda_m)^T.$$

Posługując się wartością λ , należy pamiętać, że chwilowe intensywności mogą nieraz bardzo od niej odbiegać.

Oznaczmy przez J_k stan łańcucha modulującego w chwili nadejścia k -tego pakietu, zaś przez T_k czas pomiędzy $(k-1)$ -szym a k -tym pakietem. Przyjmijmy $J_0 = J(0)$, $T_0 = 0$. Wtedy ciąg (J_n, T_n) , $n \geq 0$, tworzy markowski proces odnowy² o macierzy przejścia:

$$\begin{aligned}\hat{F}(x) &= [\hat{F}_{i,j}(x)]_{i,j}, \\ \hat{F}_{i,j}(x) &= \mathbf{P}(J_k = j, T_k \leq x | J_{k-1} = i),\end{aligned}$$

którą można wyznaczyć w następujący sposób:

$$\begin{aligned}\hat{F}(x) &= \int_0^x e^{(Q-\Lambda)u} du \cdot \Lambda \\ &= (I - e^{(Q-\Lambda)x})(\Lambda - Q)^{-1} \Lambda \\ &= (I - e^{(Q-\Lambda)x}) \hat{F}(\infty),\end{aligned}$$

gdzie I oznacza macierz jednostkową o wymiarach $m \times m$. Jasne jest, że J_k tworzy łańcuch Markowa z dyskretnym czasem. **Macierz przejścia** J_k ma postać

$$\hat{F}(\infty) = (\Lambda - Q)^{-1},$$

zaś **rozkład stacjonarny** J_k

$$p = \frac{1}{\lambda} \pi \Lambda.$$

Załóżmy teraz, że chwila $t = 0$ jest chwilą nadejścia pakietu. Wówczas transformata Laplace'a-Stieltjesa $\hat{F}(x)$ ma postać

$$\hat{F}^*(s) = \mathbf{E} e^{-sT_1} = (sI - Q + \Lambda)^{-1} \Lambda,$$

zaś transformatę łączną T_1, \dots, T_n przedstawia wzór

$$\begin{aligned}\hat{F}^*(s_1, \dots, s_n) &= \mathbf{E} \left[\exp \left(- \sum_{k=1}^n s_k T_k \right) \right] \\ &= \prod_{k=1}^n \left[(s_k I - Q + \Lambda)^{-1} \Lambda \right].\end{aligned}\quad (3.1)$$

Różniczkując (3.1), możemy wyznaczyć następujące wartości oczekiwane:

$$\mathbf{E} T_k^i = i! \left[(\Lambda - Q)^{-1} \Lambda \right]^{k-1} (\Lambda - Q)^{-(i+1)} \Lambda, \quad i, k \geq 1, \quad (3.2)$$

$$\mathbf{E}(T_1 T_{k+1}) = (\Lambda - Q)^{-2} \Lambda \left[(\Lambda - Q)^{-1} \Lambda \right]^{-(k+1)} (\Lambda - Q)^{-2} \Lambda, \quad k \geq 1. \quad (3.3)$$

²Definicje markowskiego procesu odnowy można znaleźć np. w [91].

Wówczas **macierz korelacji MMPP** można wyrazić wzorem:

$$\begin{aligned} \mathbf{E}[(T_1 - \mathbf{E}T_1)(T_{k+1} - \mathbf{E}T_{k+1})] &= \\ &= (\Lambda - Q)^{-2}\Lambda \left[(\Lambda - Q)^{-1}\Lambda \right]^{k-1} [I - (\Lambda - Q)^{-1}\Lambda](\Lambda - Q)^2\Lambda. \end{aligned} \quad (3.4)$$

Zauważmy, że wzory (3.2), (3.3) i (3.4) przedstawiają macierze, tzn. uwzględniają one również zmiany stanu łańcucha modulującego.

Dzięki (3.2) łatwo otrzymujemy wzór na **k -ty moment zwykły** czasu pomiędzy dwoma kolejnymi pakietami dla MMPP w stanie ustalonym:

$$m_k(T_i) = k! p (\Lambda - Q)^{-(k+1)}\Lambda \cdot \mathbf{1}. \quad (3.5)$$

Wykorzystując (3.5), możemy obliczyć **wariancję** czasu pomiędzy dwoma kolejnymi pakietami dla MMPP w stanie ustalonym:

$$Var = m_2(T_i) - \frac{1}{\lambda^2}.$$

Dalej, dzięki (3.4), dostajemy wzór na **autokowariancję** (dla k kroków wstecz) odstępów czasów pomiędzy pakietami dla MMPP w stanie ustalonym:

$$Cov(k) = p (\Lambda - Q)^{-2}\Lambda \left[[(\Lambda - Q)^{-1}\Lambda]^{k-1} - \mathbf{1} p \right] (\Lambda - Q)^{-2}\Lambda \mathbf{1},$$

Wówczas **autokorelacja** pomiędzy T_i oraz T_{i-k} dla MMPP w stanie ustalonym ma postać

$$Corr(k) = \frac{Cov(k)}{Var}.$$

Kolejną, niezwykle ważną charakterystyką procesu MMPP jest jego **funkcja licząca**, określona jako

$$P_{i,j}(n, t) = \mathbf{P}(N(t) = n, J(t) = j | N(0) = 0, J(0) = i),$$

gdzie $N(t)$ oznacza liczbę pakietów w strumieniu wejściowym w przedziale czasu $(0, t]$. Macierze $P(n, t)$ o elementach $P_{i,j}(n, t)$ spełniają równanie Chapmana-Kołmogorowa:

$$P'(n, t) = P(n, t)(Q - \Lambda) + P(n - 1, t)\Lambda, \quad n \geq 1, t \geq 0,$$

gdzie

$$P(0, 0) = I.$$

Wobec tego **funkcja tworząca**

$$P^*(z, t) = \sum_{n=0}^{\infty} P(n, t) z^n$$

spełnia równania:

$$\frac{d}{dt} P^*(z, t) = P^*(z, t)(Q - \Lambda) + zP^*(n-1, t)\Lambda,$$

$$P^*(z, 0) = I.$$

Ich rozwiązanie wyraża się jawnym wzorem:

$$P^*(z, t) = e^{(Q-(1-z)\Lambda)t}, \quad |z| \leq 1. \quad (3.6)$$

Wzór (3.6) tylko w wyjątkowych wypadkach można zastosować do efektywnego obliczenia funkcji liczącej dla MMPP. Na szczęście w praktyce rzadko się zdarza, by konieczne było jej wyznaczenie – w tej książce nie jest to potrzebne w ani jednym wypadku.

Natomiast często konieczne jest wyznaczenie **całkowych funkcjonałów funkcji liczącej** postaci

$$\int_0^{\infty} g(t) P_{i,j}(n, t) dt \quad (3.7)$$

dla pewnej funkcji $g(t)$. I tak, typowymi funkcjonałami postaci (3.7), którymi będziemy się posługiwać w dalszej części, są:

$$a_{k,i,j}(s) = \int_0^{\infty} e^{-st} P_{i,j}(k, t) dF(t), \quad (3.8)$$

$$d_{k,i,j}(s) = \int_0^{\infty} e^{-st} P_{i,j}(k, t) (1 - F(t)) dt, \quad (3.9)$$

$$c_{k,i,j}(s, \sigma) = \int_0^{\infty} e^{-\sigma t} P_{i,j}(k, t) dt \int_0^{\infty} e^{-sx} d_x F(x+t). \quad (3.10)$$

Funkcjonały te będziemy też zapisywać w postaci następujących macierzy $m \times m$:

$$A_k(s) = [a_{k,i,j}(s)]_{i,j}, \quad (3.11)$$

$$\bar{D}_k(s) = [d_{k,i,j}(s)]_{i,j}, \quad (3.12)$$

$$C_k(s, \sigma) = [c_{k,i,j}(s, \sigma)]_{i,j}. \quad (3.13)$$

Z praktycznego punktu widzenia niezwykle istotne jest, że macierze te można efektywnie obliczać przy pomocy tzw. metody uniformizacji. Metoda ta zostanie przedstawiona w podrozdziale 5.2.

Różniczkując macierz $P^*(z, t)$ ze względu na zmienną z , możemy wyznaczyć **średnią** $N(t)$, tzn. średnią liczbę pakietów nadchodzących w przedziale czasu $(0, t]$, jak również wyższe momenty zwykłe $N(t)$, tzn. $m_k(N(t))$. Mianowicie, oznaczając

$$M_{i,j}(t) = \sum_{n=0}^{\infty} n P_{ij}(n, t),$$

otrzymujemy

$$\begin{aligned} M(t) &= [M_{i,j}(t)]_{i,j} \\ &= \left[\frac{\partial}{\partial z} P^*(z, t) \right]_{z=1} \\ &= \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n-1} Q^k \Lambda Q^{n-1-k}. \end{aligned} \quad (3.14)$$

Stąd możemy wyznaczyć

$$m_1(N(t)) = M(t) \mathbf{1} = \mathbf{1} \lambda t + (e^{Qt} - I)(Q + \mathbf{1}\pi)^{-1} \Lambda \mathbf{1}.$$

Na i -tym miejscu wektora $m_1(N(t))$ znajduje się średnia liczba pakietów nadchodzących w przedziale $(0, t]$ przy założeniu $J(0) = i$. Jeżeli rozkładem początkowym łańcucha modulującego jest π , to

$$\mathbf{E}N(t) = \pi m_1(N(t)) = \lambda t.$$

Aby wyznaczyć **drugi moment** $N(t)$, dwukrotnie różniczkujemy $P^*(z, t)$, tzn.

$$\begin{aligned} m_2(N(t)) &= \left[\frac{\partial^2}{\partial z^2} P^*(z, t) \right]_{z=1} \cdot \mathbf{1} + m_1(N(t)) \\ &= \sum_{n=2}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n-2} Q^k \Lambda Q^{n-2-k} \Lambda \mathbf{1} + m_1(N(t)) \\ &= 2 \left[M(t)(Q + \mathbf{1}\pi)^{-1} - (e^{Qt} - I)(Q + \mathbf{1}\pi)^{-1} \Lambda (Q + \mathbf{1}\pi)^{-1} \right. \\ &\quad \left. - t \mathbf{1}\pi \Lambda (Q + \mathbf{1}\pi)^{-1} + (e^{Qt} - I - Qt)(Q + \mathbf{1}\pi)^{-2} \Lambda \mathbf{1}\pi \right. \\ &\quad \left. + \frac{t^2}{2} \mathbf{1}\pi \Lambda \mathbf{1}\pi \right] \Lambda \mathbf{1} + m_1(N(t)). \end{aligned} \quad (3.15)$$

Jeżeli rozkładem początkowym łańcucha modulującego jest π , to

$$\begin{aligned} \mathbf{E}(N(t))^2 &= t^2 \lambda^2 + 2t [\lambda^2 - \pi \Lambda (Q + \mathbf{1}\pi)^{-1} \Lambda \mathbf{1}] \\ &\quad + 2\pi \Lambda (e^{Qt} - I)(Q + \mathbf{1}\pi)^{-2} \Lambda \mathbf{1} + \lambda t. \end{aligned} \quad (3.16)$$

Bardzo istotna własność MMPP ujawnia się w wypadku **superpozycji** wielu procesów tego rodzaju. Okazuje się, że składając zdarzenia z wielu strumieni MMPP, również otrzymujemy strumień MMPP. Dokładniej, jeżeli danych jest n procesów MMPP opisanych parametrami $(Q_1, \Lambda_1), \dots, (Q_n, \Lambda_n)$, to ich superpozycja jest procesem MMPP z następującymi parametrami:

$$\begin{aligned} Q &= Q_1 \oplus Q_2 \oplus \dots \oplus Q_n, \\ \Lambda &= \Lambda_1 \oplus \Lambda_2 \oplus \dots \oplus \Lambda_n, \end{aligned}$$

gdzie \oplus oznacza sumę Kroneckera macierzy. Sumę Kroneckera definiuje się jako

$$A \oplus B = (A \otimes I_B) + (I_A \otimes B),$$

gdzie iloczyn Kroneckera \otimes jest określony następująco:

$$C \otimes D = \begin{bmatrix} c_{11}D & c_{12}D & \dots & c_{1m}D \\ \vdots & \vdots & & \vdots \\ c_{n1}D & c_{n2}D & \dots & c_{nm}D \end{bmatrix},$$

zaś I_B oraz I_A oznaczają macierze jednostkowe tego samego rzędu co A i B .

MMPP z dwoma stanami modulującymi ($m = 2$) i z jedną intensywnością 0 nazywamy też procesem typu „on-off”. Procesy „on-off” modelują ruch generowany przez źródła, w których okresy aktywności są poprzedzielane okresami milczenia i są często wykorzystywane w modelowaniu ruchu w sieciach komputerowych. Superpozycja wielu procesów „on-off” jest, oczywiście, procesem MMPP.

Jak już zostało wcześniej powiedziane, MMPP można sparametryzować tak, by dokładnie oddawał własności statystyczne zarejestrowanych śladów ruchu sieciowego. W tym celu opracowano wiele algorytmów. Jeden z nich został przedstawiony w podrozdziale 5.3. Tam również znajdują się odnośniki kierujące Czytelnika do obszernej literatury na ten temat.

Podane zostaną teraz dwie przykładowe parametryzacje MMPP, których będziemy używać później do wyznaczania charakterystyk kolejowania.

Przykład 3.1. Pierwsza, prosta parametryzacja MMPP pochodzi z [19]. Łańcuch modulujący ma tylko dwa stany ($m = 2$) oraz

$$Q = \begin{bmatrix} -8.4733 \cdot 10^{-4} & 8.4733 \cdot 10^{-4} \\ 5.0201 \cdot 10^{-6} & -5.0201 \cdot 10^{-6} \end{bmatrix},$$

$$(\lambda_1, \lambda_2) = (1.0722, 0.48976).$$

Jak widzimy intensywność λ_1 jest ponaddwukrotnie większa niż λ_2 , za to czas przebywania łańcucha modulującego w stanie 2 jest przeciętnie $\frac{8.4733 \cdot 10^{-4}}{5.0201 \cdot 10^{-6}} = 168$ razy dłuższy niż

w stanie 1. Tak więc mamy tu do czynienia z modelem ruchu, w którym od czasu do czasu występują niewielkie spiętrzenia. Rozkład stacjonarny łańcucha modulującego ma postać

$$\pi = (0.00589, 0.99411),$$

zaś średnia intensywność strumienia pakietów na jednostkę czasu (ms) wynosi

$$\lambda = 0.49319.$$

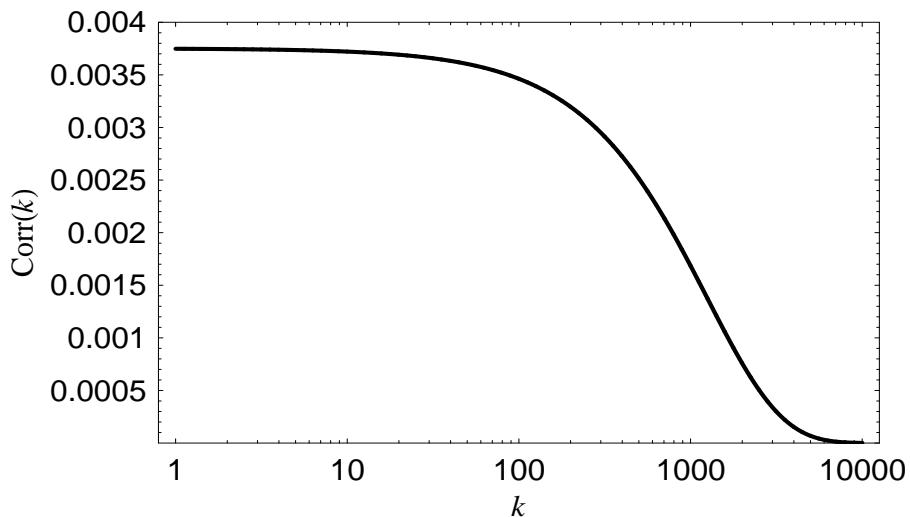
Średni odstęp czasu pomiędzy dwoma kolejnymi pakietami to

$$\frac{1}{\lambda} = 2.02761 \text{ ms},$$

zaś odchylenie standardowe tego czasu

$$\sqrt{Var} = 2.0353 \text{ ms}.$$

Funkcja autokorelacji czasów pomiędzy kolejnymi pakietami jest przedstawiona na rys. 3.1. Jak widać, autokorelacja w prezentowanym MMPP nie jest duża, jednak bardzo wolno zanika i jest istotna dla prawie czterech skal czasu.



Rys. 3.1. Autokorelacja odstępów czasu pomiędzy pakietami w funkcji opóźnienia dla parametryzacji MMPP z Przykładu 3.1

Fig. 3.1. The k -lag autocorrelation of the interarrival times for the MMPP parametrization from Example 3.1

Przykład 3.2. Druga, bardziej skomplikowana, parametryzacja MMPP jest oparta na zapisanym śladzie (ang. trace file) zagregowanego ruchu IP. Wykorzystano milion nagłówków pakietów zapisanych w ogólnodostępnym pliku FRG-1137208198-1.tsh (do pobrania z [93]), przedstawiającym ruch z 14 stycznia 2006 roku zaobserwowany w punkcie agregacji FRG (Front Range GigaPOP) monitorowanym w ramach amerykańskiego projektu PMA (Passive Measurement and Analysis Project, [155]). Średnia intensywność ruchu dla tego zapisu to $\lambda = 71732$ pakietów na sekundę, co przy średnim rozmiarze pakietu wynoszącym 850 B daje łączną intensywność równą 58.11 MB/s. Do tego zapisu ruchu IP dopasowane zostały [57] następujące parametry MMPP:

$$Q = \begin{bmatrix} -172.53 & 38.80 & 30.85 & 0.88 & 102.00 \\ 16.76 & -883.26 & 97.52 & 398.9 & 370.08 \\ 281.48 & 445.97 & -1594.49 & 410.98 & 456.06 \\ 23.61 & 205.74 & 58.49 & -598.93 & 311.09 \\ 368.48 & 277.28 & 7.91 & 32.45 & -686.12 \end{bmatrix},$$

$$(\lambda_1, \dots, \lambda_5) = (59620.6, 113826.1, 7892.6, 123563.2, 55428.2).$$

Podstawowe parametry śladu ruchu IP oraz MMPP są pokazane w tab. 3.1 i na rys. 3.2. Ważne jest tu dobre dopasowanie funkcji autokorelacji, zwłaszcza że autokorelacja ta przyjmuje dość wysokie wartości aż do trzech skal czasu. Jak zostanie później wielokrotnie pokazane, autokorelacja ma bardzo duży wpływ na charakterystyki kolejkowania ruchu. Rozkład stacjonarny łańcucha modulującego dla dopasowanego MMPP ma postać

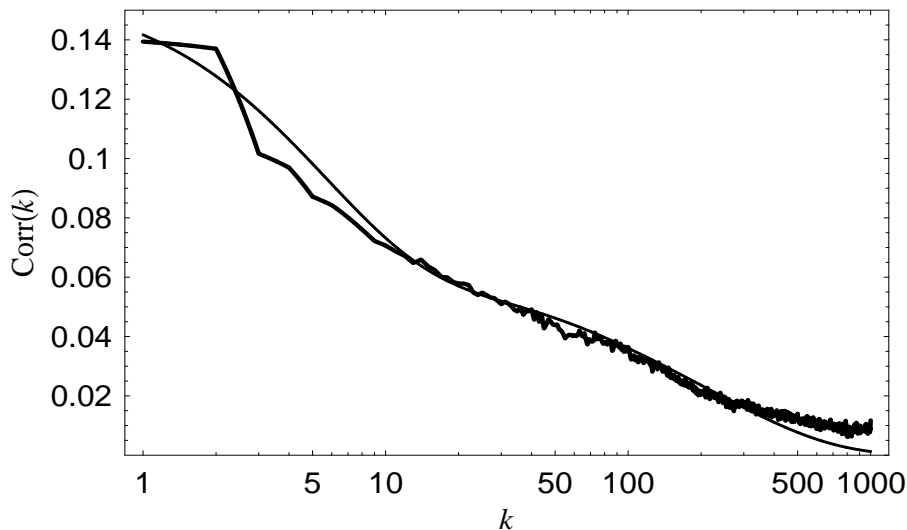
$$\pi = (0.52174, 0.12808, 0.023151, 0.11352, 0.21351).$$

Jak widać, łańcuch najczęściej przebywa w stanie 1, najrzadziej w stanie 3. Odchylenie standardowe czasu pomiędzy pakietami wynosi

$$\sqrt{Var} = 16.866 \mu s.$$

	Średni czas pomiędzy pakietami [μs]	Średnia intensywność ruchu [pakiet/s]
Zapis ruchu IP	13.940	71732
Dopasowany MMPP	13.941	71729

Tab. 3.1. Podstawowe parametry zapisanego śladu ruchu IP i dopasowanego MMPP w Przykładzie 3.2



Rys. 3.2. Autokorelacja odstępów czasu pomiędzy pakietami w funkcji opóźnienia w Przykładzie 3.2. Grubsza linia - wykres dla zarejestrowanego śladu ruchu IP, cieńsza linia - wykres dla dopasowanego MMPP

Fig. 3.2. The k -lag autocorrelation of the interarrival times for the MMPP parametrization from Example 3.2. Thick line - the autocorrelation in the IP traffic trace file, thin line - the autocorrelation of the fitted MMPP

3.2. Charakterystyki kolejkowania MMPP

Zanim przejdziemy do charakterystyk kolejkowania MMPP, wprowadzimy kilka potrzebnych oznaczeń. Szczególnie przydatne będą następujące macierze $m \times m$:

$$\bar{A}_n(s) = \sum_{k=n}^{\infty} A_k(s),$$

$$B_n(s) = A_{n+1}(s) - \bar{A}_{n+1}(s)(\bar{A}_0(s))^{-1},$$

$$Z(s) = \left[\frac{(\lambda_i - Q_{ii})p_{ij}}{s + \lambda_i - Q_{ii}} \right]_{i,j}, \quad \text{dla } p_{ij} = \begin{cases} 0 & \text{dla } i = j, \\ \frac{Q_{ij}}{\lambda_i - Q_{ii}} & \text{dla } i \neq j, \end{cases}$$

$$E(s) = \left[\frac{\Lambda_{ij}}{s + \lambda_i - Q_{ii}} \right]_{i,j},$$

$$M_b(s) = (I - Z(s)) \left[R_{b+1}(s)A_0(s) + \sum_{k=0}^b R_{b-k}(s)B_k(s) \right] - E(s) \left[R_b(s)A_0(s) + \sum_{k=0}^{b-1} R_{b-1-k}(s)B_k(s) \right],$$

gdzie $R_k(s)$ jest potencjałem dla ciągu $A_k(s)$ (wzór (1.3)), zaś zapis Λ_{ij} oznacza element macierzy Λ znajdujący się w i -tym wierszu i j -tej kolumnie.

Będziemy też używać wektora kolumnowego $z(s)$:

$$z(s) = \left(\frac{1}{s + \lambda_1 - Q_{11}}, \dots, \frac{1}{s + \lambda_m - Q_{mm}} \right)^T.$$

W dalszej części będziemy zakładać, że macierze $A_0(s)$, $\bar{A}_0(s)$ i $M_b(s)$ są nieosobliwe. Udowodnienie, że tak jest zawsze (tzn. dla dowolnej parametryzacji MMPP), może być trudne (zobacz np. [30]). Na szczęście w zastosowaniach założenie to nie ma większego znaczenia, gdyż w praktyce nie spotyka się osobliwych macierzy $A_0(s)$, $\bar{A}_0(s)$, $M_b(s)$.

3.2.1. Długość kolejki

Długość kolejki w stanie nieustalonym zależy teraz nie tylko od początkowej długości kolejki, $X(0)$, ale również od początkowego stanu łańcucha modulującego, $J(0)$. Dlatego rozkład długości kolejki będziemy zapisywać, używając indeksów n oraz i , mianowicie:

$$\Phi_{n,i}(t, l) = \mathbf{P}(X(t) = l | X(0) = n, J(0) = i), \quad 0 \leq n \leq b, 1 \leq i \leq m.$$

Udowodnimy teraz wzór na transformatę Laplace'a rozkładu długości kolejki, tzn.

$$\phi_{n,i}(s, l) = \int_0^\infty e^{-st} \Phi_{n,i}(t, l) dt.$$

Wyniki wygodnie będzie zapisywać, używając wektora kolumnowego, w którym poszczególne elementy odpowiadają kolejnym stanom łańcucha modulującego:

$$\phi_n(s, l) = (\phi_{n,1}(s, l), \dots, \phi_{n,m}(s, l))^T.$$

Twierdzenie 3.1. *Transformata Laplace'a rozkładu długości kolejki w systemie MMPP/G/1/b wyraża się wzorem*

$$\begin{aligned} \phi_n(s, l) = & \left(R_{b-n+1}(s)A_0(s) + \sum_{k=0}^{b-n} R_{b-n-k}(s)B_k(s) \right) M_b^{-1}(s)l_b(s, l) \\ & + \sum_{k=0}^{b-n} R_{b-n-k}(s)g_k(s, l), \end{aligned} \quad (3.17)$$

gdzie $R_k(s)$ jest potencjałem (wzór (1.3)) dla ciągu $A_k(s)$ określonego w (3.11), oraz

$$\begin{aligned} g_k(s, l) &= \bar{A}_{k+1}(s)(\bar{A}_0(s))^{-1}r_b(s, l) - r_{b-k}(s, l), \\ r_n(s, l) &= \begin{cases} \mathbf{0} \cdot \mathbf{1}, & \text{dla } l < n, \\ \bar{D}_{l-n}(s) \cdot \mathbf{1}, & \text{dla } n \leq l < b, \\ \frac{1-f(s)}{s} \cdot \mathbf{1} - \sum_{k=0}^{b-n-1} \bar{D}_k(s) \cdot \mathbf{1}, & \text{dla } l = b, \end{cases} \\ l_b(s, l) &= E(s) \sum_{k=0}^{b-1} R_{b-1-k}(s)g_k(s, l) - (I - Z(s)) \sum_{k=0}^b R_{b-k}(s)g_k(s, l) \\ &\quad + \delta_{0l}z(s). \end{aligned}$$

Dowód Twierdzenia 3.1. Dla uproszczenia zapisu będziemy w dowodzie pomijać zmienną l . Jeżeli początkowo kolejka nie jest pusta, tzn. $X(0) = n \in [1, b]$, to wykorzystując wzór na prawdopodobieństwo całkowite względem pierwszej chwili ukończenia obsługi, otrzymujemy

$$\begin{aligned} \Phi_{n,i}(t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \Phi_{n+k-1,j}(t-u) P_{i,j}(k, u) dF(u) \\ &\quad + \sum_{j=1}^m \sum_{k=b-n}^{\infty} \int_0^t \Phi_{b-1,j}(t-u) P_{i,j}(k, u) dF(u) \\ &\quad + \rho_{n,i}(t), \quad 1 \leq i \leq m, \end{aligned} \tag{3.18}$$

gdzie

$$\rho_{n,i}(t) = (1 - F(t)) \cdot \begin{cases} 0 & \text{dla } l < n, \\ \sum_{j=1}^m P_{i,j}(l-n, t) & \text{dla } n \leq l < b, \\ \sum_{j=1}^m \sum_{k=b-n}^{\infty} P_{i,j}(k, t) & \text{dla } l = b. \end{cases}$$

Pierwszy składnik wzoru (3.18) odpowiada za przypadek, w którym pierwszy czas ukończenia obsługi u wypada wcześniej niż t i do chwili u bufor się nie przepełnia. Drugi składnik (3.18) odpowiada za przypadek, w którym pierwszy czas ukończenia obsługi u wypada wcześniej niż t i przed chwilą u następuje przepełnienie bufora. Wreszcie trzeci składnik, $\rho_{n,i}(t)$, odpowiada za przypadek, w którym pierwszy czas ukończenia obsługi u wypada po chwili t .

W sytuacji $X(0) = 0$ użyjemy wzoru na prawdopodobieństwo całkowite względem chwili wystąpienia pierwszego zdarzenia w strumieniu MMPP (którym może być pojawienie się pakietu lub zmiana stanu łańcucha modulującego). Otrzymujemy,

że

$$\begin{aligned}\Phi_{0,i}(t) &= \sum_{j=1}^m \int_0^t \Phi_{0,j}(t-u)(\lambda_i - Q_{ii})p_{ij}e^{-(\lambda_i - Q_{ii})u} du \\ &\quad + \sum_{j=1}^m \int_0^t \Phi_{1,j}(t-u)\Lambda_{ij}e^{-(\lambda_i - Q_{ii})u} du \\ &\quad + \delta_{0i}e^{-(\lambda_i - Q_{ii})t}.\end{aligned}\quad (3.19)$$

Jeżeli do obu stron (3.18) i (3.19) zastosujemy transformatę Laplace'a oraz zapis macierzowy, to dostaniemy następujący układ równań:

$$\phi_n(s) = \sum_{k=0}^{b-n-1} A_k(s)\phi_{n+k-1}(s) + \sum_{k=b-n}^{\infty} A_k(s)\phi_{b-1}(s) + r_n(s), \quad 0 < n \leq b, \quad (3.20)$$

$$\phi_0(s) = Z(s)\phi_0(s) + E(s)\phi_1(s) + \delta_{0l}z(s). \quad (3.21)$$

Następnie, wykonując podstawienie

$$\varphi_n(s) = \phi_{b-n}(s),$$

otrzymujemy:

$$\sum_{k=-1}^n A_{k+1}(s)\varphi_{n-k}(s) - \varphi_n(s) = \psi_n(s), \quad 0 \leq n < b, \quad (3.22)$$

$$\varphi_b(s) = Z(s)\varphi_b(s) + E(s)\varphi_{b-1}(s) + \delta_{0l}z(s), \quad (3.23)$$

gdzie

$$\psi_n(s) = A_{n+1}(s)\varphi_0(s) - \sum_{k=n+1}^{\infty} A_k(s)\varphi_1(s) - r_{b-n}(s).$$

Ponieważ układ (3.22) ma taką samą postać jak (1.8), z Twierdzenia 1.3 wnioskujemy, że ma on następujące rozwiązanie

$$\varphi_n(s) = R_{n+1}(s)c(s) + \sum_{k=0}^n R_{n-k}(s)\psi_k(s), \quad n \geq 0, \quad (3.24)$$

gdzie $c(s)$ jest wektorem kolumnowym niezależnym od n , a $R_n(s)$ jest określone w (1.3).

Dla $n = 0$ wzór (3.24) daje

$$c(s) = A_0(s)\varphi_0(s), \quad (3.25)$$

zaś (3.22)

$$\varphi_0(s) = \sum_{k=0}^{\infty} A_k(s) \varphi_1(s) + r_b(s). \quad (3.26)$$

Następnie, wstawiając kolejno $n = b$ i $n = b - 1$ do (3.24) oraz wykorzystując warunek (3.23), możemy łatwo obliczyć $\varphi_0(s)$, co zakończy dowód twierdzenia. ■

Twierdzenie 3.1 możemy zastosować do znajdowania rozkładu długości kolejki zarówno w stanie nieustalonym, jak i ustalonym. Dla stanu nieustalonego konieczne jest użycie algorytmu odwracania transformaty Laplace'a (podrozdział 5.1.1), zaś dla stanu ustalonego wystarczy obliczyć

$$\Phi(l) = \lim_{s \rightarrow 0^+} s \phi_{n,i}(s, l) \quad (3.27)$$

dla dowolnego n oraz i .

Przykład 3.3. Wykonamy teraz obliczenia rozkładu długości kolejki dla parametryzacji MMPP z Przykładu 3.1. Zakładamy $b = 50$ pakietów oraz stały czas obsługi (transmisji) pakietu $d = 1$ ms. Przy średniej intensywności strumienia wejściowego $\lambda = 0.49319$ pakietów na milisekundę daje to małe obciążenie systemu $\rho = \lambda d = 49\%$. Wykorzystując (3.27) i Twierdzenie 3.1, możemy łatwo obliczyć, że średnia długość kolejki w stanie ustalonym wynosi

$$\mathbf{E}\Phi = 0.93279,$$

wariacja długości kolejki w stanie ustalonym

$$\text{Var}(\Phi) = 9.1248,$$

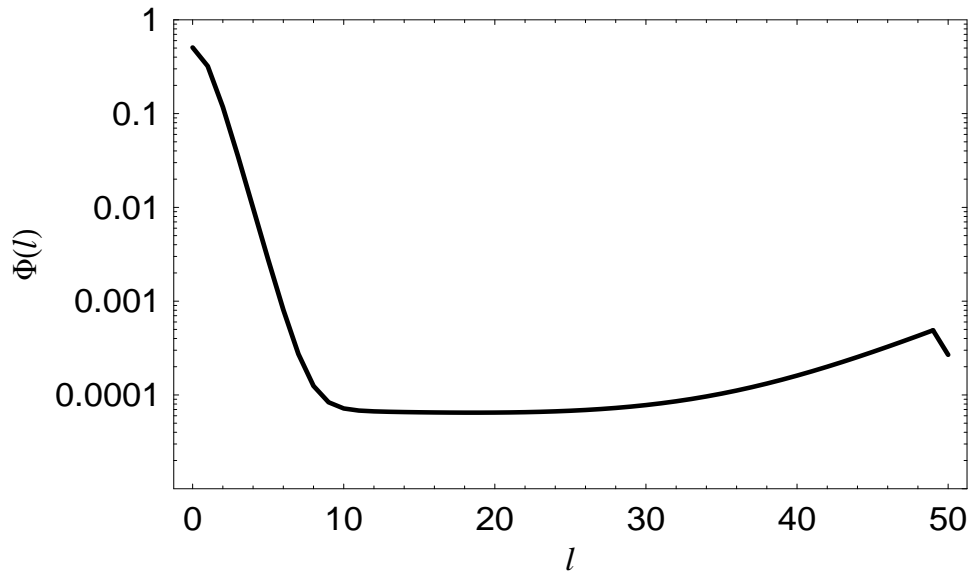
prawdopodobieństwo, że bufor jest przepełniony

$$\Phi(b) = 2.6792 \times 10^{-4},$$

zaś prawdopodobieństwo, że bufor jest pusty

$$\Phi(0) = 0.50710.$$

Jak widać, średnia długość kolejki jest mała, zaś prawdopodobieństwo zdarzenia, że bufor jest pusty - wysokie. Jest to oczywiście konsekwencją małego obciążenia systemu. Dokładna postać rozkładu długości kolejki w stanie ustalonym jest pokazana na rys. 3.3. Warto zwrócić uwagę na charakterystyczny kształt prawego końca wykresu (ostrze). Kształt taki spowodowany jest faktem, że w stanie ustalonym prawdopodobieństwo kolejki równej b pakietów jest skokowo niższe od prawdopodobieństwa kolejki równej $b - 1$ pakietów (efekt bariery odbijającej). Jest to typowe zjawisko w rozkładzie długości kolejki i będziemy je jeszcze wielokrotnie obserwować na innych rysunkach.



Rys. 3.3. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.3
 Fig. 3.3. Steady-state distribution of the queue length in Example 3.3

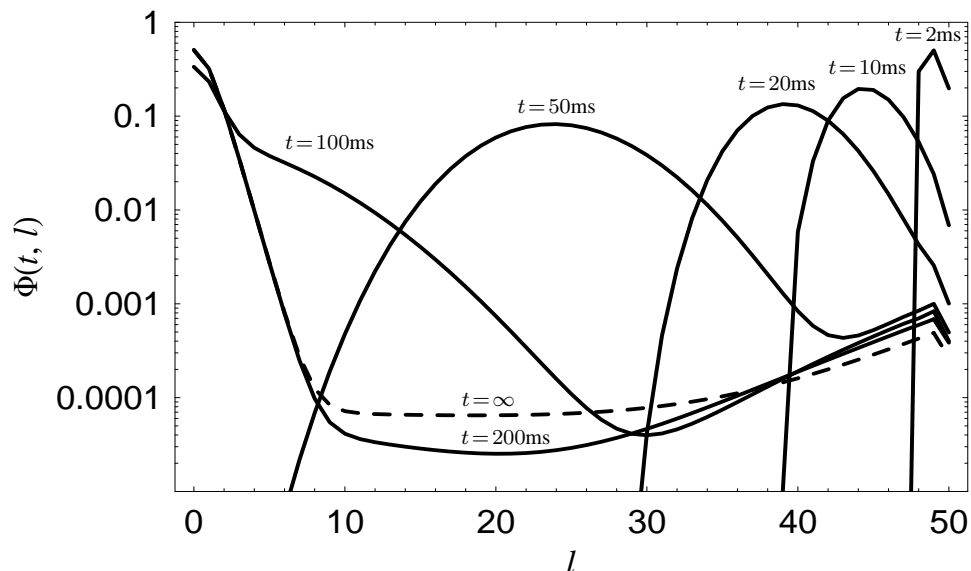
Na rys. 3.4 pokazany jest z kolei rozkład długości kolejki w różnych chwilach czasu, przy założeniu że na początku bufor jest przepelniony, tzn. $X(0) = b$. Widzimy tam wyraźnie zbieżność rozkładu długości kolejki do stanu ustalonego (przerywana linia) i jego pośrednie kształty. Kształt rozkładu zaczyna być bardzo zbliżony do kształtu w stanie ustalonym dla t większych od 200 ms. Możemy zauważyć też opisany powyżej efekt bariery odbijającej, który, początkowo nieobecny, pojawia się, w miarę jak system zbliża się do stanu nieustalonego.

Wreszcie, rys. 3.5 przedstawia prawdopodobieństwo przepelnienia bufora w funkcji rozmiaru bufora w różnych chwilach czasu, przy założeniu że początkowo bufor jest pełny, czyli $X(0) = b$.

Przykład 3.4. Zbadamy teraz długości kolejki dla ruchu MMPP z Przykładu 3.2 dla dwóch rozmiarów bufora, tzn. 100 KB (120 pakietów) oraz 1 MB (1234 pakiety). Zakładamy stały czas transmisji pakietu $d = 10.133 \mu s$, co daje umiarkowane obciążenie systemu $\rho = 72.6\%$.

Dla bufora 100 KB podstawowe parametry kolejki są następujące:

- średnia długość kolejki w stanie ustalonym: $E\Phi = 14.259 \text{ KB}$,
- odchylenie długości kolejki w stanie ustalonym: $\sqrt{\text{Var}(\Phi)} = 26.259 \text{ KB}$,
- prawdopodobieństwo, że bufor jest przepelniony: $\Phi(b) = 6.5549 \times 10^{-3}$,
- prawdopodobieństwo, że bufor jest pusty: $\Phi(0) = 0.28120$.



Rys. 3.4. Rozkład długości kolejki w Przykładzie 3.3 w różnych chwilach czasu (stan nieustalony). Początkowa długość kolejki wynosiła 50 pakietów, rozkład początkowy stanu łańcucha modulującego π

Fig. 3.4. Queue size distribution in Example 3.3 at different times (transient case). The initial queue size was 50 packets; the initial distribution of the modulating state was π

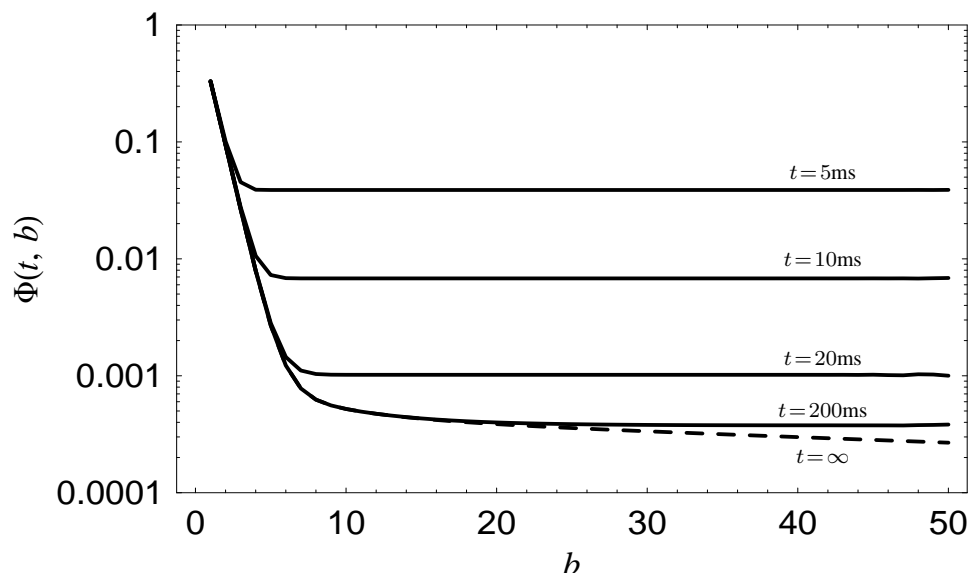
Dla bufora 1 MB mamy odpowiednio:

- $E\Phi = 20.672 \text{ KB}$,
- $\sqrt{\text{Var}(\Phi)} = 44.392 \text{ KB}$,
- $\Phi(b) = 1.5601 \times 10^{-9}$,
- $\Phi(0) = 0.27318$.

Jak widać, zwiększenie rozmiaru bufora ze 100 KB do 1 MB powoduje duży spadek prawdopodobieństwa przepełnienia bufora przy umiarkowanym wzroście średniej długości kolejki i jej odchylenia.

Kształty rozkładów długości kolejki w stanie ustalonym dla buforów 100 KB i 1 MB są przedstawione odpowiednio na rys. 3.6 i 3.7. Jak widzimy, rozkład wygląda podobnie w obu wypadkach. Taki kształt rozkładu jest często spotykany w systemach ze skończonym buforem. Masa prawdopodobieństwa jest silnie skoncentrowana wokół barier, tzn. 0 i b, zaś na pozostałej części wykres jest zbliżony do liniowego (w skali logarytmicznej). Ponadto na prawych końcach wykresów pojawia się charakterystyczne ostrze spowodowane opisanym wcześniej efektem bariery odbijającej.

Rys. 3.8 pokazuje rozkład długości kolejki dla bufora 100 KB w różnych chwilach czasu, przy założeniu że na początku bufor jest wypełniony w połowie, tzn. $X(0) = 0.5b$. Widać



Rys. 3.5. Prawdopodobieństwo przepełnienia bufora w funkcji rozmiaru bufora (Przykład 3.3) w różnych chwilach czasu (stan nieustalony). Początkowa długość kolejki wynosiła b , rozkład początkowy stanu łańcucha modulującego π

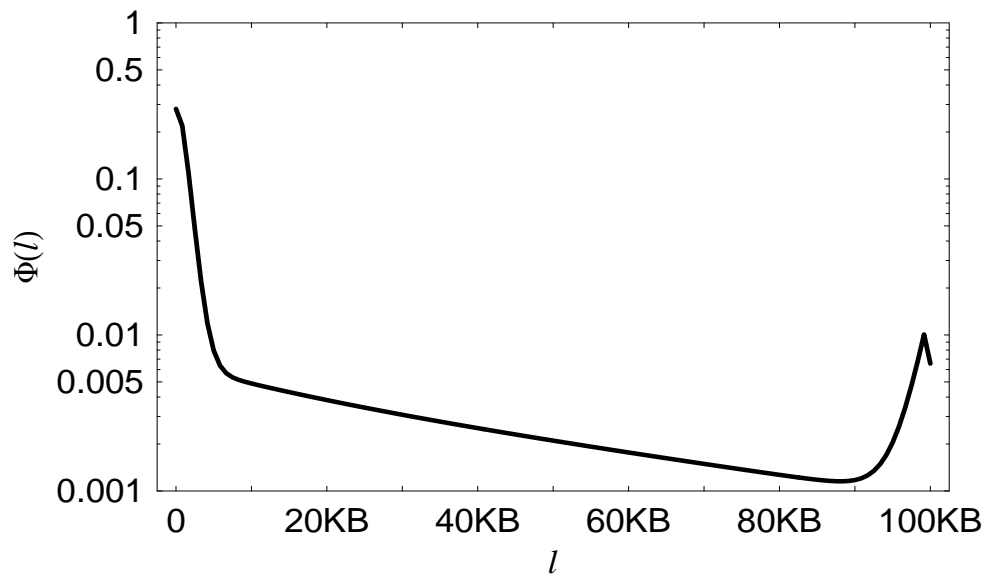
Fig. 3.5. Buffer overflow probability as a function of the buffer size (Example 3.3) at different times (transient case). The initial queue size was b packets; the initial distribution of the modulating state was π

zbieżność do rozkładu w stanie ustalonym, kształty rozkładów zbliżone do stanu ustalonego otrzymujemy dla $t > 2$ ms.

Na rys. 3.9 pokazano, jak zmienia się średnia długość kolejki w czasie dla pięciu różnych poziomów początkowego zapelnienia bufora - od 0% do 100%. Zbieżność do stanu ustalonego następuje po około 8 ms, niezależnie od początkowego zapelnienia bufora. Co ciekawe, wykres średniej długości kolejki w czasie może nie być monotoniczny (dla $X(0) = 25$ KB).

Na koniec tego przykładu zobaczymy, jak autokorelacja strumienia pakietów wpływa na charakterystyki kolejkowania. Do tego celu, oprócz rozważanego MMPP, wykorzystamy prosty proces Poissona o identycznej intensywności, tzn. $\lambda = 71729$ pakietów na sekundę. Rys. 3.10 pokazuje średnią długość kolejki dla tych dwóch różnych modeli ruchu. W stanie nieustalonym obie charakterystyki są podobne, choć dla strumienia Poissona stan ustalony jest osiąganym nieco szybciej. W stanie ustalonym średnia długość kolejki dla ruchu poissonowskiego jest dziesięciokrotnie mniejsza.

Różnica pomiędzy tymi modelami staje się drastyczna, jeżeli wziąć pod uwagę prawdopodobieństwo przepełnienia bufora. Efekt ten został pokazany na rys. 3.11. Nieuwzględnienie autokorelacji spowodowało optymistyczne zaniżenie wartości prawdopodobieństwa przepełnienia w stanie ustalonym o 30 rzędów wielkości!



Rys. 3.6. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.4 dla bufora o rozmiarze 100 KB

Fig. 3.6. Queue size distribution in steady state in Example 3.4 for the buffer size of 100 KB

3.2.2. Współczynnik strat pakietów

Podstawiając $l = b$ do (3.27), możemy łatwo znaleźć prawdopodobieństwo, że bufor jest przepełniony. Niestety, w przeciwieństwie do prostego strumienia Poissona, dla MMPP prawdopodobieństwo przepełnienia bufora nie jest równe współczynnikowi strat pakietów. Dlatego też w wypadku MMPP współczynnik strat musimy wyznaczyć oddzielnie.

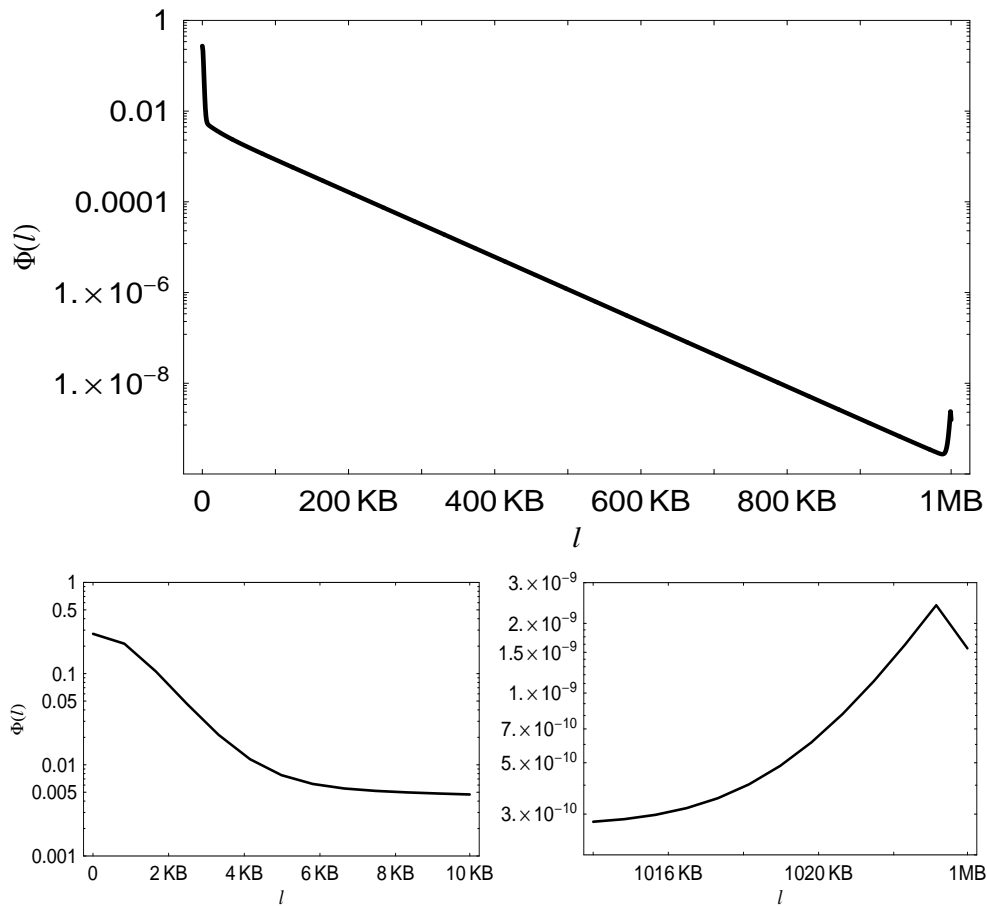
W rzeczywistości wyznaczmy tu znacznie bardziej ogólną charakterystykę, mianowicie średnią liczbę pakietów traconych w przedziale $(0, t]$. Przy jej pomocy łatwo będzie można znaleźć współczynnik strat. Ponadto będzie można analizować liczbę strat w stanie nieustalonym.

Niech $L(t)$ oznacza liczbę pakietów utraconych w przedziale $(0, t]$ oraz $\Delta_{n,i}(t)$ jej średnią wartość dla $X(0) = n$ i $J(0) = i$, tzn.

$$\Delta_{n,i}(t) = \mathbf{E}(L(t)|X(0) = n, J(0) = i).$$

Ponadto niech

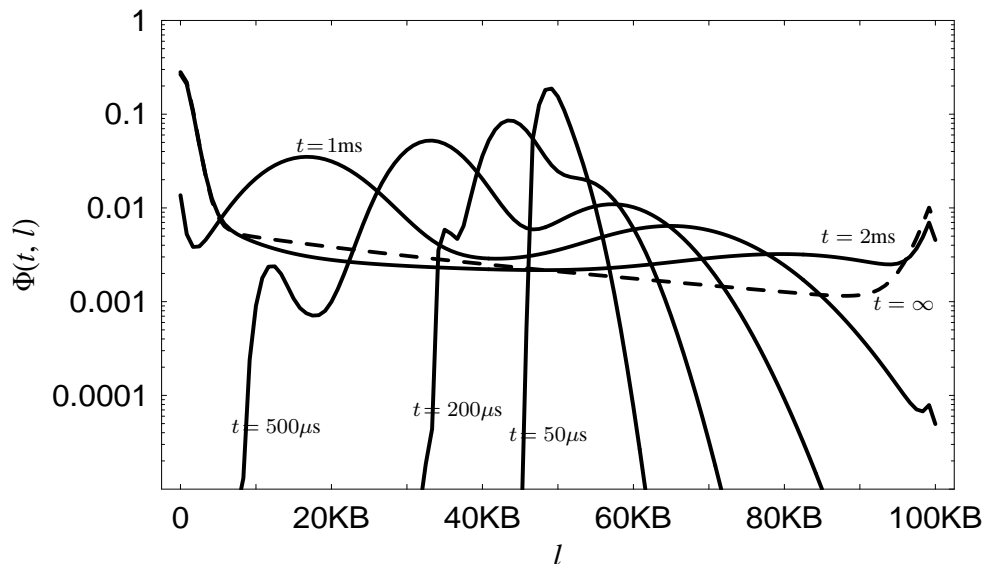
$$\begin{aligned} \delta_{n,i}(s) &= \int_0^{\infty} e^{-st} \Delta_{n,i}(t) dt, \\ \delta_n(s) &= (\delta_{n,1}(s), \dots, \delta_{n,m}(s))^T. \end{aligned}$$



Rys. 3.7. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.4 dla bufora 1 MB (u góry). Na dolnych rysunkach powiększenie zakresów 0-10 KB i 1012-1024 KB
 Fig. 3.7. Queue size distribution in steady state in Example 3.4 for the buffer size of 1 MB (above) and close-ups of the range 0-10 KB and 1012-1024 KB (beneath)

Twierdzenie 3.2. Transformata Laplace'a średniej liczby strat w przedziale $(0, t]$ w systemie MMPP/G/1/b wyraża się wzorem

$$\delta_n(s) = \left(R_{b-n+1}(s)A_0(s) + \sum_{k=0}^{b-n} R_{b-n-k}(s)B_k(s) \right) M_b^{-1}(s)y_b(s) + \sum_{k=0}^{b-n} R_{b-n-k}(s)v_k(s), \quad (3.28)$$



Rys. 3.8. Rozkład długości kolejki w Przykładzie 3.4 w różnych chwilach czasu (stan nieustalony). Początkowa długość kolejki wynosiła 50 KB, rozmiar bufora 100 KB, rozkład początkowy stanu łańcucha modulującego π

Fig. 3.8. Queue size distribution in Example 3.4 at different times (transient case). The initial queue size was 50 KB, the buffer size was 50 KB and the distribution of the modulating state was π

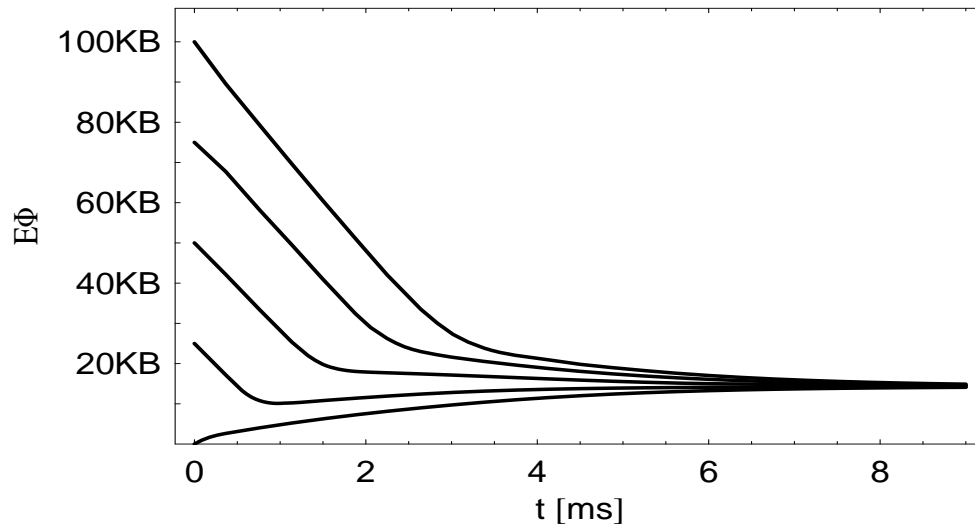
gdzie $R_k(s)$ jest potencjałem (1.3) dla ciągu $A_k(s)$ oraz

$$v_k(s) = \bar{A}_{k+1}(s)(\bar{A}_0(s))^{-1}c_b(s) - c_{b-k}(s),$$

$$c_k(s) = \frac{1}{s} \sum_{i=b-k}^{\infty} (i-b+k)A_i(s) \cdot \mathbf{1} + \sum_{i=b-k}^{\infty} (i-b+k)\bar{D}_i(s) \cdot \mathbf{1},$$

$$y_b(s) = E(s) \sum_{k=0}^{b-1} R_{b-1-k}(s)v_k(s) - (I - Z(s)) \sum_{k=0}^b R_{b-k}(s)v_k(s).$$

Dowód Twierdzenia 3.2. Stosując wzór na prawdopodobieństwo całkowite przy $X(0) = n \in [1, b]$, otrzymujemy następujący układ równań całkowych:

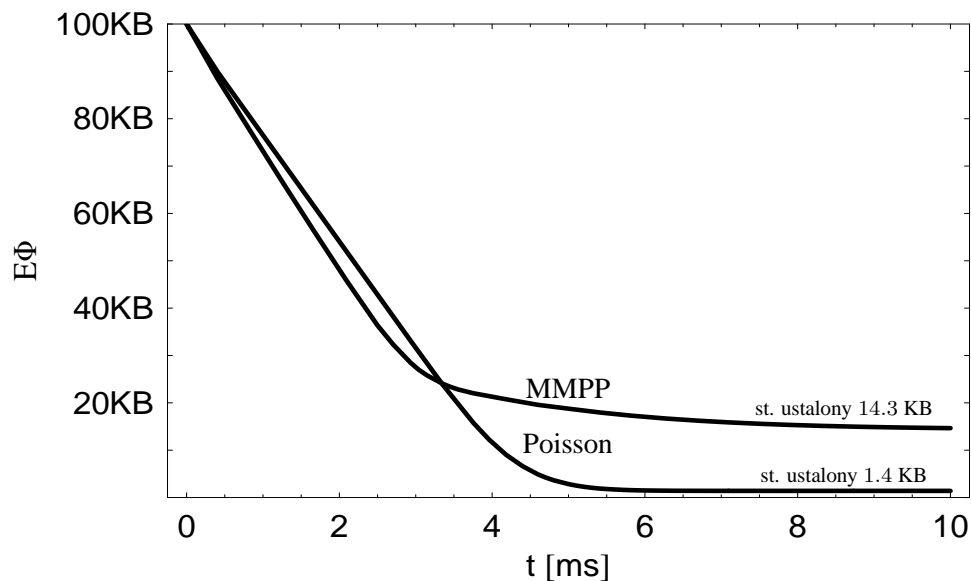


Rys. 3.9. Średnia długość kolejki w funkcji czasu w Przykładzie 3.4 dla 5 różnych początkowych długości kolejki – kolejno od dołu 0 KB, 25 KB, 50 KB, 75 KB, 100 KB. Początkowy stan łańcucha modulującego według rozkładu π , bufor 100 KB

Fig. 3.9. Average queue size as a function of time in Example 3.4 for 5 different initial queue sizes – 0 KB, 25 KB, 50 KB, 75 KB and 100 KB, counting from the bottom. The initial state of the modulating chain was set according to π , the buffer size was 100 KB

$$\begin{aligned}
\Delta_{n,i}(t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \Delta_{n+k-1,j}(t-u) P_{i,j}(k,u) dF(u) \\
&\quad + \sum_{j=1}^m \sum_{k=b-n}^{\infty} \int_0^t (k-b+n + \Delta_{b-1,j}(t-u)) P_{i,j}(k,u) dF(u) \\
&\quad + (1-F(t)) \sum_{j=1}^m \sum_{k=b-n}^{\infty} (k-b+n) P_{i,j}(k,t). \tag{3.29}
\end{aligned}$$

Pierwszy składnik (3.29) odpowiada sytuacji, w której pierwszy czas ukończenia obsługi u wypada wcześniej niż t i do chwili u do systemu przybyło $k < b - n$ pakietów. Drugi składnik (3.29) odpowiada sytuacji, w której pierwszy czas ukończenia obsługi u wypada wcześniej niż t i do chwili u przybyło do systemu $k \geq b - n$ pakietów, z czego $k - b + n$ zostało utraconych. Wreszcie ostatni składnik odpowiada



Rys. 3.10. Średnia długość kolejki w funkcji czasu dla skorelowanego (MMPP) i nieskorelowanego (Poisson) ruchu (Przykład 3.4). Rozmiar bufora 100 KB, początkowa długość kolejki 100KB

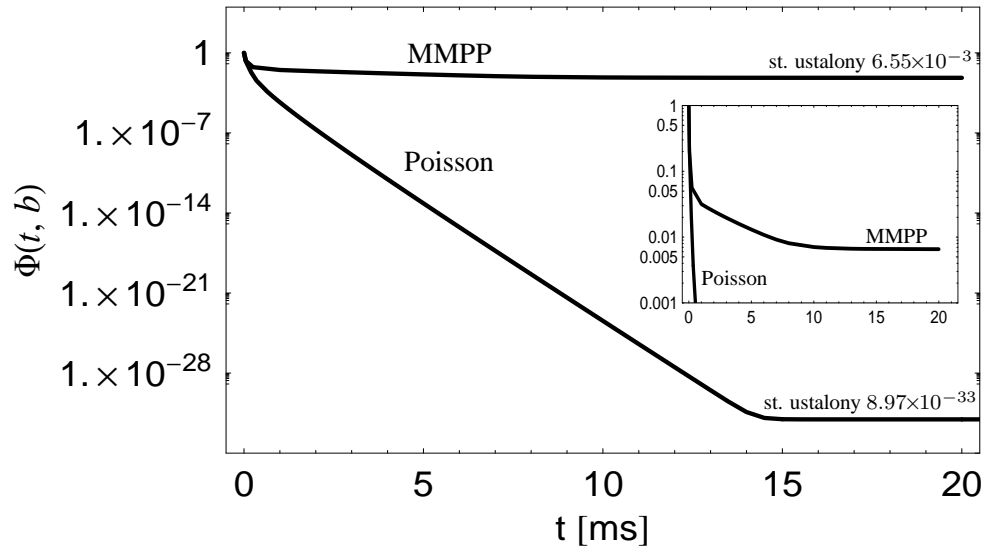
Fig. 3.10. Average queue size as a function of time for correlated (MMPP) and uncorrelated (Poisson) traffic (Example 3.4). The buffer size was 100 KB, the initial queue size was also 100 KB

sytuacji, w której pierwszy czas ukończenia obsługi u wypada po chwili t . Dzieje się tak z prawdopodobieństwem $1 - F(t)$, zaś średnia liczba utraconych pakietów jest wtedy równa

$$\sum_{j=1}^m \sum_{k=b-n}^{\infty} (k - b + n) P_{i,j}(k, t).$$

Jeżeli $X(0) = 0$, to wzór na prawdopodobieństwo całkowite daje

$$\begin{aligned} \Delta_{0,i}(t) &= \sum_{j=1}^m \int_0^t \Delta_{0,j}(t-u) (\lambda_i - Q_{ii}) p_{ij} e^{-(\lambda_i - Q_{ii})u} du \\ &+ \sum_{j=1}^m \int_0^t \Delta_{1,j}(t-u) \Lambda_{ij} e^{-(\lambda_i - Q_{ii})u} du. \end{aligned} \quad (3.30)$$



Rys. 3.11. Prawdopodobieństwo przepełnienia bufora w funkcji czasu dla skorelowanego (MMPP) i nieskorelowanego (Poisson) ruchu (Przykład 3.4). Rozmiar bufora i początkowa długość kolejki 100 KB

Fig. 3.11. Buffer overflow probability as a function of time for correlated (MMPP) and uncorrelated (Poisson) traffic (Example 3.4). The buffer size and the initial queue size were 100 KB

Stosując transformatę Laplace'a i zapis macierzowy do (3.29) i (3.30), otrzymujemy:

$$\delta_n(s) = \sum_{k=0}^{b-n-1} A_k(s) \delta_{n+k-1}(s) + \sum_{k=b-n}^{\infty} A_k(s) \delta_{b-1}(s) + c_n(s),$$

$$0 < n \leq b, \quad (3.31)$$

$$\delta_0(s) = Z(s) \delta_0(s) + E(s) \delta_1(s). \quad (3.32)$$

Zwróćmy teraz uwagę, że (3.31) i (3.32) mają taką samą postać jak (3.20) i (3.21). W związku z tym, pozostała część dowodu przebiega w taki sam sposób jak w dowodzie Twierdzenia 3.1. ■

Wykorzystując Twierdzenie 3.2 oraz własności transformaty Laplace'a, współczynnik strat możemy łatwo obliczyć jako

$$LR = \lim_{s \rightarrow 0^+} \frac{s^2 \delta_{n,i}(s)}{\lambda}, \quad (3.33)$$

gdzie n oraz i są dowolne. W praktyce najkorzystniej jest wybrać $n = b$, gdyż w tym wypadku $\delta_n(s)$ ma najprostszą postać, a mianowicie

$$\delta_b(s) = M_b^{-1}(s)y_b(s).$$

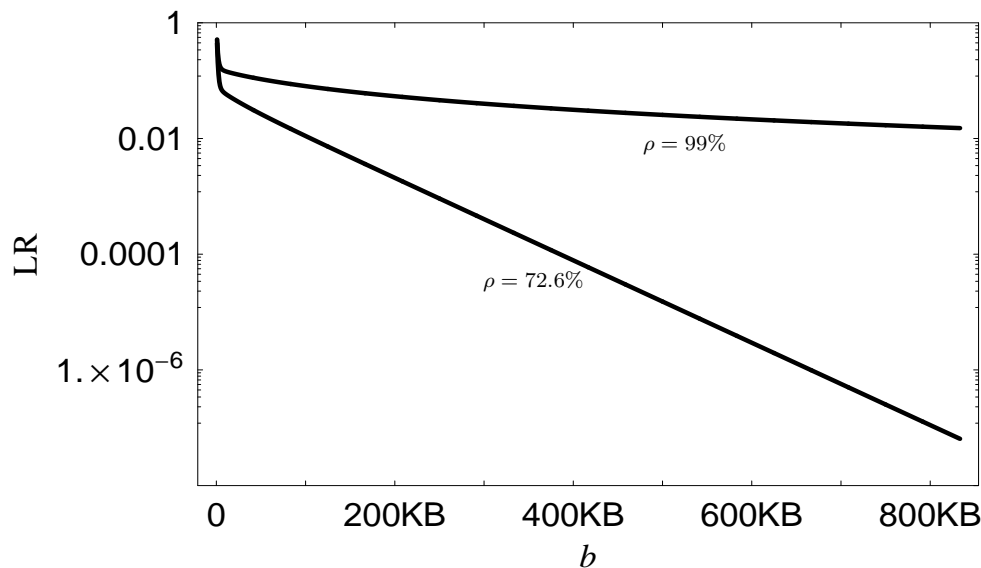
Przykład 3.5. Dla parametryzacji MMPP z Przykładu 3.2, $b = 100 \text{ KB}$ oraz czasu transmisji pakietu $d = 10.133 \mu\text{s}$ ($\rho = 72.6\%$) współczynnik strat w stanie ustalonym, obliczony wg (3.33), wynosi

$$LR = 1.1044\%.$$

Stąd średnia liczba traconych pakietów na jednostkę czasu (ms) wynosi w stanie ustalonym

$$\lambda LR = 0.79219.$$

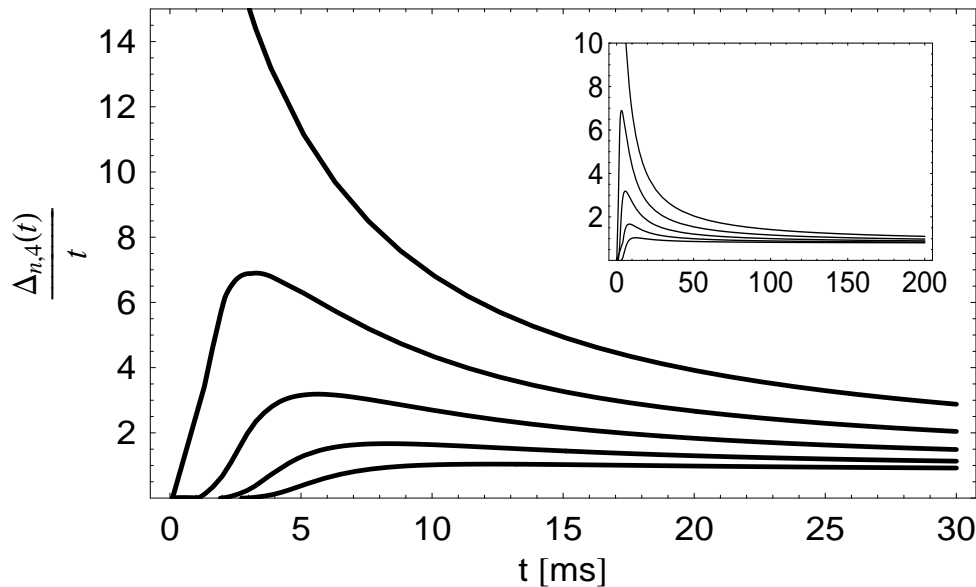
Z praktycznego punktu widzenia, bardzo istotna jest zależność współczynnika strat od rozmiaru bufora. Zależność ta jest przedstawiona na rys. 3.12 dla dwóch różnych obciążeń. Jak się można było spodziewać, wykresy te są malejące. Ponieważ w rozważanym przykładzie występuje dość silna dodatnia autokorelacja w ruchu pakietów, dla dużego obciążenia LR maleje bardzo powoli wraz z rozmiarem bufora, i nawet dla dużych wartości b straty pakietów są wciąż duże.



Rys. 3.12. Współczynnik strat w funkcji rozmiaru bufora dla dwóch różnych obciążeń systemu ρ w Przykładzie 3.5

Fig. 3.12. Loss ratio versus the buffer size for two different values of the offered load ρ in Example 3.5

Prześledźmy teraz, jak wygląda proces strat pakietów w stanie nieustalonym, a w szczególności wpływ $X(0)$ i $J(0)$ na proces strat. Zakładamy dalej $b = 100$ KB oraz obciążenie $\rho = 72.6\%$.

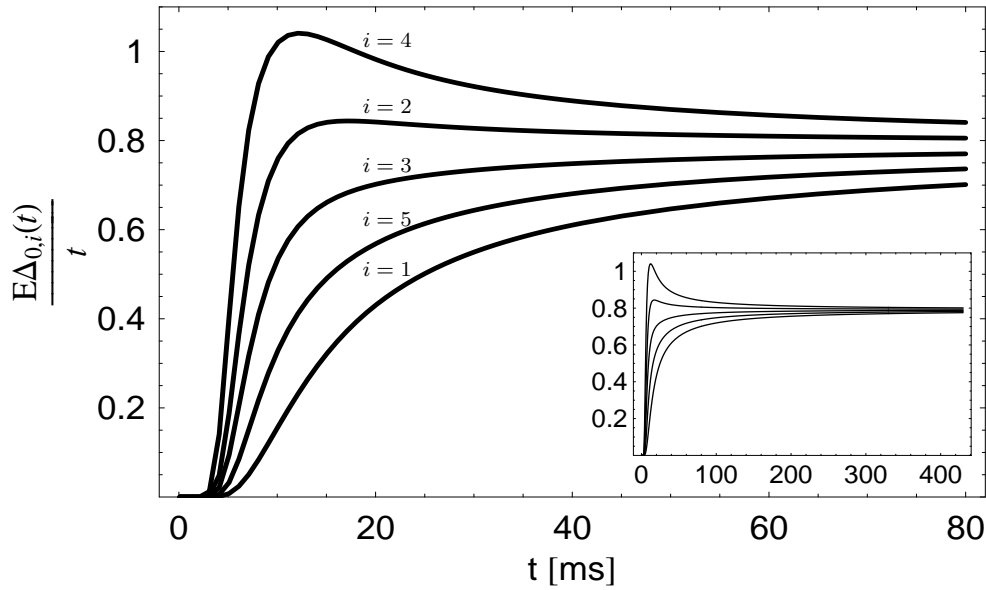


Rys. 3.13. Średnia liczba pakietów traconych w jednostce czasu (ms) w funkcji czasu dla 5 różnych początkowych długości kolejki – kolejno od dołu 0, $0.25b$, $0.5b$, $0.75b$, b . We wszystkich wypadkach przyjęto $J(0) = 4$. Przykład 3.5

Fig. 3.13. Average number of packets lost per millisecond as a function of time for 5 different initial queue sizes – from the bottom 0, $0.25b$, $0.5b$, $0.75b$ and b , counting from the bottom. In every case it was assumed that $J(0) = 4$. Example 3.5

Na rys. 3.13 pokazany jest przebieg funkcji $\Delta_{n,i}(t)/t$ dla ustalonego $i = 4$ oraz różnych wartości n , zaś na rys. 3.14, 3.15, 3.16 pokazany jest przebieg tej funkcji dla ustalonego n oraz różnych wartości i . Funkcja $\Delta_{n,i}(t)/t$ reprezentuje średnią liczbę strat pakietów w jednostce czasu. Oczywiście, dla dowolnych n oraz i funkcja ta zbiega do $LR = 0.79219$, jednak w niektórych wypadkach zbieżność ta może być wolna, np. dla wysokich początkowych długości kolejek czy dla wysokich intensywności MMPP w chwili $t = 0$ (rys. 3.16, szczególnie dla $i = 4$, $i = 2$). Ponadto, w stanie nieustalonym możemy się czasami spodziewać znacznie większych strat (o rząd wielkości) niż w stanie ustalonym (np. rys. 3.13, 3.16).

Ciekawe jest, że funkcja $\Delta_{n,i}(t)/t$ może nie tylko nie być monotoniczna, ale również posiadać więcej niż jedno ekstremum (rys. 3.16, $i = 3$).



Rys. 3.14. Średnia liczba pakietów traconych w jednostce czasu (ms) w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i (Przykład 3.5). Początkowa długość kolejki wynosiła w każdym wypadku 0

Fig. 3.14. Average number of packets lost per millisecond as a function of time for different initial states of the modulating chain i (Example 3.5). In every case the initial queue size was 0

3.2.3. Opóźnienie

Podobnie jak rozkład długości kolejki, rozkład opóźnienia będziemy zapisywać przy pomocy wektora kolumnowego transformat

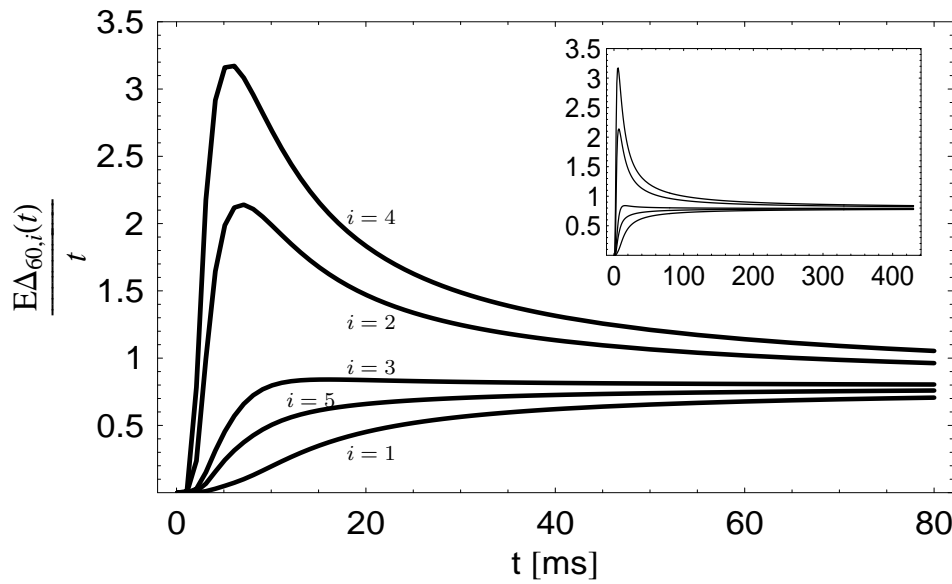
$$w_n(s, \sigma) = (w_{n,1}(s, \sigma), \dots, w_{n,m}(s, \sigma))^T,$$

gdzie

$$w_{n,i}(s, \sigma) = \int_0^\infty e^{-\sigma t} dt \int_0^\infty e^{-sx} \tilde{w}_{n,i}(x, t) dx,$$

oraz

$$\tilde{w}_{n,i}(x, t) = \mathbf{P}(V(t) > x | X(0) = n, J(0) = i). \quad (3.34)$$



Rys. 3.15. Średnia liczba pakietów traconych w jednostce czasu (ms) w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i (Przykład 3.5). Początkowa długość kolejki wynosiła w każdym wypadku 50 KB

Fig. 3.15. Average number of packets lost per millisecond as a function of time for different initial states of the modulating chain i (Example 3.5). In every case the initial queue size was 50 KB

Twierdzenie 3.3. *Transformata Laplace'a rozkładu opóźnienia w systemie MMPP/G/1/b ma postać*

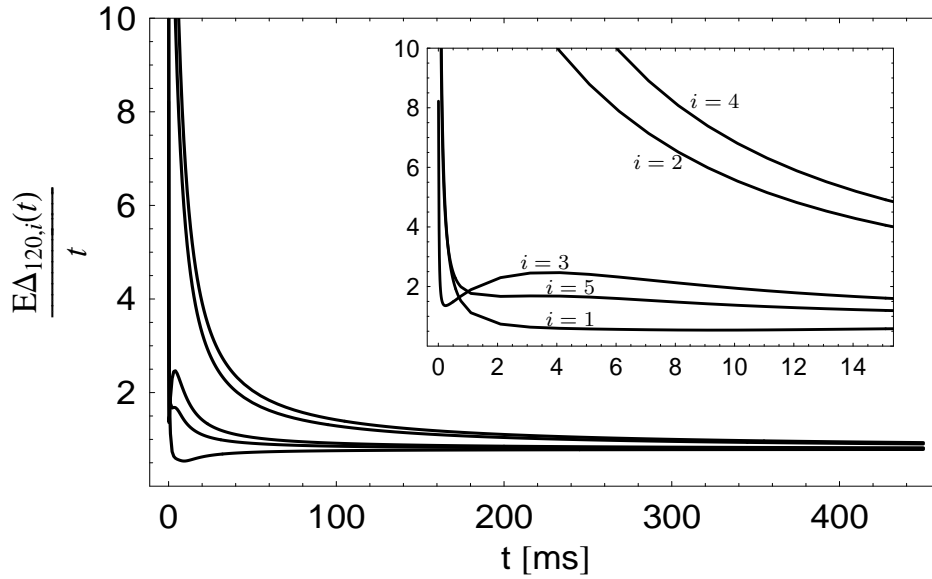
$$w_n(s, \sigma) = \left(R_{b-n+1}(\sigma)A_0(\sigma) + \sum_{k=0}^{b-n} R_{b-n-k}(\sigma)B_k(\sigma) \right) M_b^{-1}(\sigma)u_b(s, \sigma) + \sum_{k=0}^{b-n} R_{b-n-k}(\sigma)h_k(s, \sigma), \quad (3.35)$$

gdzie

$$h_n(s, \sigma) = \bar{A}_{n+1}(\sigma)(\bar{A}_0(\sigma))^{-1}q_b(s, \sigma) - q_{b-n}(s, \sigma),$$

$$q_n(s, \sigma) = \frac{1}{s} \sum_{k=0}^{b-n-1} \left(\bar{D}_k(\sigma) - C_k(s, \sigma)(f(s))^{n+k-1} \right) \cdot \mathbf{1},$$

$$u_b(s, \sigma) = E(\sigma) \sum_{k=0}^{b-1} R_{b-1-k}(\sigma)h_k(s, \sigma) - (I - Z(\sigma)) \sum_{k=0}^b R_{b-k}(\sigma)h_k(s, \sigma).$$



Rys. 3.16. Średnia liczba pakietów traconych w jednostce czasu (ms) w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i (Przykład 3.5). Początkowa długość kolejki wynosiła w każdym wypadku 100 KB

Fig. 3.16. Average number of packets lost per millisecond as a function of time for different initial states of the modulating chain i (Example 3.5). In every case the initial queue size was 100 KB

Dowód Twierdzenia 3.3. Stosując wzór na prawdopodobieństwo całkowite względem pierwszej chwili ukończenia obsługi, dla $0 < n \leq b$ mamy

$$\begin{aligned}
 \tilde{w}_{n,i}(x,t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \tilde{w}_{n+k-1,j}(x,t-u) P_{i,j}(k,u) dF(u) \\
 &+ \sum_{j=1}^m \sum_{k=b-n}^{\infty} \int_0^t \tilde{w}_{b-1,j}(x,t-u) P_{i,j}(k,u) dF(u) \\
 &+ \sum_{j=1}^m \sum_{k=0}^{b-n-1} P_{i,j}(k,t) \int_t^{\infty} (1 - F^{(n+k-1)*}(x-u+t)) dF(u).
 \end{aligned} \tag{3.36}$$

Dla $n = 0$ otrzymujemy:

$$\begin{aligned}\tilde{w}_{0,i}(x,t) &= \sum_{j=1}^m \int_0^t \tilde{w}_{0,j}(x,t-u)(\lambda_i - Q_{ii})p_{ij}e^{-(\lambda_i - Q_{ii})u} du \\ &\quad + \sum_{j=1}^m \int_0^t \tilde{w}_{1,j}(x,t-u)\Lambda_{ij}e^{-(\lambda_i - Q_{ii})u} du.\end{aligned}$$

Stosując transformaty i przechodząc do zapisu macierzowego, dla $0 < n \leq b$ dostajemy

$$w_n(s, \sigma) = \sum_{k=0}^{b-n-1} A_k(\sigma)w_{n+k-1}(s, \sigma) + \sum_{k=b-n}^{\infty} A_k(\sigma)w_{b-1}(s, \sigma) + q_n(s, \sigma), \quad (3.37)$$

oraz dla $n = 0$

$$w_0(s, \sigma) = Z(\sigma)w_0(s, \sigma) + E(\sigma)w_1(s, \sigma). \quad (3.38)$$

Jak widać, postać równań (3.37) i (3.38) jest taka sama jak (3.20) i (3.21), a zatem pozostałe obliczenia przeprowadzamy tak jak w dowodzie Twierdzenia 3.1. ■

Ponieważ w definicji $w_n(s, \sigma)$ występuje ogon rozkładu $V(t)$, do obliczenia średniej wartości opóźnienia należy podstawić $s = 0$ do (3.35). Następnie możemy odwrócić transformatę ze względu na zmienną σ i otrzymać przebieg czasowy średniej wartości opóźnienia. Jeżeli chcemy otrzymać średnią wartość opóźnienia w stanie ustalonym, obliczamy tylko

$$\lim_{\sigma \rightarrow 0^+} \sigma w_n(0, \sigma). \quad (3.39)$$

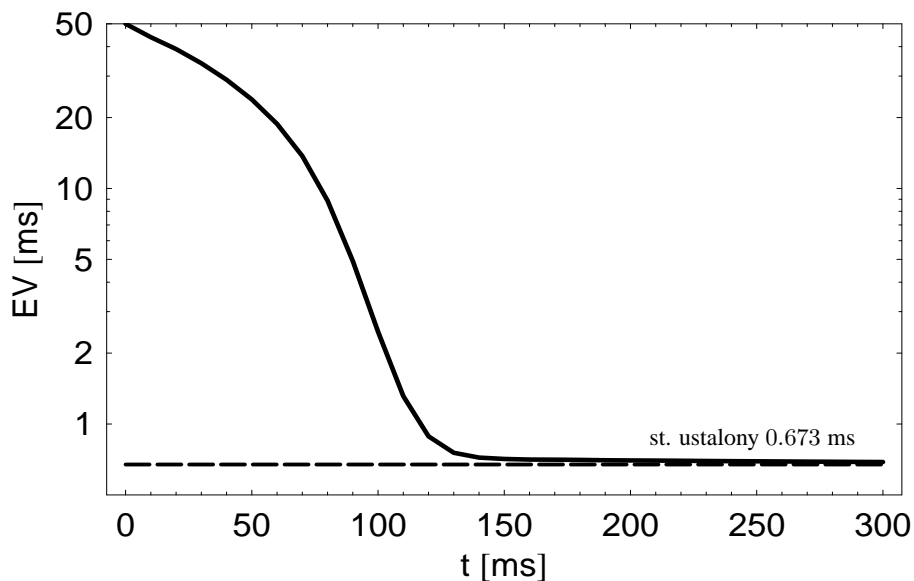
Wreszcie, jeżeli interesuje nas kształt rozkładu opóźnienia w pewnej chwili t , stosujemy algorytm do odwracania podwójnej transformaty Laplace'a (podrozdział 5.1.2).

Przykład 3.6. Dla parametryzacji MMPP z Przykładu 3.1, przyjmując $b = 50$ pakietów oraz stały czas transmisji $d = 1$ ms, średnie opóźnienie w stanie ustalonym wynosi

$$EV = 0.673 \text{ ms.}$$

Rys. 3.17 przedstawia przebieg czasowy średniego opóźnienia, przy założeniu że początkowo bufor był przepełniony. Dość wyraźnie zaznacza się czas, w którym średnie opóźnienie osiąga wartość stanu ustalonego.

Z kolei na rys. 3.18 pokazana jest zależność średniej wartości opóźnienia od rozmiaru bufora zarówno w stanie nieustalonym (linie ciągłe), jak i ustalonym (linia przerywana). Zauważmy, że opóźnienie rośnie wraz ze wzrostem rozmiaru bufora zarówno w stanie nieustalonym, jak i ustalonym.



Rys. 3.17. Średnie opóźnienie kolejkowania w czasie (Przykład 3.6). Początkowo bufor był przepelniony ($X(0) = b$), rozkład początkowy stanu łańcucha modulującego wynosił π

Fig. 3.17. Average queueing delay versus time (Example 3.6). Initially the buffer was full ($X(0) = b$) and the initial distribution of the modulating state was π

Jest to ważna z praktycznego punktu widzenia informacja – przy wyborze rozmiaru bufora należy wziąć pod uwagę fakt, że wzrost rozmiaru bufora powoduje zarówno pozytywne (zmniejszenie współczynnika strat, patrz np. rys. 3.12), jak i negatywne efekty (wzrost opóźnień).

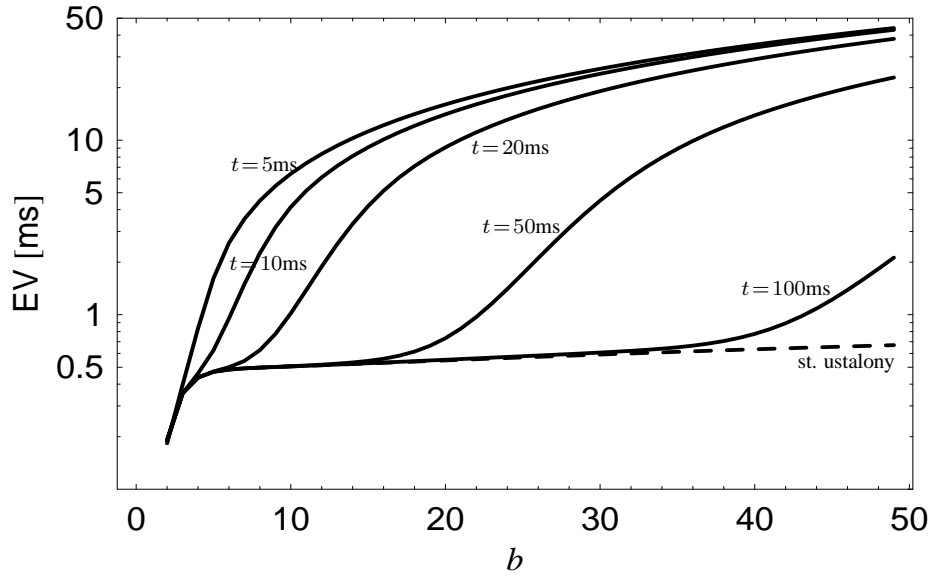
Przykład 3.7. Dla parametryzacji MMPP z Przykładu 3.2, $b = 100$ KB oraz czasu transmisji pakietu $d = 10.133 \mu s$ ($\rho = 72.6\%$), średnie opóźnienie w stanie ustalonym wynosi

$$EV = 0.162 \text{ ms.}$$

Rys. 3.19 pokazuje średnie opóźnienie kolejkowanie w stanie nieustalonym dla 5 różnych wartości $X(0)$. Po raz kolejny widzimy, że krótkoterminowa charakterystyka może przyjmować wartości znacznie różniące się od stanu ustalonego.

3.2.4. Okres przepelnienia bufora

Zajmiemy się najpierw wyznaczeniem rozkładu pierwszego okresu przepelnienia bufora, tzn. rozkładu β_1 według definicji z podrozdziału 2.2.4. Rozkład ten będziemy



Rys. 3.18. Średnie opóźnienie w funkcji rozmiaru bufora w różnych chwilach czasu (Przykład 3.6). W każdym wypadku początkowo bufor był przepełniony ($X(0) = b$) i rozkład początkowy stanu łańcucha modulującego wynosił π

Fig. 3.18. Average queueing delay versus the buffer size (Example 3.6). In every case the buffer was initially full ($X(0) = b$) and the initial distribution of the modulating state was π

zapisywać w postaci wektora kolumnowego dystrybuant, tzn.

$$H_n(t) = \mathbf{1} - \chi_n(t),$$

gdzie

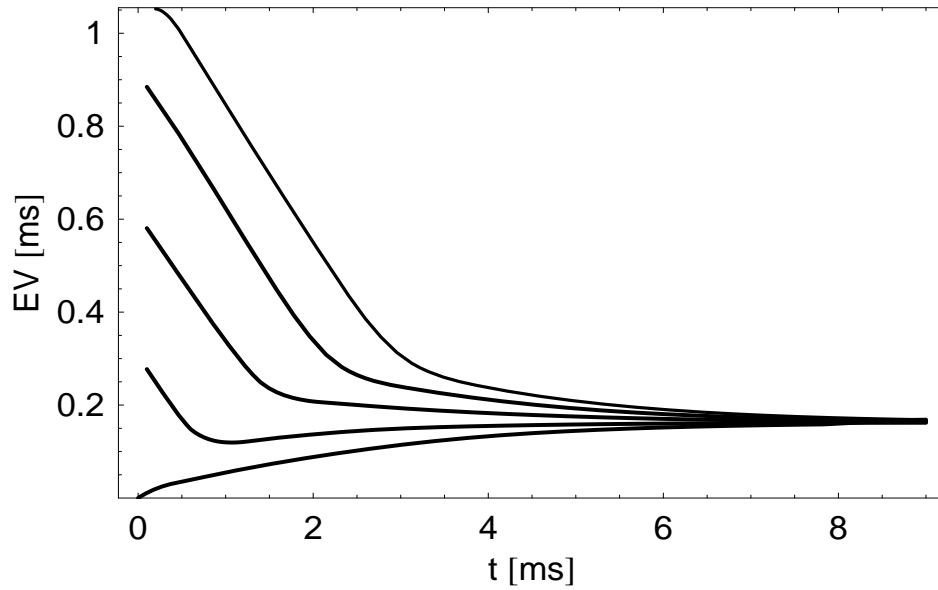
$$\begin{aligned} \chi_n(t) &= (\chi_{n,1}(t), \dots, \chi_{n,m}(t))^T, \\ \chi_{n,i}(t) &= \mathbf{P}(\beta_1 > t | X(0) = n, J(0) = i). \end{aligned}$$

Będziemy często posługiwać się operatorem K_b . Operator ten dla ciągu macierzy kwadratowych lub wektorów kolumnowych $\{X_1, X_2, X_3, \dots\}$ jest zdefiniowany jako

$$K_b(X) = (I - Z) \sum_{k=1}^b R_{b-k} X_k - E \sum_{k=1}^{b-1} R_{b-1-k} X_k, \quad (3.40)$$

gdzie

$$Z = Z(0), \quad E = E(0), \quad R = R(0).$$



Rys. 3.19. Średnie opóźnienie kolejkowania w funkcji czasu (Przykład 3.7) dla różnych początkowych wypełnień bufora – licząc od dołu 0, 0.25b, 0.5b, 0.75b, b. Rozkład początkowy stanu łańcucha modulującego wynosił π

Fig. 3.19. Average queueing delay versus time (Example 3.7) for different initial queue sizes – 0, 0.25b, 0.5b, 0.75b and b, counting from the bottom. The initial distribution of the modulating state was π

Twierdzenie 3.4. *Rozkład długości pierwszego okresu przepelnienia bufora w systemie MMPP/G/1/b ma postać*

$$H_n(t) = \mathbf{1} - \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(g(t)) + \sum_{k=1}^{b-n} R_{b-n-k} g_k(t), \quad 0 \leq n < b, \quad (3.41)$$

gdzie

$$A_k = A_k(0),$$

$$G_b = (I - Z) \sum_{k=0}^b R_{b-k} A_k - E \sum_{k=0}^{b-1} R_{b-1-k} A_k, \quad (3.42)$$

$$g_k(t) = \int_0^\infty (1 - F(u+t)) P(k-1, u) du \cdot \Lambda \cdot \mathbf{1}, \quad (3.43)$$

zaś $K_b(g(t))$ jest zdefiniowane w (3.40)³.

³Oczywiście w (3.40) należy zamienić X_k na $g_k(t)$.

Dowód Twierdzenia 3.4. Ze wzoru na prawdopodobieństwo całkowite względem pierwszego czasu ukończenia obsługi dla $0 < n < b$, $1 \leq i \leq m$ mamy:

$$\begin{aligned} & \mathbf{P}(\beta_1 > t | X(0) = n, J(0) = i) \\ &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\beta_1 > t | X(0) = n+k-1, J(0) = j) P_{ij}(k, u) dF(u) \\ & \quad + \sum_{j=1}^m \int_0^\infty dF(u) \int_0^u I(u-v > t) P_{ij}(b-n-1, v) \lambda_j dv. \end{aligned} \quad (3.44)$$

Pierwszy człon (3.44) dotyczy przypadku, gdy bufor nie przepelnia się do pierwszego czasu ukończenia obsługi u , zaś drugi człon dotyczy przypadku, gdy w pewnej chwili $v < u$ następuje przepelnienie i wtedy, oczywiście, $\beta_1 = u - v$.

Dla $n = 0$ i $1 \leq i \leq m$, warunkując względem pierwszego zdarzenia w strumieniu MMPP (pakiet lub zmiana stanu modulującego), otrzymujemy

$$\begin{aligned} & \mathbf{P}(\beta_1 > t | X(0) = 0, J(0) = i) \\ &= \sum_{j=1}^m p_{ij} \mathbf{P}(\beta_1 > t | X(0) = 0, J(0) = i) \\ & \quad + \sum_{j=1}^m \frac{\Lambda_{ij}}{\lambda_i - Q_{ii}} \mathbf{P}(\beta_1 > t | X(0) = 1, J(0) = i). \end{aligned} \quad (3.45)$$

Całkując drugi człon (3.44) przez części oraz stosując zapis przy pomocy macierzy, otrzymamy

$$\chi_n(t) = \sum_{k=0}^{b-n-1} A_k \chi_{n+k-1}(t) + g_{b-n}(t), \quad (3.46)$$

zaś z (3.45), że

$$\chi_0(t) = Z\chi_0(t) + E\chi_1(t). \quad (3.47)$$

Zamieniając indeksy

$$\varphi_n(t) = \chi_{b-n}(t),$$

dostaniemy

$$\sum_{k=-1}^{n-1} A_{k+1} \varphi_{n-k}(t) - \varphi_n(t) = \psi_n(t), \quad 0 < n < b, \quad (3.48)$$

$$\varphi_b(t) = Z\varphi_b(t) + E\varphi_{b-1}(t), \quad (3.49)$$

gdzie

$$\psi_n(t) = A_n \varphi_1(t) - g_n(t).$$

Do (3.48) możemy zastosować teraz Twierdzenie 1.1, z którego wynika, że

$$\varphi_n(t) = R_n c(t) + \sum_{k=1}^n R_{n-k} \psi_k(t), \quad n \geq 1, \quad (3.50)$$

gdzie $c(t)$ jest wektorem kolumnowym niezależnym od n .

Podstawienie $n = 1$ do (3.50) daje

$$c(t) = A_0 \varphi_1(t),$$

czyli

$$\varphi_n(t) = \sum_{k=0}^n R_{n-k} A_k \varphi_1(t) - \sum_{k=1}^n R_{n-k}(t) g_k(t).$$

Zamieniając z powrotem indeksy $\chi_n(t) = \varphi_{b-n}(t)$, otrzymujemy:

$$\chi_n(t) = \sum_{k=0}^{b-n} R_{b-n-k} A_k \chi_{b-1}(t) - \sum_{k=1}^{b-n} R_{b-n-k}(t) g_k(t). \quad (3.51)$$

Wreszcie, stosując (3.51) i warunek brzegowy (3.47), obliczamy

$$\chi_{b-1}(t) = G_b^{-1} K_b(g(t)), \quad (3.52)$$

co kończy dowód Twierdzenia 3.4. ■

Zbadamy teraz rozkład kolejnych okresów przepełnienia bufora (β_2, β_3, \dots na rys. 2.1).

Rozkłady te dla MMPP wyznacza się trudniej niż w wypadku procesu Poissona z następującego powodu. W wypadku procesu Poissona wszystkie β_k dla $k \geq 2$ mają taki sam rozkład. Jest tak, gdyż tuż po okresie przepełnienia bufora w systemie zawsze jest $b-1$ pakietów i właśnie ta liczba stanowi początkową długość kolejki dla następnego okresu przepełnienia. A zatem dla zwykłego procesu Poissona rozkład β_k , $k \geq 2$ jest taki sam jak rozkład β_1 , przy założeniu że $X(0) = b-1$, tzn.

$$\mathbf{P}(\beta_k < t) = \mathbf{P}(\beta_1 < t | X(0) = b-1) = H_{b-1}(t), \quad k \geq 2 \quad (3.53)$$

i możemy go wyznaczać, korzystając bezpośrednio ze wzoru (2.37).

W wypadku kolejkowania MMPP równość (3.53) nie zachodzi i rozkład β_k jest różny dla różnych k . Dzieje się tak, ponieważ rozkład okresu przepełnienia zależy

teraz nie tylko od długości kolejki tuż po poprzednim okresie przepełnienia (ta jest zawsze $b - 1$), ale i od stanu łańcucha modulującego w tej chwili.

A zatem znajdziemy najpierw rozkład łańcucha modulującego tuż po okresie przepełnienia.

W tym celu zdefiniujmy ciąg zmiennych losowych $\{\omega_k\}_{k=0}^{\infty}$, gdzie $\omega_0 = J(0)$ i ω_k , $k \geq 1$ oznacza stan łańcucha modulującego tuż po k -tym okresie przepełnienia. Jak łatwo zauważyć, rozkład β_1 zależy od ω_0 i $X(0)$, zaś rozkład β_k , $k \geq 2$ zależy tylko od ω_{k-1} . A zatem ciąg $\{\omega_k\}_{k=0}^{\infty}$ tworzy łańcuch Markowa z dyskretnym czasem.

Rozkład ω_1 będziemy zapisywać w zależności od $J(0)$ przy pomocy następującej macierzy $m \times m$:

$$S_n = \left[\mathbf{P}(\omega_1 = l | X(0) = n, J(0) = i) \right]_{i,l}, \quad i, l = 1, \dots, m.$$

Twierdzenie 3.5. *Rozkład stanu łańcucha modulującego tuż po pierwszym okresie przepełnienia ma w systemie MMPP/G/1/b postać*

$$S_n = \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(\bar{A}) - \sum_{k=1}^{b-n} R_{b-n-k} \bar{A}_k, \quad 0 \leq n < b, \quad (3.54)$$

gdzie

$$\bar{A}_k = \sum_{i=k}^{\infty} A_i, \quad (3.55)$$

zaś $K_b(\bar{A})$ jest zdefiniowane w (3.40). W szczególności mamy

$$S_{b-1} = G_b^{-1} K_b(\bar{A}).$$

Dowód Twierdzenia 3.5. Wzór na prawdopodobieństwo całkowite względem pierwszej chwili ukończenia obsługi dla $0 < n < b$, $1 \leq i \leq m$ daje:

$$\begin{aligned} & \mathbf{P}(\omega_1 = l | X(0) = n, J(0) = i) \\ &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^{\infty} \mathbf{P}(\omega_1 = l | X(0) = n+k-1, J(0) = j) P_{ij}(k, u) dF(u) \\ & \quad + \sum_{j=1}^m \int_0^{\infty} dF(u) \int_0^u \mathbf{P}(J(u-v) = l | J(0) = j) P_{ij}(b-n-1, v) \lambda_j dv. \end{aligned} \quad (3.56)$$

Pierwszy składnik (3.56) odpowiada za sytuację, w której do chwili pierwszego ukończenia obsługi nie wystąpiło przepełnienie bufora, tzn. liczba przybyłych w tym czasie pakietów była mniejsza niż $b-n$. Drugi składnik (3.56) odpowiada za sytuację, w której w chwili v , wcześniejszej niż u , następuje przepełnienie bufora. Wówczas mamy

$$\mathbf{P}(\omega_1 = l) = \sum_j \mathbf{P}(J(u-v) = l | J(0) = j).$$

Warunek brzegowy dla $X(0) = 0$ ma teraz postać

$$\begin{aligned} \mathbf{P}(\omega_1 = l | X(0) = 0, J(0) = i) &= \sum_{j=1}^m p_{ij} \mathbf{P}(\omega_1 = l | X(0) = 0, J(0) = i) \\ &+ \sum_{j=1}^m \frac{\Lambda_{ij}}{\lambda_i - Q_{ii}} \mathbf{P}(\omega_1 = l | X(0) = 1, J(0) = i). \end{aligned} \quad (3.57)$$

Stosując eksponentę macierzy do zapisu stanu łańcucha Markowa po czasie $u-v$ (zobacz np. [196]), a następnie stosując zapis macierzowy do (3.56) i (3.57), otrzymujemy:

$$S_n = \sum_{k=0}^{b-n-1} A_k S_{n+k-1} + \bar{A}_{b-n}, \quad 0 < n < b, \quad (3.58)$$

$$S_0 = ZS_0 + ES_1, \quad (3.59)$$

gdzie

$$\bar{A}_k = \int_0^\infty dF(u) \int_0^u P(k-1, v) \Lambda e^{Q(u-v)} dv = \sum_{i=k}^\infty A_i$$

Teraz dowód możemy skończyć tak samo jak dowód Twierdzenia 3.4. ■

Posługując się Twierdzeniami 3.4 i 3.5, możemy teraz łatwo znaleźć rozkład β_k dla dowolnego $k \geq 2$. Wystarczy zauważyć, że S_n jest macierzą przejścia łańcucha Markowa ω_k w pierwszym kroku ($\omega_0 \rightarrow \omega_1$), zaś S_{b-1} jest macierzą przejścia łańcucha ω_k w każdym kolejnym kroku ($\omega_1 \rightarrow \omega_2, \omega_2 \rightarrow \omega_3$, itd.). Wykorzystując ten fakt, otrzymujemy następujący wniosek.

Wniosek 3.1. *Dla $k \geq 2$ rozkład k -tego okresu przepełnienia bufora w systemie MMPP/G/1/b ma postać*

$$\mathbf{P}(\beta_k < t) = v S_n (S_{b-1})^{k-2} H_{b-1}(t), \quad (3.60)$$

gdzie wektor v oznacza początkowy rozkład łańcucha modulującego (rozkład $J(0)$), n oznacza początkową długość kolejki, zaś S_n i $H_n(t)$ są określone odpowiednio wzorami (3.54) i (3.41).

Posługując się Wnioskiem 3.1 możemy znaleźć rozkład okresu przepełnienia bufora w stanie ustalonym. Wygląda on następująco.

Wniosek 3.2. Rozkład graniczny β_k ma postać

$$\lim_{k \rightarrow \infty} \mathbf{P}(\beta_k < t) = w H_{b-1}(t), \quad (3.61)$$

gdzie w oznacza rozkład stacjonarny dla macierzy S_{b-1} , tzn. wektor spełniający warunki

$$w S_{b-1} = w, \quad w \mathbf{1} = 1.$$

Przykład 3.8. Zbadamy teraz okres przepełnienia bufora dla parametryzacji MMPP z Przykładu 3.1, $b = 50$ pakietów oraz stałego czasu transmisji $d = 1$ ms. Dla wizualizacji rozkładu prawdopodobieństwa zamiast dystrybuanty lepiej użyć funkcji gęstości, tzn.

$$\begin{aligned} h_n(t) &= (h_{n,1}(t), \dots, h_{n,m}(t))^T = \frac{dH_n(t)}{dt} \\ &= - \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(g'(t)) + \sum_{k=1}^{b-n} R_{b-n-k} g'_k(t). \end{aligned}$$

Rys. 3.20 przedstawia zależność gęstości pierwszego okresu przepełnienia od rozmiaru bufora dla dwóch różnych stanów początkowych łańcucha modulującego. W obu wypadkach widać bardzo szybką zbieżność do rozkładu granicznego, tzn. już dla małych b rozkłady zlewają się z granicznymi. Zbieżność ta jest również widoczna na rys. 3.21 przedstawiającym średni czas pierwszego przepełnienia bufora. Chcielibyśmy oczywiście umieć wyznaczać postać rozkładu granicznego dla dużych b , tzn. poznać jakiś analog Twierdzenia 2.10 dla procesu MMPP. Można się spodziewać, że również w tym wypadku wzór graniczny będzie miał prostszą formę. Jego zakres stosowania byłby bardzo szeroki, tzn. już do małych b , jak pokazano powyżej. Do tej pory jednak nie udało się takiego wzoru wyprowadzić.

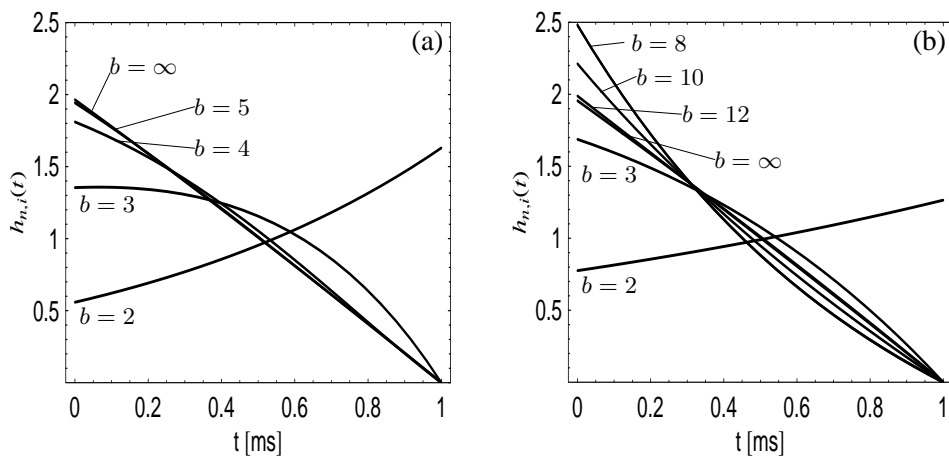
Na rys. 3.22 pokazana jest zależność gęstości pierwszego okresu przepełnienia od początkowej długości kolejki.

Wreszcie, na rys. 3.23 zaprezentowane są gęstości dla kolejnych okresów przepełnienia w stanie nieustalonym, tzn. $\beta_1, \beta_2, \beta_3, \dots$ oraz w stanie ustalonym, tzn. β_∞ . Do otrzymania tych gęstości konieczne jest wyznaczenie macierzy S_{b-1} przy pomocy Twierdzenia 3.5. Macierz ta ma postać

$$S_{b-1} = \begin{bmatrix} 0.99920 & 0.00080 \\ 0.50980 & 0.49020 \end{bmatrix},$$

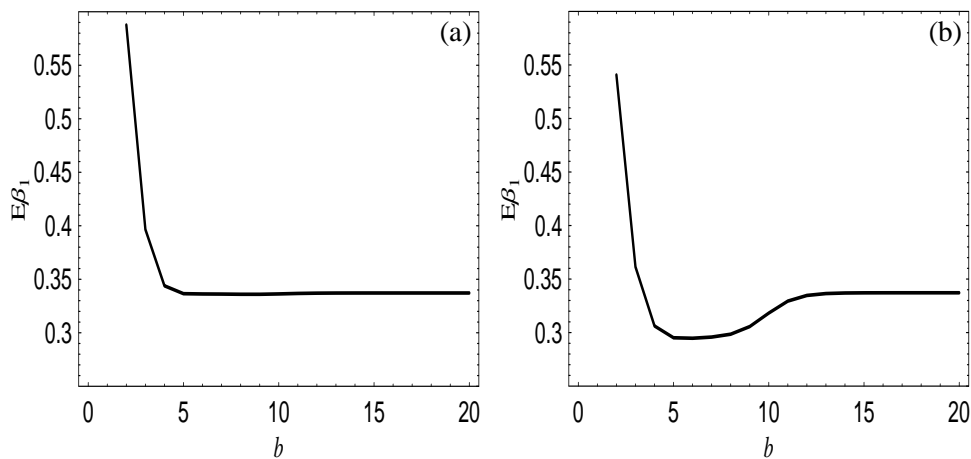
zaś wektor stacjonarny dla S_{b-1}

$$w = (0.99844, 0.00156).$$



Rys. 3.20. Gęstość rozkładu prawdopodobieństwa pierwszego okresu przepełnienia bufora ($h_{n,i}(t)$) dla różnych rozmiarów buforów w Przykładzie 3.8. Część (a): $i = 1$, $n = 0$. Część (b): $i = 2$, $n = 0$

Fig. 3.20. Probability density of the first buffer overflow period ($h_{n,i}(t)$) for different buffer sizes in Example 3.8. Part (a): $i = 1$, $n = 0$. Part (b): $i = 2$, $n = 0$



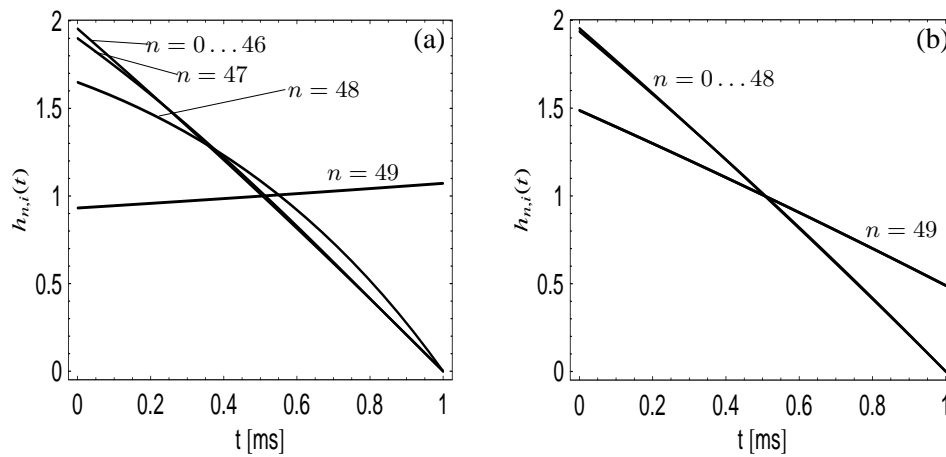
Rys. 3.21. Średnia długość pierwszego okresu przepełnienia bufora w funkcji rozmiaru bufora w Przykładzie 3.8. Część (a): $i = 1$, $n = 0$. Część (b): $i = 2$, $n = 0$

Fig. 3.21. Average duration of the first buffer overflow period versus the buffer size in Example 3.8. Part (a): $i = 1$, $n = 0$. Part (b): $i = 2$, $n = 0$

Porównajmy teraz wektory w oraz

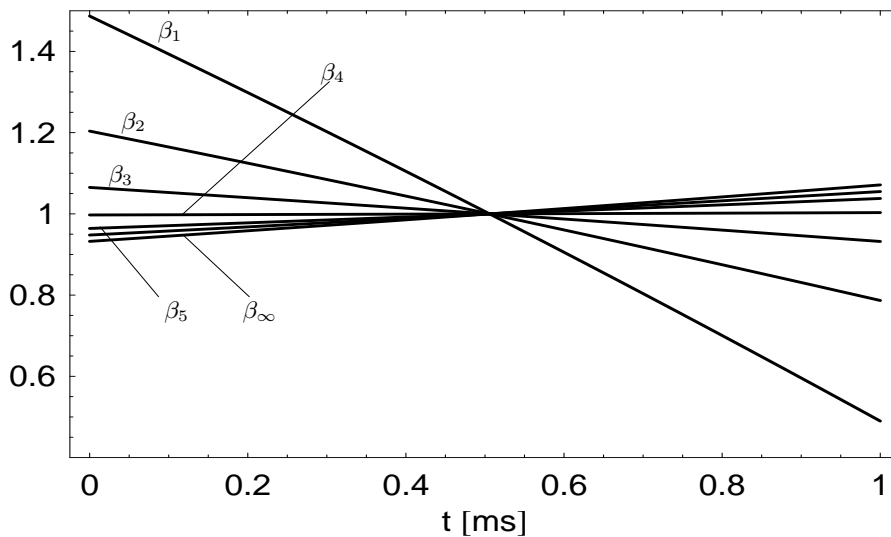
$$\pi = (0.00589, 0.99411).$$

Patrząc na π , widzimy, że stan łańcucha modulującego wynosi najczęściej 2. Mimo to stan łańcucha modulującego na końcu okresu przepełnienia bufora wynosi najczęściej 1, co wy-



Rys. 3.22. Gęstość rozkładu prawdopodobieństwa pierwszego okresu przepełnienia bufora ($h_{n,i}(t)$) dla różnych n w Przykładzie 3.8. Część (a): $i = 1$, $b = 50$. Część (b): $i = 2$, $b = 50$

Fig. 3.22. Probability density of the first buffer overflow period ($h_{n,i}(t)$) for different values of n in Example 3.8. Part (a): $i = 1$, $b = 50$. Part (b): $i = 2$, $b = 50$



Rys. 3.23. Gęstości rozkładu prawdopodobieństwa dla kolejnych okresów przepełnienia bufora w Przykładzie 3.8. Parametry początkowe systemu: $J(0) = 2$, $X(0) = 49$, $b = 50$

Fig. 3.23. Probability densities for consecutive buffer overflow periods in Example 3.8. Initial settings: $J(0) = 2$, $X(0) = 49$, $b = 50$

nika z postaci w . Jest to konsekwencją faktu, że przepełnienia bufora następują najczęściej, gdy łańcuch modulujący jest w stanie 1, gdyż $\lambda_1 > \lambda_2$.

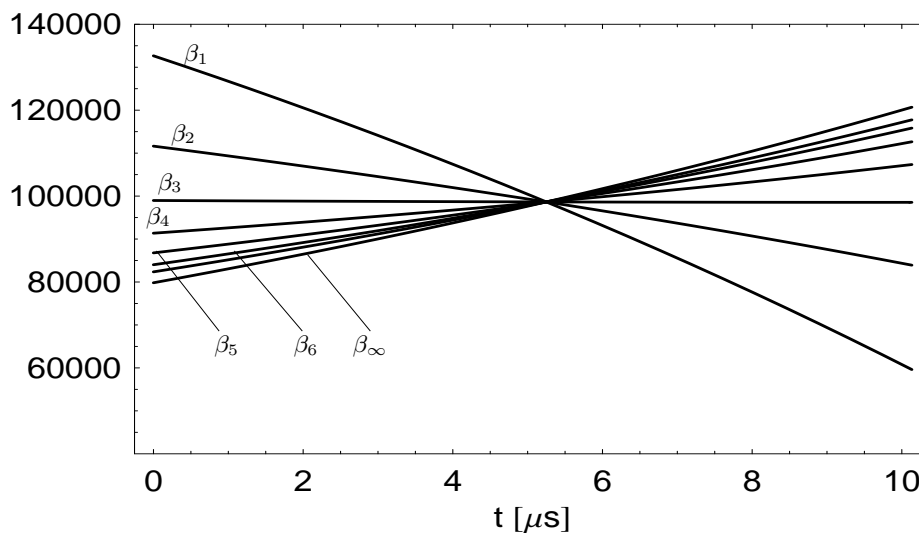
Przykład 3.9. Dla parametryzacji MMPP z Przykładu 3.2, $b = 120$ pakietów oraz czasu transmisji pakietu $d = 10.133 \mu s$ ($\rho = 72.6\%$) macierz S_{b-1} wygląda następująco

$$S_{b-1} = \begin{bmatrix} 0.60242 & 0.09548 & 0.00020 & 0.29941 & 0.00247 \\ 0.00021 & 0.97342 & 0.00054 & 0.02204 & 0.00377 \\ 0.00260 & 0.24492 & 0.07887 & 0.66704 & 0.00657 \\ 0.00026 & 0.00677 & 0.00033 & 0.98956 & 0.00305 \\ 0.00398 & 0.13003 & 0.00014 & 0.30881 & 0.55703 \end{bmatrix},$$

zaś

$$w = (0.00070, 0.23479, 0.00041, 0.75684, 0.00723).$$

Gęstości $\beta_1, \beta_2, \beta_3, \dots$ oraz β_∞ są przedstawione na rys. 3.24. Chociaż zbieżność do β_∞ jest dość szybka, dla kilku pierwszych okresów przepełnienia rozkład może się dość znacznie różnić od rozkładu dla stanu ustalonego.

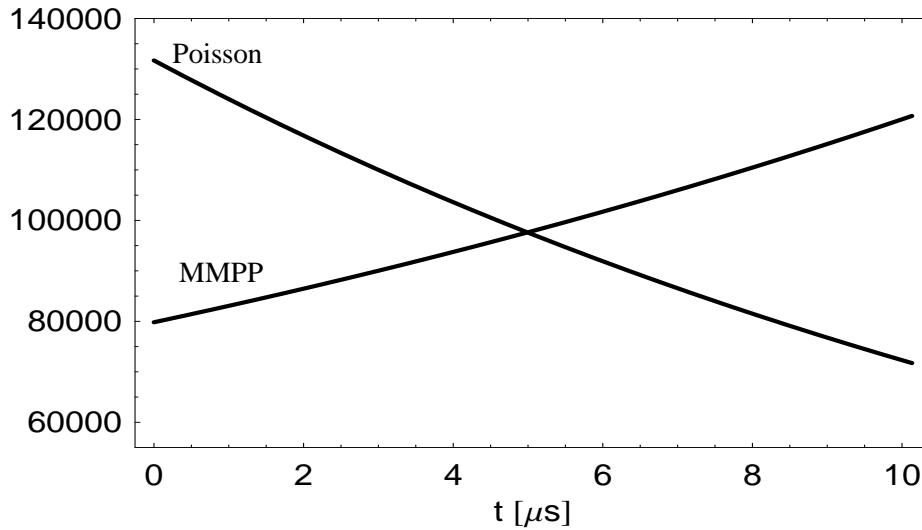


Rys. 3.24. Gęstości rozkładu prawdopodobieństwa dla kolejnych okresów przepełnienia bufora w Przykładzie 3.9. Parametry początkowe: $J(0) = 1, X(0) = b - 1$

Fig. 3.24. Probability densities for consecutive buffer overflow periods in Example 3.9. Initial settings: $J(0) = 1, X(0) = b - 1$

Zbadamy teraz, czy autokorelacja w strumieniu pakietów ma wpływ na długość okresu przepełnienia. Intuicyjnie jest to trudniejsze do rozstrzygnięcia niż np. w wypadku długości kolejki. Rozważmy więc dla porównania strumień Poissona o takiej samej intensywności jak MMPP. Rys. 3.25 przedstawia porównanie gęstości czasu przepełnienia bufora dla tych

dwóch modeli ruchu. Widać, że dla MMPP masa probabilistyczna skoncentrowana jest wokół większych wartości niż w wypadku procesu Poissona. A zatem również na tę charakterystykę kolejkowania dodatnia autokorelacja ma negatywny wpływ.



Rys. 3.25. Gęstości rozkładu prawdopodobieństwa okresu przepełnienia bufora w stanie ustalonym dla kolejek ze skorelowanym (MMPP) i nieskorelowanym (Poisson) strumieniem pakietów (Przykład 3.8)

Fig. 3.25. Probability densities for the buffer overflow period in steady state for queues with correlated (MMPP) and uncorrelated (Poisson) traffic in Example 3.8

3.2.5. Liczba strat w okresie przepełnienia bufora

Dla zwykłego procesu Poissona prawdopodobieństwo, że w k -tym okresie przepełnienia zostanie utraconych i pakietów, można łatwo policzyć, i dla $k \geq 2$ wynosi ono

$$\int_0^{\infty} \frac{e^{-\lambda t} (\lambda t)^i}{i!} dH_{b-1}(t), \quad (3.62)$$

zaś dla $k = 1$

$$\int_0^{\infty} \frac{e^{-\lambda t} (\lambda t)^i}{i!} dH_n(t), \quad (3.63)$$

gdzie $n = X(0)$.

W wypadku kolejkowania MMPP nie możemy zastosować wzorów (3.62), (3.63) ani nawet żadnych ich analogów. Jest tak dlatego, że nie znamy rozkładu łańcucha

modulującego na początku okresu przepelnienia, a od tego stanu zależy zachowanie MMPP w okresie przepelnienia i, w konsekwencji, liczba strat.

Problem ten możemy rozwiązać, obliczając liczbę kolejnych strat pakietów podobnie jak poprzednie charakterystyki kolejkowania.

Przez γ_k oznaczamy liczbę pakietów utraconych w k -tym okresie przepelnienia oraz

$$\begin{aligned} q_{n,i}(l) &= \mathbf{P}(\gamma_1 = l | X(0) = n, J(0) = i), \\ q_n(l) &= (q_{n,1}(l), \dots, q_{n,m}(l))^T. \end{aligned} \quad (3.64)$$

Twierdzenie 3.6. *Rozkład liczby pakietów utraconych w pierwszym okresie przepelnienia bufora w systemie MMPP/G/1/b ma postać*

$$q_n(l) = \sum_{k=0}^{b-n} R_{b-n-k} A_k G_b^{-1} K_b(v(l)) - \sum_{k=1}^{b-n} R_{b-n-k} v_k(l), \quad 0 \leq n < b, \quad (3.65)$$

gdzie

$$v_k(l) = A_{k+l} \cdot \mathbf{1},$$

zaś $K_b(v(l))$ jest określone w (3.40).

Dowód Twierdzenia 3.6. Dla $0 < n < b$, $1 \leq i \leq m$ otrzymujemy

$$\begin{aligned} &\mathbf{P}(\gamma_1 = l | X(0) = n, J(0) = i) \\ &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^\infty \mathbf{P}(\gamma_1 = l | X(0) = n+k-1, J(0) = j) P_{ij}(k, u) dF(u) \\ &\quad + \sum_{j=1}^m \int_0^\infty P_{ij}(b-n+l, u) dF(u), \end{aligned}$$

zaś dla $n = 0$ i $1 \leq i \leq m$

$$\begin{aligned} &\mathbf{P}(\gamma_1 = l | X(0) = 0, J(0) = i) \\ &= \sum_{j=1}^m p_{ij} \mathbf{P}(\gamma_1 = l | X(0) = 0, J(0) = i) \\ &\quad + \sum_{j=1}^m \frac{\Lambda_{ij}}{\lambda_i - Q_{ii}} \mathbf{P}(\gamma_1 = l | X(0) = 1, J(0) = i). \end{aligned}$$

W zapisie macierzowym równania te wyglądają następująco:

$$q_n(l) = \sum_{k=0}^{b-n-1} A_k q_{n+k-1}(l) + v_{b-n}(l), \quad 0 < n < b, \quad (3.66)$$

$$q_0(l) = Zq_0(l) + Eq_1(l). \quad (3.67)$$

Można je rozwiązać tak samo jak w dowodzie Twierdzenia 3.4. ■

Wykorzystując ostatni wynik oraz macierz S_{b-1} , możemy teraz łatwo określić liczbę kolejnych strat w następnych okresach przepełnienia i w stanie ustalonym.

Wniosek 3.3. *Dla $k \geq 2$ rozkład liczby pakietów utraconych w k -tym okresie przepełnienia bufora w systemie MMPP/G/1/b ma postać*

$$\mathbf{P}(\gamma_k = l) = vS_n(S_{b-1})^{k-2}q_{b-1}(l), \quad (3.68)$$

gdzie wektor v oznacza początkowy rozkład łańcucha modulującego (rozkład $J(0)$), n oznacza początkową długość kolejki, S_n i $q_n(l)$ są określone odpowiednio w (3.54) i (3.65). Ponadto, rozkład liczby pakietów utraconych w okresie przepełnienia w stanie ustalonym ma postać

$$\lim_{k \rightarrow \infty} \mathbf{P}(\gamma_k = l) = wq_{b-1}(l), \quad (3.69)$$

gdzie wektor w oznacza rozkład stacjonarny dla macierzy S_{b-1} .

Przykład 3.10. *Dla parametryzacji MMPP z Przykładu 3.1, $b = 50$ pakietów oraz stałego czasu transmisji $d = 1$ ms prawdopodobieństwa utraty l kolejnych pakietów w k -tym okresie przepełnienia przedstawia tab. 3.2.*

Przykład 3.11. *Dla parametryzacji MMPP z Przykładu 3.2, $b = 100$ KB oraz czasu transmisji pakietu $d = 10.133 \mu s$, prawdopodobieństwa utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym zawiera tab. 3.3. Dla porównania pokazane są też wyniki dla procesu Poissona o identycznej intensywności. Zaobserwowane różnice są, oczywiście, konsekwencją różnic w rozkładzie okresu przepełnienia bufora pokazanych w Przykładzie 3.9. Różnice te są statystycznie istotne, np. prawdopodobieństwo utraty kolejno 5 pakietów jest o rząd wielkości większe w modelu uwzględniającym autokorelację.*

$l \backslash k$	1	2	3	4	∞
0	7.5377×10^{-1}	6.7843×10^{-1}	6.4156×10^{-1}	6.2352×10^{-1}	6.0622×10^{-1}
1	1.9772×10^{-1}	2.3715×10^{-1}	2.5644×10^{-1}	2.6589×10^{-1}	2.7494×10^{-1}
2	4.0314×10^{-2}	6.5763×10^{-2}	7.8217×10^{-2}	8.4313×10^{-2}	9.0155×10^{-2}
3	6.9565×10^{-3}	1.5120×10^{-2}	1.9116×10^{-2}	2.1072×10^{-2}	2.2946×10^{-2}
4	1.0618×10^{-3}	2.9472×10^{-3}	3.8699×10^{-3}	4.3215×10^{-3}	4.7544×10^{-3}
5	1.4607×10^{-4}	4.9485×10^{-4}	6.6554×10^{-4}	7.4908×10^{-4}	8.2915×10^{-4}
6	1.8237×10^{-5}	7.2606×10^{-5}	9.9214×10^{-5}	1.1223×10^{-4}	1.2471×10^{-4}
7	2.0731×10^{-6}	9.4304×10^{-6}	1.3031×10^{-5}	1.4793×10^{-5}	1.6482×10^{-5}
8	2.1530×10^{-7}	1.0968×10^{-7}	1.5282×10^{-6}	1.7394×10^{-6}	1.9417×10^{-6}

Tab. 3.2. Prawdopodobieństwa utraty l kolejnych pakietów w k -tym okresie przepełnienia bufora w Przykładzie 3.10. $J(0) = 2$, $X(0) = b - 1$

l	MMPP	Poisson
0	6.8230×10^{-1}	8.0571×10^{-1}
1	2.3465×10^{-1}	1.6359×10^{-1}
2	6.4755×10^{-2}	2.6638×10^{-2}
3	1.4826×10^{-2}	3.5958×10^{-3}
4	2.8914×10^{-3}	4.1307×10^{-4}
5	4.9020×10^{-4}	4.1217×10^{-5}
6	7.3432×10^{-5}	3.6307×10^{-6}
7	9.8470×10^{-6}	2.8608×10^{-7}
8	1.1947×10^{-6}	2.0379×10^{-8}
9	1.3233×10^{-7}	1.3242×10^{-9}
10	1.3480×10^{-8}	7.9087×10^{-11}

Tab. 3.3. Prawdopodobieństwa utraty l kolejnych pakietów w okresie przepełnienia bufora (w stanie ustalonym) dla skorelowanego (MMPP) i nieskorelowanego (Poisson) modelu ruchu (Przykład 3.11)

3.2.6. Czas do przepełnienia bufora

Czas do przepełnienia bufora zależy od początkowej długości kolejki n oraz od początkowego stanu łańcucha modulującego i i jest zdefiniowany jako

$$\tau_{n,i} = \inf\{t > 0 : X(t) = b | X(0) = n, J(0) = i\}.$$

W praktyce interesuje nas najczęściej $\tau_{0,i}$, tzn. czas do przepełnienia bufora, przy założeniu że początkowo system jest pusty.

Rozkład czasu do przepełnienia bufora wyznaczymy w postaci wektora kolumnowego transformat:

$$l_n(s) = (l_{n,1}(s), l_{n,2}(s), \dots, l_{n,m}(s))^T,$$

gdzie

$$l_{n,i}(s) = \int_0^\infty e^{-st} \mathbf{P}(\tau_{n,i} > t) dt.$$

Twierdzenie 3.7. Transformata rozkładu czasu do przepełnienia bufora w systemie MMPP/G/1/b ma postać

$$l_n(s) = \sum_{k=0}^{b-n} R_{b-n-k}(s) A_k(s) G_b^{-1}(s) h_b(s) - \sum_{k=1}^{b-n} R_{b-n-k}(s) \tilde{d}_k(s), \quad 0 \leq n < b, \quad (3.70)$$

gdzie

$$\begin{aligned} \tilde{d}_k(s) &= \sum_{i=0}^{k-1} \bar{D}_i(s) \cdot \mathbf{1}, \\ G_b(s) &= (I - Z(s)) \sum_{k=0}^b R_{b-k}(s) A_k(s) - E(s) \sum_{k=0}^{b-1} R_{b-1-k}(s) A_k(s), \end{aligned} \quad (3.71)$$

$$\begin{aligned} h_b(s) &= (I - Z(s)) \sum_{k=1}^b R_{b-k}(s) \tilde{d}_k(s) - E(s) \sum_{k=1}^{b-1} R_{b-1-k}(s) \tilde{d}_k(s) - z(s), \end{aligned} \quad (3.72)$$

zaś $R_k(s)$ jest potencjałem dla ciągu $A_k(s)$.

Dowód Twierdzenia 3.7. Jeżeli początkowo system nie jest pusty, $0 < n < b$, to stosując wzór na prawdopodobieństwo całkowite względem pierwszej chwili

ukończenia obsługi, otrzymujemy

$$\begin{aligned} \mathbf{P}(\tau_{n,i} > t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \mathbf{P}(\tau_{n+k-1,j} > t-u) P_{i,j}(k,u) dF(u) \\ &+ (1-F(t)) \sum_{j=1}^m \sum_{k=0}^{b-n-1} P_{i,j}(k,t). \end{aligned} \quad (3.73)$$

Pierwsza część (3.73) odpowiada przypadkowi, w którym pierwszy czas ukończenia obsługi u wypada przed t i do czasu ukończenia obsługi przybywa najwyżej $b-n-1$ pakietów (więc bufor się nie przepełnia). W sytuacji gdy do chwili $u < t$ do systemu przybywa $b-n$ lub więcej pakietów, przepełnienie następuje i

$$\mathbf{P}(\tau_{n,i} > t) = 0,$$

więc odpowiedni człon nie występuje w równaniu. Druga część (3.73) odpowiada przypadkowi, gdy pierwszy czas ukończenia obsługi wypada po chwili t i do chwili t przybywa najwyżej $b-n-1$ pakietów. Wtedy

$$\mathbf{P}(\tau_{n,i} > t) = 1.$$

Jeżeli początkowo system jest pusty, to ze wzoru na prawdopodobieństwo całkowite względem chwili pierwszego zdarzenia (tzn. pakietu lub zmiany stanu łańcucha) otrzymujemy

$$\begin{aligned} \mathbf{P}(\tau_{0,i} > t) &= \sum_{j=1}^m \mathbf{P}(\tau_{0,j} > t-u) p_{ij} (\lambda_i - Q_{ii}) e^{-(\lambda_i - Q_{ii})u} du \\ &+ \sum_{j=1}^m \int_0^t \mathbf{P}(\tau_{1,j} > t-u) \Lambda_{ij} e^{-(\lambda_i - Q_{ii})u} du \\ &+ e^{-(\lambda_i - Q_{ii})t}. \end{aligned} \quad (3.74)$$

Stosując transformaty i zapis macierzowy do (3.73) i (3.74), otrzymujemy:

$$l_n(s) = \sum_{k=0}^{b-n-1} A_k(s) l_{n+k-1}(s) + \tilde{d}_{b-n}(s), \quad 0 < n < b,$$

$$l_0(s) = Z(s)l_0(s) + E(s)l_1(s) + z(s).$$

Podstawienie

$$l_n(s) = u_{b-n}(s)$$

daje następnie:

$$\sum_{k=-1}^{n-1} A_{k+1}(s)u_{n-k}(s) - u_n(s) = \psi_n(s), \quad 0 < n < b, \quad (3.75)$$

$$u_b(s) = Z(s)u_b(s) + E(s)u_{b-1}(s) + z(s), \quad (3.76)$$

gdzie

$$\psi_n(s) = A_n(s)u_1(s) - \tilde{d}_n(s).$$

Zgodnie z Twierdzeniem 1.1 każde rozwiązanie układu (3.75) jest postaci:

$$u_n(s) = R_n(s)c(s) + \sum_{k=1}^n R_{n-k}(s)\psi_k(s). \quad (3.77)$$

Podstawiając $n = 1$ do (3.77), obliczamy

$$c(s) = A_0(s)u_1(s)$$

oraz

$$u_n(s) = \sum_{k=0}^n R_{n-k}(s)A_k(s)u_1(s) - \sum_{k=1}^n R_{n-k}(s)\tilde{d}_k(s).$$

W końcu, przy pomocy (3.76) możemy obliczyć

$$u_1(s) = G_b^{-1}(s)h_b(s)$$

i to kończy dowód Twierdzenia 3.7. ■

Z praktycznego punktu widzenia wygodne jest, że w definicji $l_{n,i}(s)$ występuje ogon rozkładu. Dzięki temu możemy łatwo obliczać wartość oczekiwaną czasu do wypełnienia:

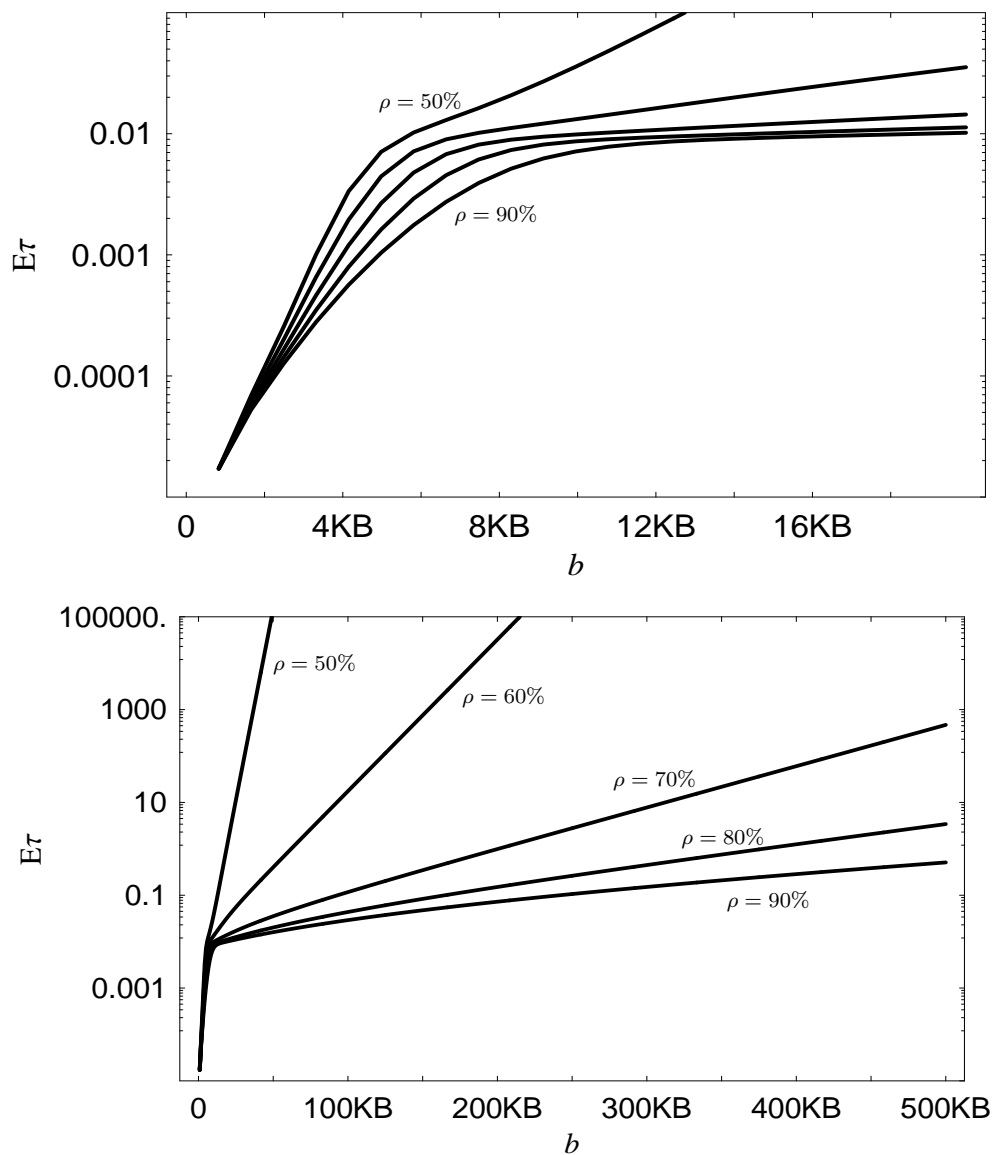
$$\mathbf{E}\tau_{n,i} = l_{n,i}(0). \quad (3.78)$$

W celu znalezienia transformat dystrybucyjny czy gęstości wystarczy proste przekształcenia transformaty ogona. Mianowicie, dla dystrybucyjny czasu do wypełnienia mamy

$$\int_0^\infty e^{-st} \mathbf{P}(\tau_{n,i} < t) dt = \frac{1}{s} - l_{n,i}(s), \quad (3.79)$$

zaś dla gęstości

$$\int_0^\infty e^{-st} \mathbf{P}'_t(\tau_{n,i} < t) dt = 1 - sl_{n,i}(s). \quad (3.80)$$



Rys. 3.26. Średni czas [s] do przepięnienia bufora w funkcji rozmiaru bufora dla różnych wartości obciążenia systemu (Przykład 3.12). $X(0) = 0$, rozkład $J(0)$ zgodnie z π

Fig. 3.26. Average time [s] to buffer overflow versus the buffer size for different values of the offered load (Example 3.12). $X(0) = 0$, $J(0)$ distributed according to π

Przykład 3.12. Zbadamy teraz czas do przepełnienia bufora dla parametryzacji MMPP z Przykładu 3.2. Obliczenia zostaną przeprowadzone dla różnych rozmiarów bufora oraz różnych wartości czasu transmisji d - w ten sposób otrzymamy też różne wartości obciążenia systemu $\rho = \lambda d$.

Rys. 3.26 przedstawia średni czas do przepełnienia bufora w funkcji b dla obciążeń od 50% do 90%. Widzimy, że kształt wykresu jest podobny dla wszystkich wartości ρ , tzn. dla $b = 4 - 8$ KB następuje zagięcie wykresu, a następnie funkcja przyjmuje kształt zbliżony do liniowego (przy logarytmicznym wyskalowaniu osi pionowej). Nachylenie wykresu zależy od ρ i dla dużych ρ jest małe. Dlatego czasy do przepełnienia bufora mogą być krótkie, nawet przy dużych rozmiarach buforów, o ile tylko obciążenie systemu jest duże.

Jak się okazuje, czas do przepełnienia bufora (a więc i częstość przepełnień) jest charakterystyką szczególnie wrażliwą na autokorelację strumienia pakietów. Na rys. 3.27 pokazane są czasy do przepełnienia bufora dla strumienia Poissona o tej samej intensywności co rozważany MMPP. Już na pierwszy rzut oka widać różnicę w wartościach z rys. 3.26 i 3.27.

Przykładowo, dla $\rho = 80\%$ i $b = 50$ pakietów średni czas do przepełnienia bufora dla strumienia Poissona wynosi

$$E\tau = 220210 \text{ s,}$$

a więc praktycznie przepełnień się nie obserwuje. Dla rozważanego MMPP czas ten wynosi odpowiednio

$$E\tau = 0.01739 \text{ s,}$$

czyli jest o 7 rzędów wielkości mniejszy i w tym wypadku przepełnienia będą już w dużym stopniu wpływały na wydajność mechanizmu kolejkowania.

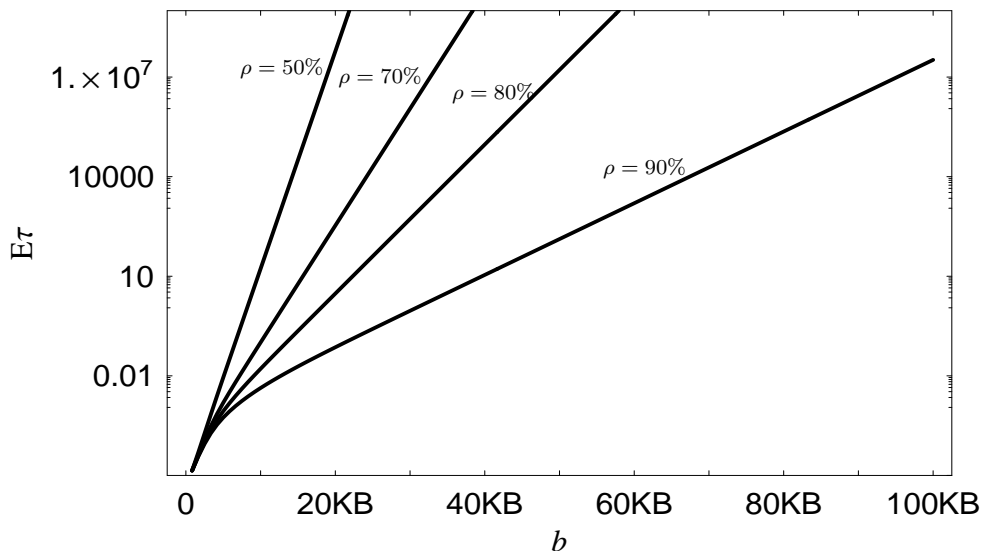
3.2.7. Złożoność obliczeniowa

Poświęćmy teraz trochę uwagi na zbadanie złożoności obliczeniowej rozwiązań przedstawionych w Twierdzeniach 3.1–3.7. Łatwo zauważyć, że złożoność jest podobna w każdym z tych twierdzeń, dlatego omówimy ją na przykładzie $\phi_n(s, l)$ z Twierdzenia 3.1.

Jeżeli chodzi o złożoność pamięciową, to do obliczenia jednej wartości $\phi_n(s, l)$ musimy zająć w pamięci miejsce na $3b$ macierzy o wymiarach $m \times m$, a mianowicie macierze $A_k(s)$, $R_k(s)$ i $B_k(s)$. Ponadto konieczne jest miejsce na b wektorów $g_k(s, l)$ rozmiaru m . A zatem złożoność pamięciowa jest rzędu $O(bm^2)$.

Jeżeli chodzi o złożoność czasową, to zakładając, że mnożenie i odwracanie macierzy kwadratowych jest rzędu $O(m^3)$, łatwo zobaczyć, że liczba operacji zmienoprzecinkowych rośnie jak $O(m^3b^2)$ i główny koszt jest związany z wyznaczeniem potencjału $R_k(s)$.

Wprawdzie zaawansowane metody, jak np. Coppersmitha i Winograda, pozwalają zredukować złożoność mnożenia macierzy nawet do $O(m^{2.376})$, ale realny zysk otrzymuje się w praktyce dopiero dla bardzo dużych macierzy.



Rys. 3.27. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla strumienia Poissona i różnych wartości obciążenia systemu (Przykład 3.12). $X(0) = 0$

Fig. 3.27. Average time to buffer overflow versus the buffer size for Poisson arrivals and different values of the offered load (Example 3.12). $X(0) = 0$

Otrzymana złożoność czasowa $O(m^3b^2)$ nie jest zła, jeżeli weźmiemy pod uwagę, że siłowe rozwiązanie układu (3.20) wymaga $O(m^3b^3)$ operacji zmiennoprzecinkowych.

Zwróćmy jeszcze uwagę na dwie ważne przy wykonywaniu obliczeń numerycznych sprawy. Po pierwsze, można zauważyć, że część wyznaczonych charakterystyk kolejkowania wraz ze wzrostem rozmiaru bufora ma postać asymptotycznie liniową (przy logarytmicznej skali osi pionowej). Do takich charakterystyk można zaliczyć np. prawdopodobieństwo przepełnienia bufora (rys. 3.5), współczynnik strat pakietów (rys. 3.12), czas do przepełnienia bufora (rys. 3.26). Dzięki tej własności, możemy łatwo otrzymywać przybliżone charakterystyki dla bardzo dużych buforów, wykorzystując wyniki dla małych lub umiarkowanych buforów, jednocześnie redukując nakład obliczeniowy.

Po drugie, obliczanie charakterystyk kolejkowania przy pomocy metody potencjału ma następującą zaletę. Dla dowolnej charakterystyki możemy wyznaczyć potencjał $R_k(s)$ do pewnego maksymalnego indeksu k_{max} , zapamiętać wyniki, a następnie wielokrotnie obliczać wartości interesującej nas charakterystyki dla różnych b , już bez dodatkowego nakładu obliczeniowego.

3.3. Nota bibliograficzna do Rozdziału 3

Przedstawione na początku rozdziału podstawowe informacje o MMPP można znaleźć w doskonałej „książce kucharskiej” (jak ją nazywają sami autorzy) poświęconej temu procesowi [91].

Ze względu za swoją względną prostotę, możliwość kształtowania funkcji autokorelacji oraz stosunkową łatwość analizy, proces MMPP stał się bardzo popularny w modelowaniu ruchu w sieciach pakietowych [5, 13, 108, 136, 156, 178, 183, 207, 206], szczególnie w sieciach ATM [18, 112, 116, 170, 169, 174, 188, 200, 205] oraz w modelowaniu ruchu multimedialnego [185, 105, 201].

Do popularności MMPP przyczyniła się też dostępność algorytmów umożliwiających dopasowanie parametrów MMPP do zarejestrowanych śladów ruchu (temat ten zostanie szerzej omówiony w Rozdziale 5).

Jeżeli chodzi o podstawowe charakterystyki kolejkowania ruchu MMPP, to uwaga badaczy skupiona była głównie na systemach z nieograniczonym buforem (np. [137, 138, 139, 134, 15, 153]).

Systemy ze skończonym buforem analizowane były albo jedynie w stanie ustalonym [30, 19], albo przy szczególnych założeniach dotyczących rozkładu czasu obsługi [124, 133, 17].

Oprócz wymienionych wyżej prac poświęconych metodom dokładnym rozwiązywania kolejek dla ruchu MMPP, rozwijane były też metody przybliżone, np. [37, 198].

Przedstawione w tym rozdziale wyniki pochodzą z następujących prac. Rozkład długości kolejki i opóźnienia (podrozdziały 3.2.1 i 3.2.3) zostały zbadane w [57]. Współczynnik strat pakietów oraz średnia liczba pakietów utraconych w przedziale $(0, t]$ (podrozdział 3.2.2) zostały przedstawione w [66]. Analiza okresu przepełnienia bufora oraz liczby kolejnych strat pakietów (podrozdziały 3.2.4 i 3.2.5) została przeprowadzona w [68]. Wreszcie rozkład czasu do przepełnienia bufora (podrozdział 3.2.6) został wyznaczony w [69].

Na zakończenie warto dodać, że prowadzone są też badania, w których charakterystyki kolejkowania oblicza się nie w postaci jawnych wzorów, lecz przy pomocy numerycznego rozwiązywania dużych układów równań liniowych. W szczególności z MMPP wiążą się metody rozwiązywania układów występujących przy badaniu tzw. łańcuchów Markowa typu M/G/1, gdyż struktura tych układów jest podobna do (3.20). Literatura poświęcona numerycznym rozwiązaniom łańcuchów typu M/G/1 jest obszerna, np. [152, 7, 6, 26, 27, 25, 24, 94, 95, 102, 101, 127, 126, 128, 146, 145, 147, 148, 171, 168, 176].

4. KOLEJKOWANIE STRUMIENI BMAP

W tym rozdziale przedstawione będą charakterystyki kolejkowania strumienia typu BMAP. Proces ten posiada najbogatszą strukturę wśród procesów markowskich wykorzystywanych do modelowania ruchu w sieciach pakietowych i, co za tym idzie, największe możliwości naśladowania zjawisk statystycznych występujących w ruchu. Pokazane zostaną najważniejsze charakterystyki kolejkowania BMAP dla stanu ustalonego i nieustalonego, w tym długość kolejki, opóźnienie, czas do przepełnienia bufora.

4.1. Definicja i własności BMAP

Największą wadą omówionego w poprzednim rozdziale MMPP jako modelu ruchu w sieciach pakietowych jest to, że proces ten nie pozwala uwzględnić różnych rozmiarów pakietów (w modelowaniu wykorzystuje się wtedy tylko średni rozmiar pakietu). Nie ma to znaczenia dla sieci ATM, ale może mieć znaczenie w modelowaniu ruchu IP, zwłaszcza wtedy, gdy wariancja rozkładu rozmiaru pakietu jest duża lub występuje korelacja pomiędzy rozmiarami pakietów a intensywnością strumienia pakietów (zobacz np. [182]). Wady tej pozbawiony jest proces BMAP (*ang.* batch Markovian arrival process¹).

Nieco upraszczając, można powiedzieć, że BMAP jest rozszerzoną wersją MMPP, w której zdarzenia mogą się pojawiać również w chwilach zmiany stanu łańcucha modulującego, a także mogą się pojawiać w grupach.

Podobnie jak w wypadku złożonego procesu Poissona grupowa struktura może zostać wykorzystana albo do modelowania rozkładu rozmiaru pakietu (wtedy grupa zdarzeń to jeden pakiet o rozmiarze równym rozmiarowi grupy), albo do modelowania spiętrzeń w ruchu sieciowym (wtedy każde zdarzenie z grupy to jeden pakiet i grupa zdarzeń odpowiada grupie pakietów). W pierwszym wypadku, w odróżnieniu od złożonego procesu Poissona, przy pomocy BMAP możemy modelować nie tylko

¹Tzn. markowski proces zdarzeń o grupowej strukturze. Dla wygody będziemy używać po prostu skrótu BMAP.

rozmiar pakietu, lecz również autokorelację czasów pomiędzy pakietami i korelację rozmiaru pakietu z bieżącą intensywnością strumienia pakietów. Jak pokazano w pracy [182], zjawisko to może istotnie wpływać na charakterystyki kolejowania ruchu. Podobnie jak MMPP, BMAP daje też możliwość dokładnego dopasowania kształtu rozkładu brzegowego czasu pomiędzy pakietami, a co za tym idzie – również jego wartości średniej, wariancji i wyższych momentów.

Z powyższych względów BMAP szczególnie dobrze nadaje się do modelowania zagregowanego ruchu IP, gdyż dzięki jego bogatej strukturze można modelować praktycznie wszystkie własności statystyczne ruchu pakietów,

Proces BMAP formalnie definiuje się jako dwuwymiarowy łańcuch Markowa $(N(t), J(t))$ w przestrzeni stanów $\{(i, j) : i \geq 0, 1 \leq j \leq m\}$, o macierzy intensywności postaci

$$R = \begin{bmatrix} D_0 & D_1 & D_2 & D_3 & \cdots \\ & D_0 & D_1 & D_2 & \cdots \\ & & D_0 & D_1 & \cdots \\ & & & \cdot & \cdots \end{bmatrix},$$

gdzie $D_k, k \geq 0$ są macierzami o wymiarach $m \times m$ takimi, że:

- $D_k, k \geq 1$ są nieujemne,
- D_0 ma nieujemne elementy poza przekątną i ujemne na przekątnej,
- $D = \sum_{k=0}^{\infty} D_k$ jest macierzą intensywności (zerowa suma elementów w każdym wierszu),
- $D \neq D_0$.

$N(t)$ oznacza całkowitą liczbę zdarzeń w przedziale $(0, t]$ i ma postać niemalejącej funkcji schodkowej, w której skoki mają miejsce w punktach pojawiania się grup zdarzeń, zaś $J(t)$ oznacza modulujący łańcuch Markowa z ciągłym czasem o przestrzeni stanów $\{1, \dots, m\}$ i macierzy intensywności D .

Z powyższej definicji wynika, że do kompletnej parametryzacji BMAP konieczne i wystarczające jest podanie ciągu D_0, D_1, D_2, \dots macierzy $m \times m$ spełniających powyższe warunki.

Wyrobienie sobie intuicyjnego obrazu struktury BMAP ułatwia tzw. konstruktywna definicja tego procesu. Wygląda ona następująco.

Założmy, że modulujący łańcuch Markowa jest początkowo w stanie i , czyli $J(0) = i$. Łańcuch ten przebywa następnie w tym stanie przez pewien czas o rozkładzie wykładniczym z parametrem μ_i . Na końcu tego okresu następuje przejście łańcucha modulującego do innego stanu, być może (choć niekoniecznie) z równoczesnym pojawieniem się grupy zdarzeń. W szczególności, z prawdopodobieństwem

$p_i(j, k)$, gdzie $1 \leq k \leq m$, $j \geq 0$, następuje przejście do stanu k łącznie z pojawieniem się grupy zdarzeń o rozmiarze j . Zakładamy, że prawdopodobieństwa $p_i(j, k)$ spełniają warunki:

- $p_i(0, i) = 0$,
- $\sum_{j=0}^{\infty} \sum_{k=1}^m p_i(j, k) = 1$, dla każdego $0 \leq i \leq m$.

Jak widać, powyższa, konstruktywna definicja BMAP oparta jest na innej parametryzacji niż macierze D_0, D_1, D_2, \dots , a mianowicie na prawdopodobieństwach $p_i(j, k)$ i intensywnościach μ_i . Pomiedzy tymi dwoma parametryzacjami zachodzą następujące relacje:

$$\mu_i = -(D_0)_{ii}, \quad 1 \leq i \leq m, \quad (4.1)$$

$$p_i(0, k) = \frac{1}{\mu_i} (D_0)_{ik}, \quad 1 \leq i, k \leq m, \quad k \neq i, \quad (4.2)$$

$$p_i(j, k) = \frac{1}{\mu_i} (D_j)_{ik}, \quad 1 \leq i, k \leq m, \quad j \geq 1. \quad (4.3)$$

Konstruktywna definicja BMAP umożliwia bardzo prostą symulację tego procesu. Przykładowy kod w języku C jest podany w podrozdziale 5.5.

Podstawowe charakterystyki BMAP w stanie ustalonym obliczamy następująco.

Średnia intensywność (wliczając rozmiary grup) wyraża się wzorem

$$\lambda = \pi \sum_{k=1}^{\infty} k D_k \mathbf{1},$$

gdzie π jest rozkładem stacjonarnym dla D , tzn. wektorem spełniającym

$$\pi D = (0, \dots, 0), \quad \pi \mathbf{1} = 1.$$

Intensywność strumienia grup zdarzeń oblicza się jako

$$\lambda_g = \pi(-D_0)\mathbf{1}.$$

Wariancja czasu pomiędzy wpływem kolejnych grup zdarzeń ma postać

$$Var = -\frac{2}{\lambda_g} \pi D_0^{-1} \mathbf{1} - \frac{1}{\lambda_g^2}.$$

Oznaczając przez T_i czas pomiędzy $(i-1)$ -szą i i -tą grupą zdarzeń, możemy obliczyć **autokowariancję** BMAP jako

$$\begin{aligned} Cov(k) &= \mathbf{E}[(T_i - \mathbf{E}T_i)(T_{i+k} - \mathbf{E}T_{i+k})] \\ &= p D_0^{-1} C (C^{k-1} - \mathbf{1} p) D_0^{-1} C \mathbf{1}, \end{aligned} \quad (4.4)$$

gdzie

$$C = -D_0^{-1}(D - D_0)$$

oraz p jest rozkładem stacjonarnym dla macierzy $C - I$, tzn.

$$p(C - I) = (0, \dots, 0), \quad p \mathbf{1} = 1.$$

Autokorelacja BMAP wynosi wtedy

$$Corr(k) = \frac{Cov(k)}{Var}.$$

Funkcja licząca jest określona dla BMAP jako

$$P_{i,j}(n, t) = \mathbf{P}(N(t) = n, J(t) = j | N(0) = 0, J(0) = i),$$

zaś **funkcja tworząca funkcji liczącej**

$$P^*(z, t) = \sum_{n=0}^{\infty} P(n, t) z^n$$

wyraża się wzorem

$$P^*(z, t) = e^{D(z)t}, \quad |z| \leq 1,$$

gdzie

$$D(z) = \sum_{k=0}^{\infty} z^k D_k.$$

W wypadku BMAP również będziemy posługiwać się **funkcjonałami całkowymi** funkcji liczącej:

$$a_{k,i,j}(s) = \int_0^{\infty} e^{-st} P_{i,j}(k, t) dF(t), \quad (4.5)$$

$$d_{k,i,j}(s) = \int_0^{\infty} e^{-st} P_{i,j}(k, t) (1 - F(t)) dt, \quad (4.6)$$

$$c_{k,i,j}(s, \sigma) = \int_0^{\infty} e^{-\sigma t} P_{i,j}(k, t) dt \int_0^{\infty} e^{-sx} d_x F(x+t), \quad (4.7)$$

zwykle w postaci macierzy $m \times m$:

$$A_k(s) = [a_{k,i,j}(s)]_{i,j}, \quad (4.8)$$

$$\bar{D}_k(s) = [d_{k,i,j}(s)]_{i,j}, \quad (4.9)$$

$$C_k(s, \sigma) = [c_{k,i,j}(s, \sigma)]_{i,j}. \quad (4.10)$$

Macierze te obliczamy przy pomocy metody uniformizacji przedstawionej w podrozdziale 5.2. Ponadto będziemy używać następujących macierzy $m \times m$:

$$Y_k(s) = \left[\frac{\mu_i p_i(k, j)}{s + \mu_i} \right]_{i, j},$$

$$\bar{A}_k(s) = \sum_{i=k}^{\infty} A_i(s),$$

$$B_k(s) = A_{k+1}(s) - \bar{A}_{k+1}(s)(\bar{A}_0(s))^{-1},$$

oraz wektora kolumnowego:

$$z(s) = ((s + \mu_1)^{-1}, \dots, (s + \mu_m)^{-1})^T.$$

Oczywiście, podstawową własnością BMAP wykorzystywaną do wyprowadzania wzorów dla charakterystyk kolejkowania jest jego **markowska struktura**, która oznacza, że znając wartość łańcucha modulującego w pewnej chwili t_0 , możemy określić zachowanie probabilistyczne tego procesu po chwili t_0 jedynie na podstawie $J(t_0)$ – bez znajomości historii procesu przed t_0 .

BMAP jest najbardziej ogólnym procesem punktowym o markowskiej strukturze. Stosując odpowiednie parametryzacje, można z BMAP otrzymać wiele innych prostszych procesów markowskich używanych w modelowaniu ruchu sieciowego. W szczególności:

- przyjmując $m = 1$ oraz $D_0 = -\lambda$, $D_1 = \lambda$, otrzymujemy proces Poissona;
- przyjmując $m = 1$, $D_0 = -\lambda$ oraz dla $j \geq 1$ $D_k = p_j \lambda$, otrzymujemy złożony proces Poissona;
- przyjmując $D_0 = T$ oraz $D_1 = -T\mathbf{1}\alpha$, otrzymujemy proces odnowy, w którym rozkład czasu pomiędzy zdarzeniami jest typu fazowego z parametrami (α, T) (zobacz [150, 113]). Procesy tego rodzaju mają bardzo duże znaczenie praktyczne, gdyż przy pomocy rozkładu fazowego można przybliżyć każdy inny rozkład z dowolnie dużą dokładnością;
- przyjmując $D_0 = T$ oraz $D_k = -p_k T\mathbf{1}\alpha$, $k \geq 1$, otrzymujemy proces odnowy o grupowej strukturze i fazowym rozkładzie czasu pomiędzy zdarzeniami (gdzie p_k oznacza prawdopodobieństwo grupy o rozmiarze k);
- przyjmując, że $D_k = \mathbf{0}$ dla $k \geq 2$, otrzymujemy proces MAP (*ang.* Markovian arrival process, zobacz też [140]). Różni się on od BMAP jedynie tym, że zdarzenia pojawiają się pojedynczo, a nie w grupach;

- przyjmując $D_0 = Q - \Lambda$ i $D_1 = \Lambda$, otrzymujemy proces MMPP z parametrami Q i Λ ;
- przyjmując $D_0 = Q - \Lambda$ i $D_k = p_k \Lambda$, $k \geq 1$, otrzymujemy proces o grupowej strukturze (p_k - prawdopodobieństwo grupy o rozmiarze k), w którym czasy pomiędzy grupami zdarzeń tworzą MMPP;
- przyjmując $D_j = p_j D_1$, $j \geq 1$, otrzymujemy proces o skorelowanych czasach pomiędzy zdarzeniami i grupowym wpływie zdarzeń, w którym rozmiary kolejnych grup są zmiennymi niezależnymi;
- przyjmując $D_0 = -\lambda I$ oraz $(D_k)_{ij} = \lambda P_{ij} q_i(k)$, gdzie $\{q_i(k), k \geq 1, 1 \leq i \leq m\}$ jest rodziną dyskretnych rozkładów prawdopodobieństwa, zaś $P = [P_{ij}]$ jest macierzą przejścia łańcucha Markowa z dyskretnym czasem, otrzymujemy, odwrotnie niż poprzednio, proces o nieskorelowanych czasach pomiędzy grupami zdarzeń i skorelowanych rozmiarach grup zdarzeń. Ściśle mówiąc, rozmiar kolejnej grupy wybierany jest zgodnie ze stanem łańcucha Markowa o macierzy przejścia P .

Do pokazania przykładowych wyników numerycznych będziemy w dalszej części używali dwóch parametryzacji BMAP - jednej prostej, sztucznie skonstruowanej w celu zademonstrowania pewnych cech BMAP, oraz drugiej, bardziej skomplikowanej, zbudowanej w oparciu o zarejestrowany ślad ruchu IP oraz estymację parametrów przy pomocy metody EM (przedstawionej w podrozdziale 5.3).

Przykład 4.1. Pierwsza parametryzacja BMAP wygląda następująco:

$$D_0 = \begin{bmatrix} -32 & 3 & 1 \\ 4 & -54 & 4 \\ 1 & 1 & -127 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 14 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 3 & 5 \end{bmatrix},$$

$$D_3 = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 32 & 1 \\ 4 & 1 & 2 \end{bmatrix}, \quad D_{10} = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 4 & 1 \\ 3 & 0 & 105 \end{bmatrix}.$$

A zatem mamy $m = 3$, grupy mogą przyjmować rozmiary 1, 3, 10. Macierz D ma postać

$$D = \begin{bmatrix} -15 & 9 & 6 \\ 7 & -15 & 8 \\ 10 & 5 & -15 \end{bmatrix},$$

zaś jej stacjonarny rozkład prawdopodobieństwa

$$\pi = (0.36133, 0.32227, 0.31641).$$

Alternatywna parametryzacja, oparta na definicji konstruktywnej, wygląda następująco:

$$\begin{aligned}
 \mu_1 &= 32, & \mu_2 &= 54, & \mu_3 &= 127, \\
 p_1(1, 1) &= 0.437500, & p_1(3, 1) &= 0.031250, & p_1(10, 1) &= 0.06250, \\
 p_1(0, 2) &= 0.093750, & p_1(1, 2) &= 0.031250, & p_1(3, 2) &= 0.062500, \\
 p_1(10, 2) &= 0.09375, & p_1(0, 3) &= 0.031250, & p_1(1, 3) &= 0.062500, \\
 p_1(3, 3) &= 0.031250, & p_1(10, 3) &= 0.06250, & p_2(0, 1) &= 0.074074, \\
 p_2(1, 1) &= 0.018519, & p_2(3, 1) &= 0.018519, & p_2(10, 1) &= 0.018519, \\
 p_2(1, 2) &= 0.055556, & p_2(3, 2) &= 0.592593, & p_2(10, 2) &= 0.074074, \\
 p_2(0, 3) &= 0.074074, & p_2(1, 3) &= 0.037037, & p_2(3, 3) &= 0.018519, \\
 p_2(10, 3) &= 0.018519, & p_3(0, 1) &= 0.007874, & p_3(1, 1) &= 0.015748, \\
 p_3(3, 1) &= 0.031496, & p_3(10, 1) &= 0.023622, & p_3(0, 2) &= 0.007874, \\
 p_3(1, 2) &= 0.023622, & p_3(3, 2) &= 0.007874, & p_3(1, 3) &= 0.039370, \\
 p_3(3, 3) &= 0.015748, & p_3(10, 3) &= 0.826772.
 \end{aligned}$$

Przedstawiony BMAP został tak zbudowany, by zademonstrować wpływ na charakterystyki kolejkowania zjawiska dodatniego skorelowania rozmiaru grupy z bieżącą intensywnością strumienia grup. Patrząc na wartości π , μ_i oraz $p_i(k, j)$, widzimy, że łańcuch modulujący przebywa mniej więcej tyle samo czasu w każdym ze stanów, jednak intensywność strumienia pojawiających się grup rośnie wraz ze wzrostem numeru stanu. Ponadto średnie rozmiary grup pojawiających się w stanach 1, 2, 3 wynoszą odpowiednio 3.1, 3.1, 8.8. A zatem w stanie 3 występuje największa intensywność pojawiania się grup oraz największy średni rozmiar grupy.

Podstawowe charakterystyki omawianego BMAP wyglądają następująco. Intensywność strumienia grup wynosi

$$\lambda_g = 64.492,$$

zaś całkowita średnia intensywność

$$\lambda = 441.44.$$

Daje to średni rozmiar grupy

$$\frac{\lambda}{\lambda_g} = 6.8449.$$

Wariancja czasu pomiędzy dwoma kolejnymi grupami wynosi

$$Var = 4.3293 \times 10^{-4}.$$

Przykład 4.2. Druga prezentowana tu parametryzacja BMAP odwzorowuje zapisany przebieg zagregowanego ruchu IP. Do dopasowania parametrów wykorzystanych zostało milion nagłówków pakietów zapisanych w ogólnodostępnym pliku FRG-1137458803-1.tsh (zobacz [93]), przedstawiającym ruch z 17 stycznia 2006 roku, zarejestrowany w punkcie agregacji

Front Range GigaPOP, monitorowanym w ramach projektu *Passive Measurement and Analysis Project*, [155].

Rozmiar grupy w BMAP został tu wykorzystany do modelowania rozmiaru pakietu IP. Ślady ruchu IP są często zdominowane przez kilka czy kilkanaście rozmiarów pakietów i tyle różnych rozmiarów grup w BMAP wystarcza zwykle do precyzyjnego modelowania. W wykorzystywanej tu próbce ruchu siedem najczęstszych rozmiarów pakietów to 40, 52, 552, 1300, 1420, 1488 i 1500 B. Wymienione rozmiary wyczerpują już 97% całego ruchu i do nich tylko ograniczymy model.

Parametry BMAP dopasowane do zapisanego śladu IP wyglądają następująco (zobacz [61]):

$$\begin{aligned}
 D_0 &= \begin{bmatrix} -90020.6 & 5300.2 & 11454.4 \\ 9132.5 & -126814.5 & 14807.8 \\ 2923.1 & 198.6 & -94942.6 \end{bmatrix}, \\
 D_{40} &= \begin{bmatrix} 2898.0 & 3415.6 & 1365.3 \\ 3649.6 & 1510.9 & 1044.6 \\ 1943.3 & 1954.9 & 7696.6 \end{bmatrix}, \\
 D_{52} &= \begin{bmatrix} 10980.1 & 8180.1 & 3284.0 \\ 5875.6 & 19512.8 & 19443.1 \\ 4064.1 & 9956.0 & 2744.3 \end{bmatrix}, \\
 D_{552} &= \begin{bmatrix} 1259.3 & 1143.0 & 960.4 \\ 1343.6 & 1541.2 & 3903.5 \\ 2470.5 & 1607.7 & 665.2 \end{bmatrix}, \\
 D_{1300} &= \begin{bmatrix} 115.3 & 174.5 & 188.1 \\ 333.8 & 28.0 & 405.9 \\ 337.3 & 124.4 & 552.4 \end{bmatrix}, \\
 D_{1420} &= \begin{bmatrix} 878.9 & 65.6 & 506.4 \\ 916.4 & 31.1 & 999.1 \\ 205.0 & 138.1 & 1603.7 \end{bmatrix}, \\
 D_{1488} &= \begin{bmatrix} 95.1 & 93.2 & 61.3 \\ 199.5 & 175.3 & 129.7 \\ 192.2 & 34.6 & 107.5 \end{bmatrix}, \\
 D_{1500} &= \begin{bmatrix} 13997.0 & 14089.9 & 9514.9 \\ 31145.3 & 9632.8 & 1052.4 \\ 17681.9 & 14814.8 & 22926.4 \end{bmatrix}.
 \end{aligned}$$

W tab. 4.1 wyszczególnione są najważniejsze parametry wykorzystanej próbki ruchu oraz dopasowanego BMAP.

Ponieważ w urządzeniach sieciowych pakiety są zwykle dzielone na segmenty o stałej wielkości, powyższą parametryzację można jeszcze uprościć tak, by rozmiar grupy był wyskalowany w segmentach, a nie w bajtach. Przyjmując typowy rozmiar segmentu równy 64 B [109], otrzymujemy wówczas następującą parametryzację BMAP:

	śląd IP	BMAP
średni czas pomiędzy pakietami [μs]	11.467	11.467
odchylenie std. czasu pom. pakietami [μs]	11.599	11.594
średni rozmiar pakietu [B]	869.18	869.38
λ_g - średnia intensywność strumienia pakietów [pakiety/s]	87205	87204
λ - średnia całkowita intensywność ruchu [MB/s]	72.286	72.301

Tab. 4.1. Podstawowe parametry zapisanego ruchu IP i dopasowanego BMAP (Przykład 4.2)

$$D'_1 = D_{40} + D_{52}, \quad D'_9 = D_{552}, \quad D'_{21} = D_{1300},$$

$$D'_{23} = D_{1420}, \quad D'_{24} = D_{1488} + D_{1500},$$

której będziemy używać w dalszych obliczeniach.

4.2. Charakterystyki kolejkowania BMAP

4.2.1. Długość kolejki

Wyznaczanie charakterystyk kolejkowania BMAP rozpoczniemy od wyprowadzenia wzoru na transformatę długości kolejki, w postaci wektora kolumnowego

$$\phi_n(s, l) = (\phi_{n,1}(s, l), \dots, \phi_{n,m}(s, l))^T,$$

gdzie

$$\phi_{n,i}(s, l) = \int_0^\infty e^{-st} \Phi_{n,i}(t, l) dt,$$

oraz

$$\Phi_{n,i}(t, l) = \mathbf{P}(X(t) = l | X(0) = n, J(0) = i).$$

Twierdzenie 4.1. *Transformata Laplace'a rozkładu długości kolejki w systemie BMAP/G/1/b wyraża się wzorem*

$$\begin{aligned} \phi_n(s, l) = & \sum_{k=0}^{b-n} R_{b-n-k}(s) g_k(s, l) \\ & + \left(R_{b-n+1}(s) A_0(s) + \sum_{k=0}^{b-n} R_{b-n-k}(s) B_k(s) \right) T_b^{-1}(s) m_b(s, l), \end{aligned} \quad (4.11)$$

gdzie $R_k(s)$ jest potencjałem (wzór (1.3)) dla ciągu $A_k(s)$ oraz

$$\begin{aligned}
g_k(s, l) &= \bar{A}_{k+1}(s)(\bar{A}_0(s))^{-1}r_b(s, l) - r_{b-k}(s, l), \\
r_n(s, l) &= \begin{cases} \mathbf{0} \cdot \mathbf{1}, & \text{dla } l < n, \\ \bar{D}_{l-n}(s) \cdot \mathbf{1}, & \text{dla } n \leq l < b, \\ \frac{1-f(s)}{s} \cdot \mathbf{1} - \sum_{k=0}^{b-n-1} \bar{D}_k(s) \cdot \mathbf{1}, & \text{dla } l = b, \end{cases} \\
m_b(s, l) &= \sum_{k=0}^b Y_{b-k}(s) \sum_{i=0}^k R_{k-i}(s) g_i(s, l) - \sum_{k=0}^b R_{b-k}(s) g_k(s, l) + \delta_{0l} z(s), \\
T_b(s) &= R_{b+1}(s) A_0(s) + \sum_{k=0}^b R_{b-k}(s) B_k(s) - \sum_{k=b+1}^{\infty} Y_k(s) \\
&\quad - \sum_{k=0}^b Y_{b-k}(s) \left[R_{k+1}(s) A_0(s) + \sum_{i=0}^k R_{k-i}(s) B_i(s) \right]. \tag{4.12}
\end{aligned}$$

Dowód Twierdzenia 4.1. Stosując wzór na prawdopodobieństwo całkowite względem pierwszej chwili ukończenia obsługi, otrzymujemy dla $0 < n \leq b$ i $1 \leq i \leq m$ następujący układ równań:

$$\begin{aligned}
\Phi_{n,i}(t, l) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \Phi_{n+k-1,j}(t-u, l) P_{i,j}(k, u) dF(u) \\
&\quad + \sum_{j=1}^m \sum_{k=b-n}^{\infty} \int_0^t \Phi_{b-1,j}(t-u, l) P_{i,j}(k, u) dF(u) \\
&\quad + \rho_{n,i}(t, l), \tag{4.13}
\end{aligned}$$

gdzie

$$\rho_{n,i}(t, l) = (1 - F(t)) \cdot \begin{cases} 0, & \text{dla } l < n, \\ \sum_{j=1}^m P_{i,j}(l-n, t), & \text{dla } n \leq l < b, \\ \sum_{j=1}^m \sum_{k=b-n}^{\infty} P_{i,j}(k, t), & \text{dla } l = b. \end{cases}$$

Z kolei dla $n = 0$ i $1 \leq i \leq m$ mamy:

$$\begin{aligned}
\Phi_{0,i}(t, l) &= \sum_{j=1}^m \sum_{k=0}^b \int_0^t \Phi_{k,j}(t-u, l) p_i(k, j) \mu_i e^{-\mu_i u} du \\
&\quad + \sum_{j=1}^m \sum_{k=b+1}^{\infty} \int_0^t \Phi_{b,j}(t-u, l) p_i(k, j) \mu_i e^{-\mu_i u} du \\
&\quad + \delta_{0l} e^{-\mu_i t}. \tag{4.14}
\end{aligned}$$

Stosując transformatę Laplace'a do obu stron (4.13) i (4.14), otrzymamy odpowiednio

$$\begin{aligned} \phi_{n,i}(s, l) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} a_{k,i,j}(s) \phi_{n+k-1,j}(s, l) \\ &+ \sum_{j=1}^m \sum_{k=b-n}^{\infty} a_{k,i,j}(s) \phi_{b-1,j}(s, l) + \int_0^{\infty} e^{-st} \rho_{n,i}(t, l) dt, \end{aligned} \quad (4.15)$$

i

$$\begin{aligned} \phi_{0,i}(s, l) &= \sum_{j=1}^m \sum_{k=0}^b p_i(k, j) \phi_{k,j}(s, l) \frac{\mu_i}{s + \mu_i} \\ &+ \sum_{j=1}^m \sum_{k=b+1}^{\infty} p_i(k, j) \phi_{b,j}(s, l) \frac{\mu_i}{s + \mu_i} + \delta_{0l} \frac{1}{s + \mu_i}. \end{aligned} \quad (4.16)$$

W zapisie macierzowym równania (4.15) i (4.16) mają postać

$$\phi_n(s, l) = \sum_{k=0}^{b-n-1} A_k(s) \phi_{n+k-1}(s, l) + \sum_{k=b-n}^{\infty} A_k(s) \phi_{b-1}(s, l) + r_n(s, l), \quad 0 < n \leq b, \quad (4.17)$$

$$\phi_0(s, l) = \sum_{k=0}^b Y_k(s) \phi_k(s, l) + \sum_{k=b+1}^{\infty} Y_k(s) \phi_b(s, l) + \delta_{0l} z(s). \quad (4.18)$$

Po zastosowaniu zamiany zmiennych

$$\varphi_n(s, l) = \phi_{b-n}(s, l)$$

układ (4.17), (4.18) przyjmuje postać

$$\sum_{k=-1}^n A_{k+1}(s) \varphi_{n-k}(s, l) - \varphi_n(s, l) = \psi_n(s, l), \quad 0 \leq n < b, \quad (4.19)$$

$$\varphi_b(s, l) = \sum_{k=0}^b Y_{b-k}(s) \varphi_k(s, l) + \sum_{k=b+1}^{\infty} Y_k(s) \varphi_0(s, l) + \delta_{0l} z(s), \quad (4.20)$$

gdzie

$$\psi_n(s, l) = A_{n+1}(s) \varphi_0(s, l) - \sum_{k=n+1}^{\infty} A_k(s) \varphi_1(s, l) - r_{b-n}(s, l).$$

Zgodnie z Twierdzeniem 1.3, każde rozwiązanie układu (4.19) jest postaci

$$\varphi_n(s, l) = R_{n+1}(s)c(s, l) + \sum_{k=0}^n R_{n-k}(s)\psi_k(s, l), \quad n \geq 0. \quad (4.21)$$

Wstawiając $n = 0$ do (4.21), mamy

$$c(s, l) = A_0(s)\varphi_0(s, l), \quad (4.22)$$

więc pozostało jedynie znaleźć funkcje $\varphi_0(s, l)$ i $\varphi_1(s, l)$, konieczne do obliczenia $\psi_k(s, l)$.

Podstawiając $n = 0$ do (4.19), otrzymujemy

$$\varphi_1(s, l) = (\overline{A_0}(s))^{-1}(\varphi_0(s, l) - r_b(s, l)), \quad (4.23)$$

co sprowadza problem do znalezienia $\varphi_0(s, l)$. Posługując się warunkiem (4.20), łatwo otrzymujemy

$$\varphi_0(s, l) = T_b^{-1}(s)m_b(s, l),$$

co kończy dowód Twierdzenia 4.1. ■

Twierdzenie 4.1 stosujemy w praktyce do znajdowania rozkładu długości kolejki zarówno w stanie ustalonym, tzn. obliczając tylko

$$\Phi(l) = \lim_{s \rightarrow 0^+} s\phi_{n,i}(s, l)$$

dla dowolnego n oraz i , jak i w stanie nieustalonym, wykorzystując algorytm odwracania jednowymiarowej transformaty Laplace'a (podrozdział 5.1.1).

Przykład 4.3. Wyznamy teraz rozkład długości kolejki dla parametryzacji BMAP z Przykładu 4.1, $b = 100$, oraz stałego czasu transmisji $d = 1$ ms. Mamy wtedy do czynienia z małym obciążeniem systemu, $\rho = 44\%$.

Średnia długość kolejki w stanie ustalonym wynosi

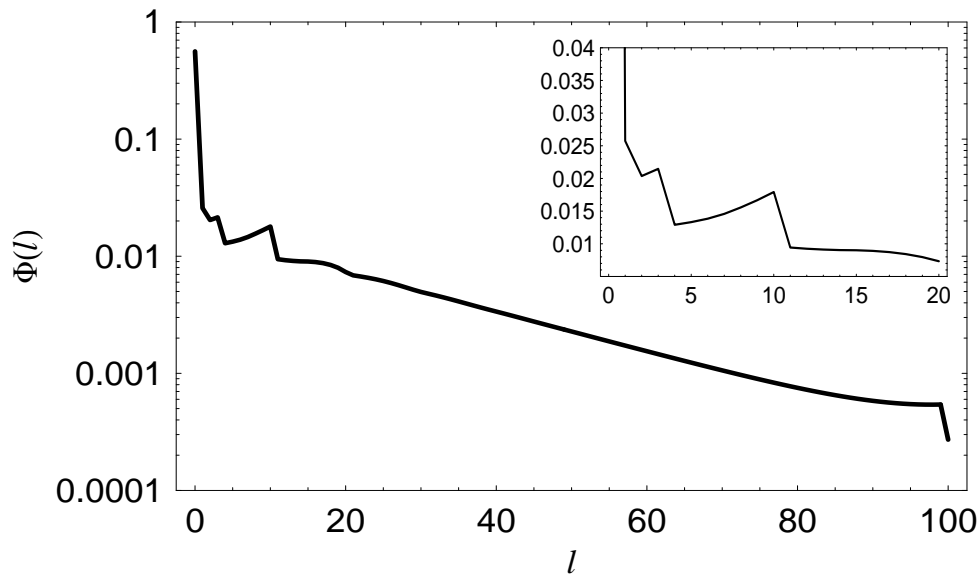
$$\mathbf{E}\Phi = 9.9415,$$

odchylenie standardowe długości kolejki

$$\sqrt{\text{Var}(\Phi)} = 18.108,$$

prawdopodobieństwo, że bufor jest przepełniony

$$\Phi(b) = 2.7147 \times 10^{-4},$$



Rys. 4.1. Rozkład długości kolejki w stanie ustalonym – Przykład 4.3

Fig. 4.1. Steady-state distribution of the queue length in Example 4.3

zaś prawdopodobieństwo, że bufor jest pusty

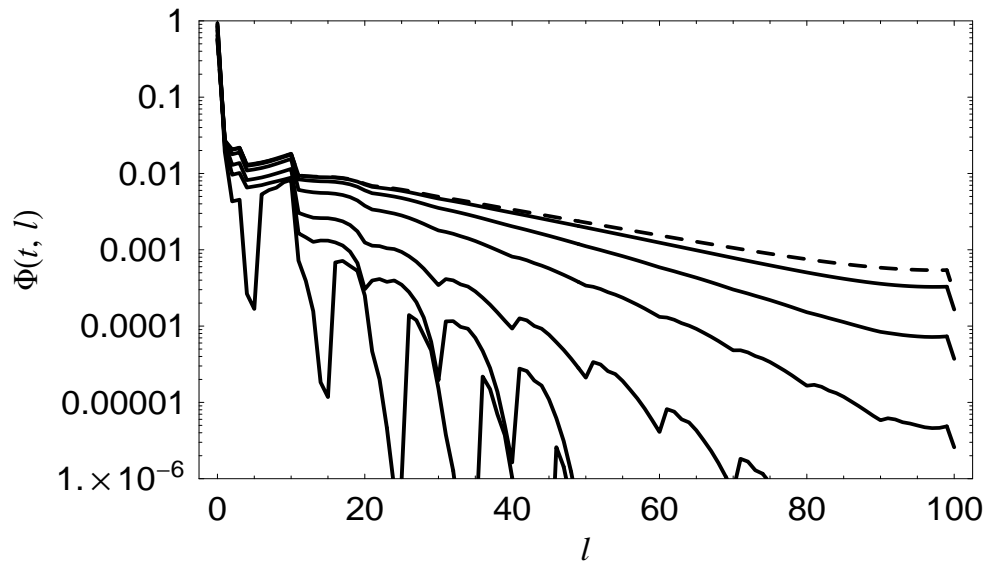
$$\Phi(0) = 0.56124.$$

Na rys. 4.1 przedstawiono rozkład długości kolejki w stanie ustalonym. Jak widzimy, dla małych l funkcja prawdopodobieństwa przyjmuje nieregularny kształt, w którym występują ostre piki. Jest to typowy efekt dla kolejek z procesami BMAP, związany z możliwymi rozmiarami pojawiających się grup (pakietów). W omawianym przykładzie te rozmiary to 1, 3, 10 i dlatego w tych miejscach na wykresie obserwujemy wysokie prawdopodobieństwa. Niekiedy możemy też zaobserwować piki odpowiadające sumom dwóch lub więcej rozmiarów grup, jednak w stanie ustalonym zanikają one szybko wraz ze wzrostem l .

Piki takie są też szczególnie wyraźne w stanie nieustalonym dla małych t . Widać to na rys. 4.2. Krótco po rozpoczęciu obserwacji jedynie niektóre długości kolejek związane z sumami rozmiarów grup mają wysokie prawdopodobieństwa. Z czasem wykres coraz bardziej się wygładza.

Z kolei na rys. 4.3 pokazano prawdopodobieństwo przepełnienia bufora w funkcji czasu dla różnych początkowych długości kolejki. Jak widać, wykres może nie być monotoniczny (np. przy $X(0) = 0.5b$). Stan ustalony jest osiągany po około 0.5 s.

Używana w tym przykładzie parametryzacja BMAP została dobrana tak, by nadać modelowi ruchu specjalną strukturę polegającą na dodatnim skorelowaniu bieżącej intensywności z rozmiarem grupy (zobacz Przykład 4.1). Zobaczymy teraz, jak struktura ta wpływa na



Rys. 4.2. Rozkład długości kolejki w różnych chwilach czasu w Przykładzie 4.3. Kolejno od dołu: $t = 5$ ms, $t = 10$ ms, $t = 20$ ms, $t = 50$ ms, $t = 100$ ms, $t = 200$ ms. Przerywaną linią zaznaczono stan ustalony, $t = \infty$. Początkowe parametry systemu: $X(0) = 0$, rozkład $J(0)$ wg π

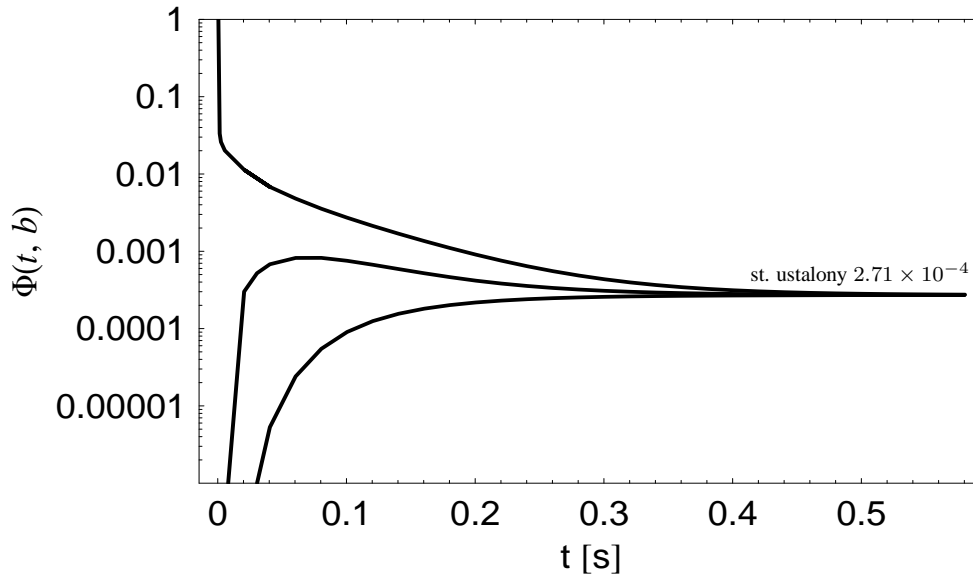
Fig. 4.2. Queue size distribution at different times, namely at $t = 5$ ms, $t = 10$ ms, $t = 20$ ms, $t = 50$ ms, $t = 100$ ms and $t = 200$ ms, counting from the bottom (Example 4.3). The dashed line represents the steady state. Initial settings: $X(0) = 0$, $J(0)$ distributed according to π

prawdopodobieństwo przepełnienia bufora. Tym razem dla porównania użyjemy nie prostego, ale złożonego strumienia Poissona (M^X) o takiej samej jak w BMAP średniej intensywności. Przyjmujemy, że rozmiar grupy w złożonym strumieniu Poissona jest stały i równy 10, czyli jest nawet większy niż średni rozmiar grupy w BMAP (wynoszący 6.8).

Wyniki są pokazane na rys. 4.4 (stan nieustalony) i 4.5 (stan ustalony) dla różnych rozmiarów bufora. Różnica, na niekorzyść BMAP, jest w każdym wypadku bardzo duża. W stanie ustalonym dla $b = 100$ sięga 4 rzędów wielkości (2.71×10^{-4} versus 2.32×10^{-8}).

Przykład 4.4. Zbadamy teraz długości kolejki dla ruchu z Przykładu 4.2 dla dwóch rozmiarów bufora, tzn. 100 KB oraz 200 KB. Zakładamy stały czas obsługi i obciążenie systemu wynoszące $\rho = 93.23\%$.

Rozkład długości kolejki w stanie ustalonym dla bufora 100 KB jest przedstawiony na rys. 4.6, zaś dla bufora 200 KB na rys. 4.7. Z wykresów tych możemy wyciągnąć kilka wniosków. Po pierwsze, w obu wypadkach widzimy charakterystyczną dla BMAP nieregularność wykresu dla małych l , szczególnie widoczny jest duży pik w okolicy 1.5 KB związany z dużym prawdopodobieństwem pakietu o rozmiarze 1500 B. Po drugie, kształt rozkładów



Rys. 4.3. Prawdopodobieństwo przepełnienia bufora w funkcji czasu w Przykładzie 4.3 dla różnych początkowych długości kolejki. Kolejno od dołu: $X(0) = 0$, $X(0) = 0.5b$, $X(0) = b$. Rozkład $J(0)$ wg π

Fig. 4.3. Buffer overflow probability as a function of time for different initial queue sizes, namely for $X(0) = 0$, $X(0) = 0.5b$ and $X(0) = b$, counting from the bottom. Example 4.3, $J(0)$ distributed according to π

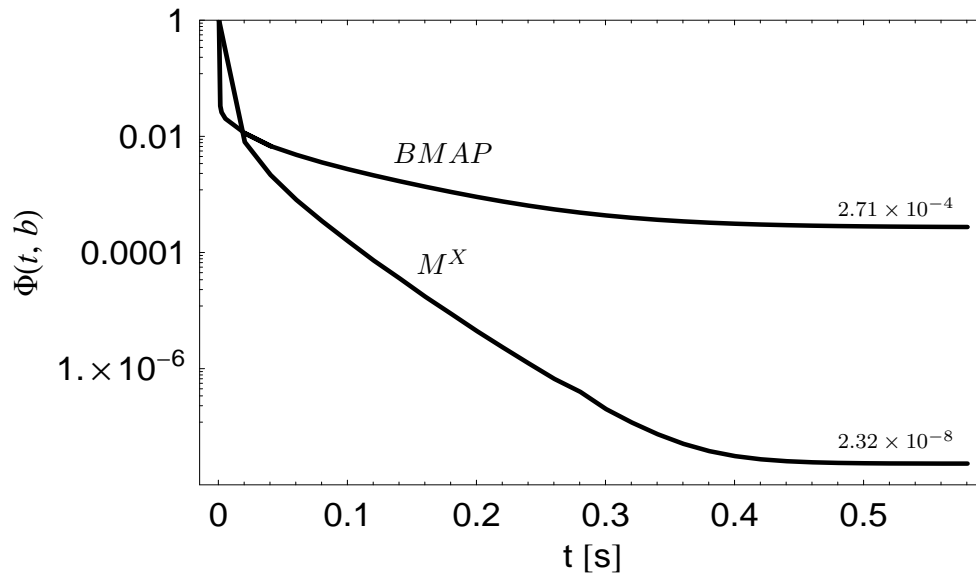
jest dla buforów 100 KB i 200 KB bardzo podobny, dla małych l prawie identyczny. Po trzeciej, pomijając okolice barier 0 i b , wykresy są liniowe (w skali logarytmicznej). Ta własność może zostać wykorzystana do szybkiego wyznaczenia rozkładu dla bardzo dużych buforów. Wystarczy wykonać obliczenia dla wartości l wokół barier 0 i b , zaś środkową część wykresu aproksymować funkcją liniową. Procedura taka daje bardzo dokładne rezultaty, ponieważ, jak widać na rys. 4.6 i 4.7, środkowy zakres jest niemal idealnie liniowy.

Na rys. 4.8 przedstawione jest prawdopodobieństwo przepełnienia bufora w stanie ustalonym w zależności od rozmiaru bufora i obciążenia systemu. Jak się można było spodziewać, prawdopodobieństwo to rośnie ze wzrostem obciążenia i maleje ze wzrostem rozmiaru bufora. Co istotne, zależność od rozmiaru bufora również może być z powodzeniem aproksymowana funkcją liniową.

4.2.2. Opóźnienie

Wprowadzimy teraz wzór na dwuwymiarową transformatę opóźnienia, tzn.

$$w_{n,i}(s, \sigma) = \int_0^{\infty} e^{-\sigma t} dt \int_0^{\infty} e^{-sx} \tilde{w}_{n,i}(x, t) dx,$$



Rys. 4.4. Prawdopodobieństwo przepełnienia bufora w funkcji czasu dla skorelowanego (BMAP) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.3). W obu wypadkach $X(0) = b$, rozkład $J(0)$ w BMAP wg π

Fig. 4.4. Buffer overflow probability as a function of time for correlated (BMAP) and uncorrelated (M^X) traffic (Example 4.3). In each case was $X(0) = b$, $J(0)$ in the BMAP case was distributed according to π

gdzie

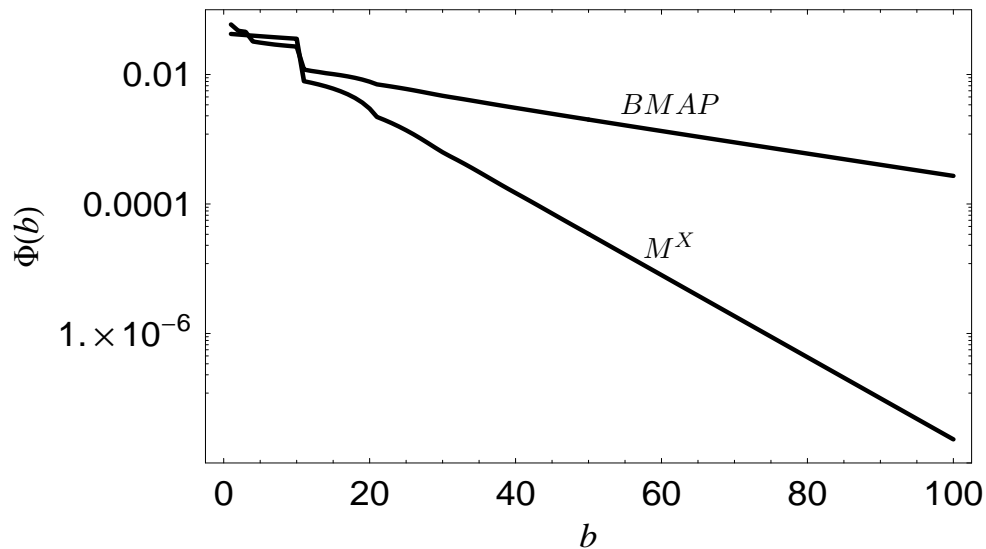
$$\tilde{w}_{n,i}(x, t) = \mathbf{P}(V(t) > x | X(0) = n, J(0) = i). \quad (4.24)$$

Jak poprzednio, będziemy zapisywać ją w postaci wektora kolumnowego:

$$w_n(s, \sigma) = (w_{n,1}(s, \sigma), \dots, w_{n,m}(s, \sigma))^T.$$

Twierdzenie 4.2. Transformata Laplace'a rozkładu opóźnienia w systemie BMAP/G/1/b wyraża się wzorem:

$$w_n(s, \sigma) = \sum_{k=0}^{b-n} R_{b-n-k}(\sigma) h_k(s, \sigma) + \left(\sum_{k=-1}^{b-n} R_{b-n-k}(\sigma) B_k(\sigma) + R_{b-n+1}(\sigma) \right) T_b^{-1}(\sigma) u_b(s, \sigma), \quad (4.25)$$



Rys. 4.5. Prawdopodobieństwo przepełnienia bufora w stanie ustalonym w funkcji rozmiaru bufora dla skorelowanego (BMAP) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.3)

Fig. 4.5. Steady-state overflow probability as a function of the buffer size for correlated (BMAP) and uncorrelated (M^X) traffic (Example 4.3)

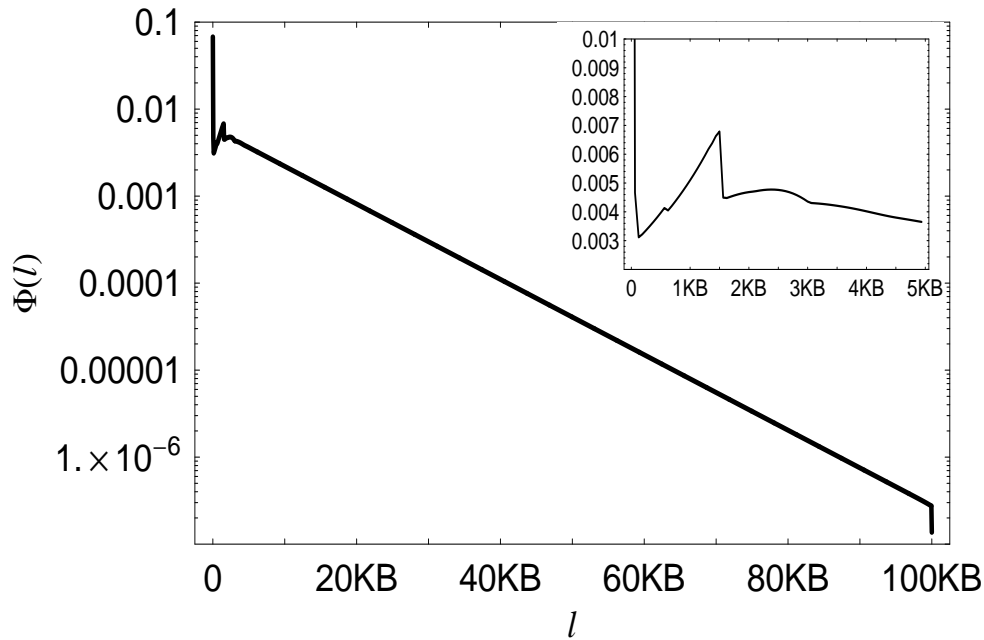
gdzie $R_k(s)$ jest potencjałem dla ciągu $A_k(s)$, funkcje $u_b(s, \sigma)$, $h_n(s, \sigma)$ mają następujące postaci

$$u_b(s, \sigma) = \sum_{k=0}^b Y_{b-k}(\sigma) \sum_{i=0}^k R_{k-i}(\sigma) h_i(s, \sigma) - \sum_{k=0}^b R_{b-k}(\sigma) h_k(s, \sigma),$$

$$h_n(s, \sigma) = \bar{A}_{n+1}(\sigma) (\bar{A}_0(\sigma))^{-1} q_b(s, \sigma) - q_{b-n}(s, \sigma),$$

$$q_n(s, \sigma) = \frac{1}{s} \sum_{k=0}^{b-n-1} [\bar{D}_k(\sigma) - f^{n+k-1}(s) C_k(s, \sigma)] \cdot \mathbf{1},$$

oraz $T_b(s)$ jest określone przez (4.12).

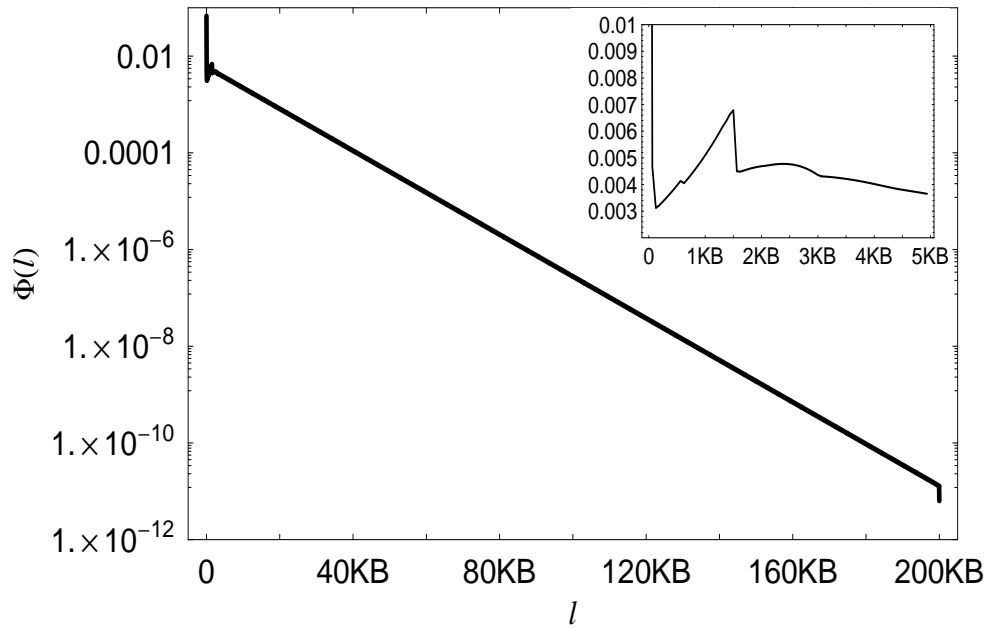


Rys. 4.6. Rozkład długości kolejki w stanie ustalonym w Przykładzie 4.4 dla bufora 100 KB.
W rogu – zbliżenie zakresu 0-5 KB

Fig. 4.6. Steady-state queue size distribution in Example 4.4 for the buffer size of 100 KB.
In the corner – a close-up of the range 0-5 KB

Dowód Twierdzenia 4.2. Wykorzystując wzór na prawdopodobieństwo całkowite ze względu na pierwszą chwilę ukończenia obsługi, otrzymujemy dla $0 < n \leq b$ i $1 \leq i \leq m$, że

$$\begin{aligned}
 \tilde{w}_{n,i}(x,t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \tilde{w}_{n+k-1,j}(x,t-u) P_{i,j}(k,u) dF(u) \\
 &+ \sum_{j=1}^m \sum_{k=b-n}^{\infty} \int_0^t \tilde{w}_{b-1,j}(x,t-u) P_{i,j}(k,u) dF(u) \\
 &+ \sum_{j=1}^m \sum_{k=0}^{b-n-1} P_{i,j}(k,t) \int_t^{\infty} (1 - F^{(n+k-1)*}(x-u+t)) dF(u).
 \end{aligned} \tag{4.26}$$



Rys. 4.7. Rozkład długości kolejki w stanie ustalonym w Przykładzie 4.4 dla bufora 200 KB.
W rogu – zbliżenie zakresu 0-5 KB

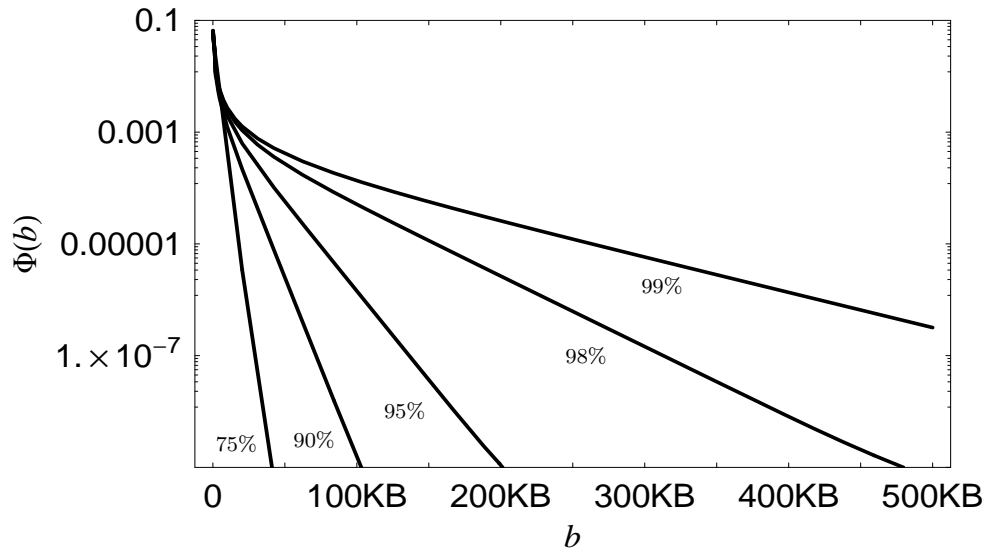
Fig. 4.7. Steady-state queue size distribution in Example 4.4 for the buffer size of 200 KB.
In the corner – a close-up of the range 0-5 KB

Dla $n = 0$ i $1 \leq i \leq m$ mamy

$$\begin{aligned} \tilde{w}_{0,i}(x,t) &= \sum_{j=1}^m \sum_{k=0}^b \int_0^t \tilde{w}_{k,j}(x,t-u) p_i(k,j) \mu_i e^{-\mu_i u} du \\ &+ \sum_{j=1}^m \sum_{k=b+1}^{\infty} \int_0^t \tilde{w}_{b,j}(x,t-u) p_i(k,j) \mu_i e^{-\mu_i u} du. \end{aligned}$$

Po zastosowaniu transformat i zapisu macierzowego dostajemy

$$\begin{aligned} w_n(s,\sigma) &= \sum_{k=0}^{b-n-1} A_k(\sigma) w_{n+k-1}(s,\sigma) \\ &+ \sum_{k=b-n}^{\infty} A_k(\sigma) w_{b-1}(s,\sigma) + q_n(s,\sigma), \quad 0 < n \leq b, \quad (4.27) \end{aligned}$$



Rys. 4.8. Prawdopodobieństwo przepełnienia bufora w stanie ustalonym w funkcji rozmiaru bufora dla obciążenia ρ od 75% do 99% (Przykład 4.4)

Fig. 4.8. Steady-state buffer overflow probability versus the buffer size for the offered load from 75% to 99% (Example 4.4)

oraz

$$w_0(s, \sigma) = \sum_{k=0}^b Y_k(\sigma) w_k(s, \sigma) + \sum_{k=b+1}^{\infty} Y_k(\sigma) w_b(s, \sigma). \quad (4.28)$$

Wykorzystując Twierdzenie 1.3, otrzymujemy dalej:

$$w_n(s, \sigma) = R_{b-n+1}(\sigma) c(\sigma) + \sum_{k=0}^{b-n} R_{b-n-k}(\sigma) \psi_k(s, \sigma), \quad (4.29)$$

$$\psi_n(s, \sigma) = A_{n+1}(\sigma) w_b(s, \sigma) - \sum_{k=n+1}^{\infty} A_k(\sigma) w_{b-1}(s, \sigma) - q_{b-n}(s, \sigma).$$

Pozostałe do obliczenia $c(\sigma)$, $w_{b-1}(s, \sigma)$ i $w_b(s, \sigma)$ wyznaczamy tak samo jak w dowodzie Twierdzenia 4.1. ■

Twierdzenie 4.2 wykorzystujemy w praktyce na trzy sposoby:

- średnią wartość opóźnienia w stanie ustalonym obliczamy jako

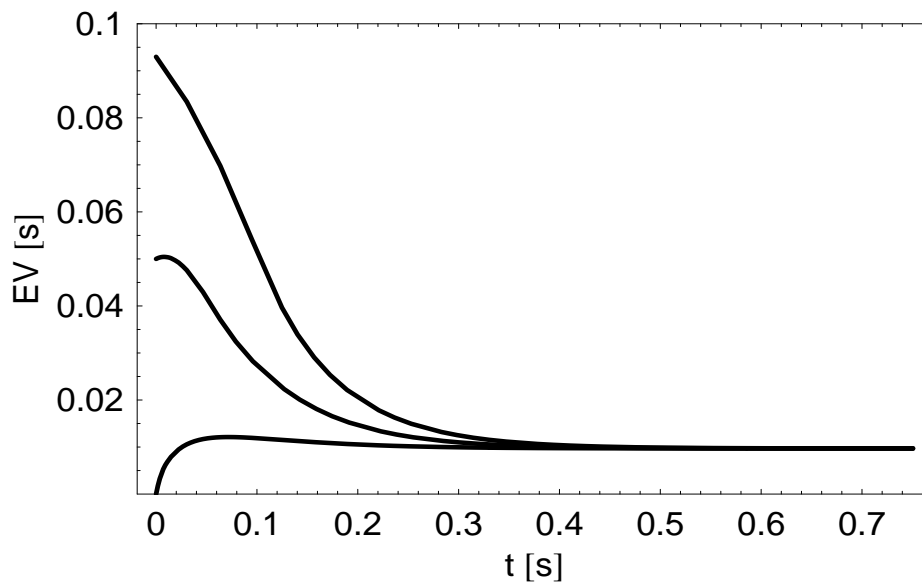
$$\lim_{\sigma \rightarrow 0^+} \sigma w_n(0, \sigma);$$

- średnią wartość opóźnienia w stanie nieustalonym obliczamy, odwracając $w_n(0, \sigma)$ przy pomocy algorytmu z podrozdziału 5.1.1;
- dokładną postać rozkładu opóźnienia w stanie nieustalonym obliczamy, odwracając $w_n(s, \sigma)$ przy pomocy algorytmu z podrozdziału 5.1.2.

Przykład 4.5. Średnie opóźnienie w stanie ustalonym dla parametryzacji BMAP z Przykładu 4.1, $b = 100$ oraz stałego czasu transmisji $d = 1$ ms ($\rho = 44\%$) wynosi

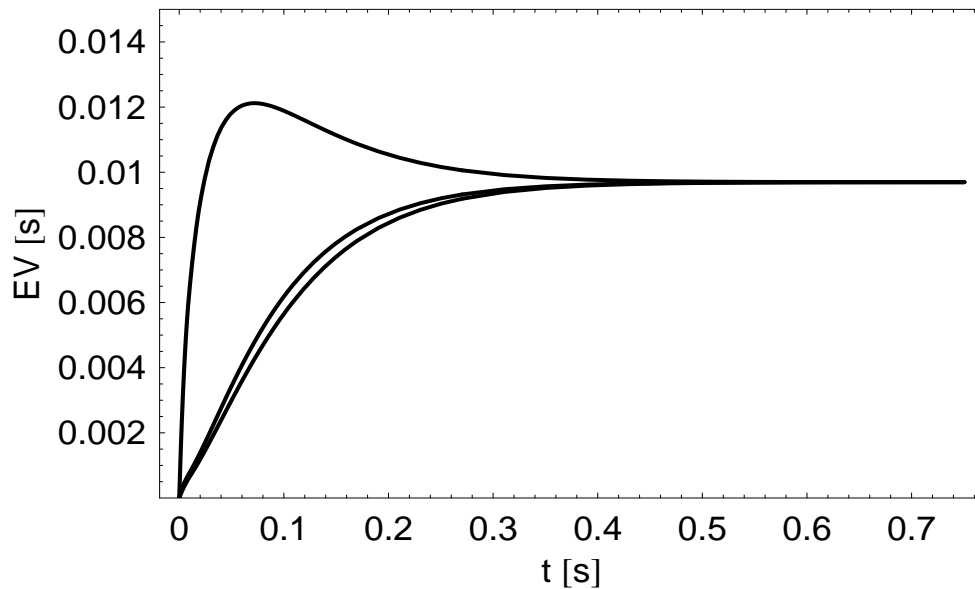
$$EV = 9.694 \text{ ms.}$$

Przebiegi średniego opóźnienia w funkcji czasu dla różnych początkowych długości kolejek oraz stanów łańcucha modulującego są przedstawione na rys. 4.9 i 4.10.



Rys. 4.9. Średnie opóźnienie w funkcji czasu w Przykładzie 4.5 dla różnych wartości $X(0)$ i $J(0) = 3$. Kolejno od dołu: $X(0) = 0$, $X(0) = 0.5b$, $X(0) = b$

Fig. 4.9. Average delay versus time in Example 4.5 for $J(0) = 3$ and different values of $X(0)$, namely for $X(0) = 0$, $X(0) = 0.5b$ and $X(0) = b$, counting from the bottom

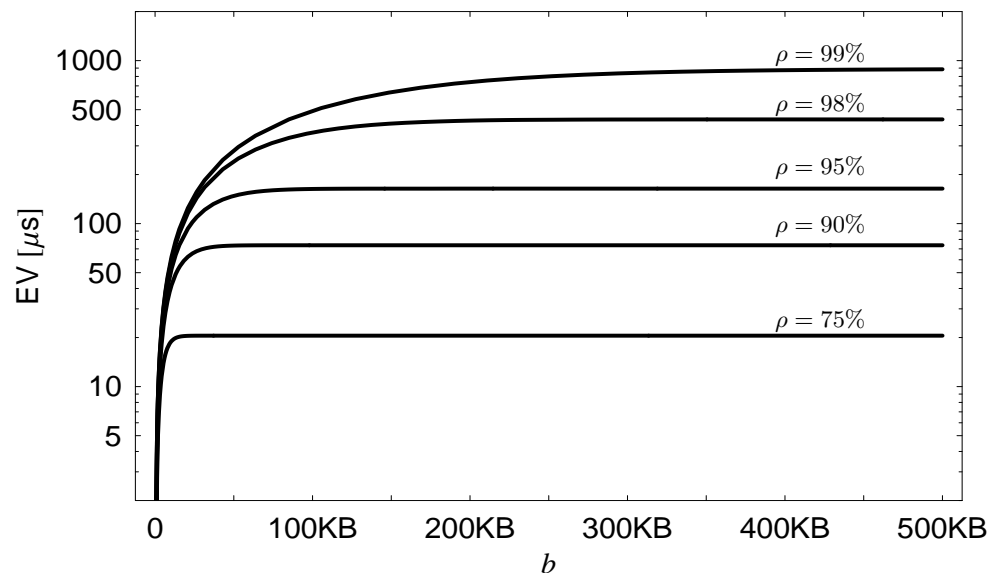


Rys. 4.10. Średnie opóźnienie w funkcji czasu w Przykładzie 4.5 dla różnych wartości $J(0)$. Kolejno od dołu: $J(0) = 1$, $J(0) = 2$, $J(0) = 3$. W każdym wypadku $X(0) = 0$

Fig. 4.10. Average delay versus time in Example 4.5 for different values of $J(0)$, namely for $J(0) = 1$, $J(0) = 2$ and $J(0) = 3$, counting from the bottom. In every case was $X(0) = 0$

Przykład 4.6. Dla parametryzacji BMAP z Przykładu 4.2 średnie opóźnienie w stanie ustalonym w zależności od rozmiaru bufora i obciążenia systemu jest pokazane na rys. 4.11. Jak widać, wraz ze wzrostem obciążenia opóźnienie rośnie i takiego efektu należało się spodziewać. Ciekawsza jest zależność opóźnienia od rozmiaru bufora. Ma ona charakterystyczny kształt - początkowo szybko narasta, następnie wykres zagina się („kolano”), by potem, od pewnego progowego rozmiaru bufora, przyjąć stałą wartość. Taki kształt funkcji opóźnienia ma swoje praktyczne konsekwencje. Poniżej progowego rozmiaru bufora (powiedzmy 200 KB dla $\rho = 98\%$) zmniejszanie rozmiaru bufora powoduje zmniejszanie opóźnienia. W zakresie powyżej wartości progowej zmiany rozmiaru bufora nie mają wpływu na opóźnienie.

Jak nietrudno odgadnąć, taka postać zależności opóźnienia od rozmiaru bufora wiąże się z zanikaniem wpływu skończonego rozmiaru bufora na pracę systemu. Dla dostatecznie dużych buforów przepełnienia następują na tyle rzadko, że system właściwie pracuje tak jak system z nieograniczonym rozmiarem bufora i stała wartość funkcji opóźnienia na wykresie jest właśnie równa opóźnieniu w systemie z nieograniczonym buforem.



Rys. 4.11. Średnie opóźnienie w stanie ustalonym w funkcji rozmiaru bufora dla różnych wartości obciążenia ρ (Przykład 4.6)

Fig. 4.11. Average steady-state delay versus the buffer size for different values of the offered load ρ (Example 4.6)

4.2.3. Czas do przepelnienia bufora

Wyprowadzimy teraz wzór na transformatę czasu do przepelnienia bufora, tzn.

$$l_{n,i}(s) = \int_0^{\infty} e^{-st} \mathbf{P}(\tau_{n,i} > t) dt,$$

$$l_n(s) = (l_{n,1}(s), l_{n,2}(s), \dots, l_{n,m}(s))^T,$$

gdzie

$$\tau_{n,i} = \inf\{t > 0 : X(t) = b | X(0) = n, J(0) = i\}.$$

Twierdzenie 4.3. Transformata Laplace'a rozkładu czasu do przepelnienia bufora w systemie BMAP/G/1/b ma postać

$$l_n(s) = \sum_{k=0}^{b-n} R_{b-n-k}(s) A_k(s) W_b^{-1}(s) v_b(s) - \sum_{k=1}^{b-n} R_{b-n-k}(s) \tilde{d}_k(s), \quad 0 \leq n < b, \quad (4.30)$$

gdzie

$$W_b(s) = \sum_{k=1}^b Y_{b-k}(s) \sum_{i=0}^k R_{k-i}(s) A_i(s) - \sum_{k=0}^b R_{b-k}(s) A_k(s), \quad (4.31)$$

$$v_b(s) = \sum_{k=1}^b Y_{b-k}(s) \sum_{i=1}^{k-1} R_{k-i}(s) \tilde{d}_i(s) - \sum_{k=1}^b R_{b-k}(s) \tilde{d}_k(s) - z(s), \quad (4.32)$$

$$\tilde{d}_k(s) = \sum_{i=0}^{k-1} \bar{D}_i(s) \cdot \mathbf{1},$$

oraz $R_k(s)$ jest potencjałem dla ciągu $A_k(s)$.

Dowód Twierdzenia 4.3. Wykorzystując wzór na prawdopodobieństwo całkowite, otrzymujemy dla $0 < n < b$, $1 \leq i \leq m$:

$$\begin{aligned} \mathbf{P}(\tau_{n,i} > t) &= \sum_{j=1}^m \sum_{k=0}^{b-n-1} \int_0^t \mathbf{P}(\tau_{n+k-1,j} > t-u) P_{i,j}(k,u) dF(u) \\ &\quad + (1-F(t)) \sum_{j=1}^m \sum_{k=0}^{b-n-1} P_{i,j}(k,t) \end{aligned} \quad (4.33)$$

oraz

$$\mathbf{P}(\tau_{0,i} > t) = \sum_{j=1}^m \sum_{k=0}^{b-1} \int_0^t \mathbf{P}(\tau_{k,j} > t-u) p_i(k,j) \mu_i e^{-\mu_i u} du + e^{-\mu_i t}. \quad (4.34)$$

Ze wzorów (4.33) i (4.34) – po zastosowaniu transformaty Laplace'a – obliczamy:

$$l_{n,i}(s) = \sum_{j=1}^m \sum_{k=0}^{b-n-1} a_{k,i,j}(s) l_{n+k-1,j}(s) + \tilde{d}_{b-n,i}(s), \quad 0 < n < b, 1 \leq i \leq m,$$

$$l_{0,i}(s) = \sum_{j=1}^m \sum_{k=0}^{b-1} p_i(k,j) l_{k,j}(s) \frac{\mu_i}{s + \mu_i} + \frac{1}{s + \mu_i}, \quad 1 \leq i \leq m,$$

gdzie

$$\tilde{d}_{n,i}(s) = \sum_{j=1}^m \sum_{k=0}^{n-1} \int_0^\infty e^{-st} P_{i,j}(k,t) (1-F(t)) dt.$$

Powyższe równania w zapisie macierzowym mają postać

$$l_n(s) = \sum_{k=0}^{b-n-1} A_k(s)l_{n+k-1}(s) + \tilde{d}_{b-n}(s), \quad 0 < n < b,$$

$$l_0(s) = \sum_{k=0}^{b-1} Y_k(s)l_k(s) + z(s).$$

Następnie, po podstawieniu

$$l_n(s) = u_{b-n}(s),$$

mamy:

$$\sum_{k=-1}^{n-1} A_{k+1}(s)u_{n-k}(s) - u_n(s) = \psi_n(s), \quad 0 < n < b, \quad (4.35)$$

$$u_b(s) = \sum_{k=1}^b Y_{b-k}(s)u_k(s) + z(s), \quad (4.36)$$

gdzie

$$\psi_n(s) = A_n(s)u_1(s) - \tilde{d}_n(s).$$

Na mocy Twierdzenia 1.1 otrzymujemy:

$$u_n(s) = R_n(s)c(s) + \sum_{k=1}^n R_{n-k}(s)\psi_k(s), \quad n \geq 1, \quad (4.37)$$

gdzie $c(s)$ jest wektorem niezależnym od n .

Podstawiając $n = 1$ do (4.37), dostajemy $c(s) = A_0(s)u_1(s)$ oraz

$$u_n(s) = \sum_{k=0}^n R_{n-k}(s)A_k(s)u_1(s) - \sum_{k=1}^n R_{n-k}(s)\tilde{d}_k(s).$$

Wreszcie przy pomocy (4.36) możemy łatwo wyprowadzić

$$u_1(s) = W_b^{-1}(s)v_b(s),$$

co kończy dowód Twierdzenia 4.3. ■

Wykorzystując Twierdzenie 4.3, możemy łatwo znajdować różne charakterystyki pochodne, takie jak wartość średnia czasu do przepełnienia, dystrybuanta czasu do

przepełnienia czy gęstość prawdopodobieństwa czasu do przepełnienia. Odpowiednie wzory wyglądają podobnie jak w wypadku MMPP (zobacz (3.78), (3.79), (3.80)).

Warto przy tej okazji zwrócić uwagę na jeszcze jedną charakterystykę, którą otrzymujemy natychmiast, jeżeli znamy rozkład τ_n . Mianowicie, zdarzenie $\{\tau_n > t\}$ zachodzi wtedy i tylko wtedy, gdy w przedziale czasu $(0, t]$ nie nastąpi przepełnienie bufora. A zatem funkcja

$$\chi(t) = \mathbf{P}(\tau_n > t) \quad (4.38)$$

wyraża prawdopodobieństwo, że w przedziale $(0, t]$ ani razu nie nastąpi przepełnienie. Wartości $\chi(t)$ otrzymamy łatwo – odwracając $l_n(s)$.

Przykład 4.7. Dla parametryzacji BMAP z Przykładu 4.1, $b = 100$, oraz stałego czasu transmisji $d = 1$ ms ($\rho = 44\%$), średni czas do przepełnienia bufora, przy założeniu że bufor jest początkowo pusty, wynosi

$$\mathbf{E}\tau = 8.9931 \text{ s.}$$

Na rys. 4.12 pokazane jest porównanie czasu do przepełnienia bufora dla BMAP oraz złożonego procesu Poissona o stałym rozmiarze grupy równym 10 i takiej samej jak w BMAP średniej intensywności. Przykładowo, dla bufora $b = 100$ średni czas do przepełnienia dla złożonego strumienia Poissona wynosi 37741 s, czyli o 3 rzędy wielkości więcej niż dla BMAP.

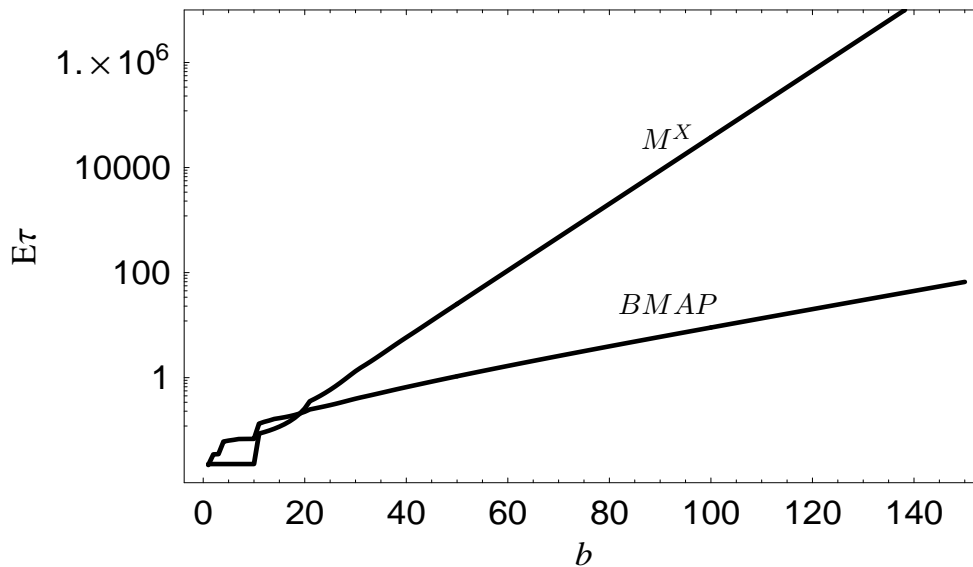
Przykład 4.8. Dla parametryzacji BMAP z Przykładu 4.2 zależność czasu do przepełnienia bufora od rozmiaru bufora i obciążenia systemu jest pokazana na rys. 4.13. Jak widać, jest to kolejny wypadek, w którym wykres ma postać liniową dla dużych b , co może zostać wykorzystane do szybkiego obliczania czasu do przepełnienia dla dużych buforów w oparciu o wyniki otrzymane dla małych buforów.

4.2.4. Złożoność obliczeniowa

Złożoność obliczeniowa przedstawionych tu rezultatów dla strumieni BMAP jest tego samego rzędu co dla MMPP i wyznaczamy ją w taki sam sposób jak w podrozdziale 3.2.7.

4.3. Nota bibliograficzna do Rozdziału 4

BMAP został zaproponowany przez M. Neutsa [151] i od jego nazwiska proces ten początkowo nazywano N -procesem. Oryginalna parametryzacja tego procesu jest skomplikowana, dlatego D. Lucantoni [137] zdefiniował na nowo N -proces, posługując się bardziej przejrzystym zapisem. Od tego też czasu używa się zwykle nazwy BMAP. Podstawowe własności BMAP można znaleźć np. w [43, 70, 64].



Rys. 4.12. Średni czas do przepięnienia bufora w funkcji rozmiaru bufora dla skorelowanego ($BMAP$) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.7). W obu wypadkach $X(0) = 0$, rozkład $J(0)$ dla BMAP wg π

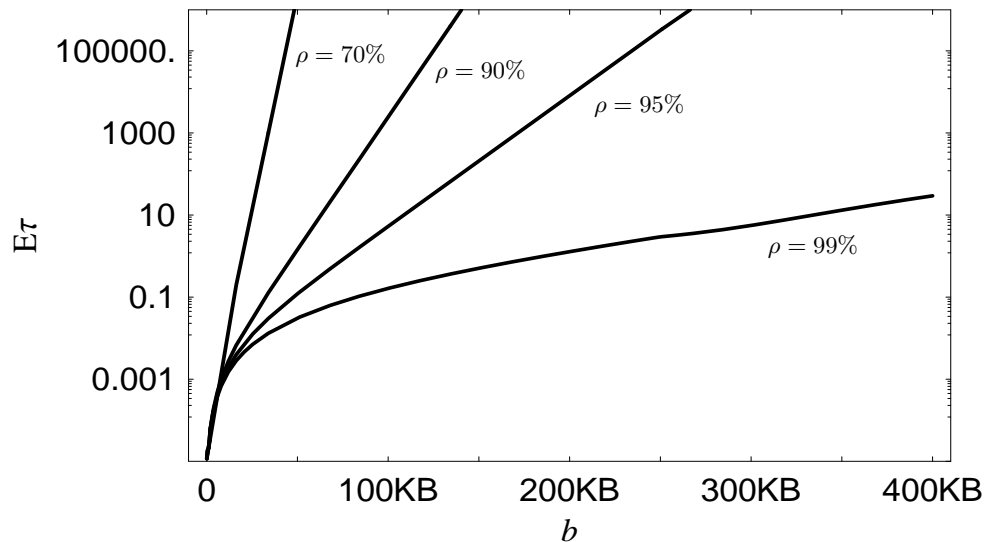
Fig. 4.12. Average time to buffer overflow versus the buffer size for correlated ($BMAP$) and uncorrelated (M^X) traffic (Example 4.7). In each case $X(0) = 0$, $J(0)$ in the BMAP case was distributed according to π

Do tej pory napisano wiele artykułów poświęconych modelowaniu ruchu w sieciach pakietowych przy pomocy BMAP lub wykorzystujących BMAP do szacowania ich parametrów wydajnościowych, np. [182, 121, 190, 173, 29, 208, 3, 120, 118, 119, 89, 125, 184]. W części badań wykorzystywane są nieco uproszczone wersje BMAP, takie jak MAP² (zobacz np. [154, 113, 114, 106, 165, 189, 12, 11, 195, 23, 180]) czy D-BMAP³ (zobacz np. [20, 28, 92, 96, 107, 197]).

Jeżeli chodzi o charakterystyki kolejkowania procesu BMAP, to dość dobrze zbadane zostały systemy z nieograniczonym buforem zarówno w stanie ustalonym [167, 137, 129], jak i nieustalonym [138, 139, 134]. Prace poświęcone systemom z ograniczonym buforem są rzadsze (np. [30, 100]). Część badań poświęcono różnego rodzaju specjalnym systemom kolejkowym, np. z grupową obsługą [38, 39], z wieloma trybami obsługi [84, 85], ze szczególnymi typami strumieni wejściowych, jak

²Tzn. BMAP, w którym $D_k = 0$ dla $k \geq 2$. Proces taki zachowuje wszystkie cechy BMAP oprócz modelowania rozmiaru pakietu.

³Tzn. BMAP z dyskretnym czasem, więcej informacji np. w [28].



Rys. 4.13. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla różnych wartości obciążenia (Przykład 4.8). $X(0) = 0$, rozkład $J(0)$ dla BMAP wg π
 Fig. 4.13. Average time to buffer overflow versus the buffer size for different values of the offered load (Example 4.8). $X(0) = 0$, $J(0)$ was distributed according to π

D-BMAP [132, 130, 131], MAP [191, 142, 143, 194, 141, 16, 149, 104], czy wreszcie ze specjalnymi typami czasu obsługi [40, 10, 42, 8, 41, 22].

Przedstawione w tym rozdziale wyniki pochodzą z następujących źródeł. Długość kolejki dla procesu BMAP (podrozdział 4.2.1) była tematem artykułów [61, 60, 64]. W szczególności, ogólna postać rozkładu długości kolejki została pokazana w [61], praca [60] jest poświęcona szczegółowym studiom nad prawdopodobieństwem przepełnienia bufora $\Phi(b)$ w stanie ustalonym i nieustalonym, zaś w [64] zademonstrowany został na przykładach obliczeniowych wpływ autokorelacji BMAP na długość kolejki. Wzór na opóźnienie kolejkowania (podrozdział 4.2.2) został udowodniony w [62], natomiast wzór na rozkład czasu do przepełnienia bufora (podrozdział 4.2.3) w pracy [67]. Dodatkowo w pracy [65] można znaleźć przykłady obliczeniowe dla czasu do przepełnienia bufora w zależności od parametrów systemu.

Na zakończenie warto wspomnieć, że BMAP bywa wykorzystywany również w innych dziedzinach, na przykład do modelowania ruchu pojazdów na drogach [9].

5. NARZĘDZIA

W rozdziale tym przedstawione zostaną narzędzia umożliwiające efektywne wykorzystanie wyników teoretycznych otrzymanych w poprzednich rozdziałach w obliczeniach numerycznych. Przede wszystkim zostanie pokazane, jak numerycznie odwracać transformaty Laplace'a i funkcje tworzące, jak obliczać funkcyjony całkowe funkcji liczącej dla procesów markowskich typu MMPP i BMAP, czy wreszcie jak dobierać parametry MMPP i BMAP tak, by otrzymane przy ich pomocy charakterystyki kolejkowania były wiarygodne. Dodatkowo rozdział ten zawiera stosowane wielokrotnie w dowodach twierdzenie o prawdopodobieństwie całkowitym oraz kod do symulacji BMAP.

5.1. Odwracanie transformat

5.1.1. Odwracanie transformaty Laplace'a

W tym podrozdziale zajmiemy się efektywnym wyznaczeniem oryginału, przy założeniu że znamy jego transformatę Laplace'a. Innymi słowy, będziemy poszukiwać funkcji $f(t)$, określonej na półosi rzeczywistej $t > 0$ i przyjmującej wartości rzeczywiste, znając funkcję $f^*(s)$ zdefiniowaną jako

$$f^*(s) = \int_0^{\infty} e^{-st} f(t) dt, \quad (5.1)$$

dla $\text{Re}(s) > 0$.

W zastosowaniach związanych z rachunkiem prawdopodobieństwa transformata postaci (5.1) zwykle oznacza transformatę gęstości rozkładu ($f(t)$ oznacza gęstość). Spotyka się również transformaty postaci

$$\int_0^{\infty} e^{-st} dF(t), \quad (5.2)$$

nazywane transformatami Laplace'a-Stieltjesa. Jeżeli funkcja $F(t)$ oznacza dystrybuantę dla gęstości $f(t)$, to oczywiście całki w (5.1) i (5.2) są równoważne.

Znając transformatę Laplace'a-Stieltjesa dla dystrybuanty $F(t)$, możemy łatwo obliczyć jej (zwykłą) transformatę Laplace'a:

$$F^*(s) = \int_0^{\infty} e^{-st} F(t) dt = \frac{1}{s} \int_0^{\infty} e^{-st} dF(t), \quad (5.3)$$

a zatem nie ma potrzeby zajmować się osobno odwracaniem transformat Laplace'a-Stieltjesa.

Podstawą do prezentowanego tu algorytmu odwracania, pochodzącego z [2], jest wzór całkowy Bromwicha¹, który wyraża oryginał $f(t)$ przy pomocy jego transformaty w następujący sposób:

$$f(t) = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} e^{st} f^*(s) ds, \quad (5.4)$$

gdzie b jest dowolną liczbą rzeczywistą położoną na prawo od wszystkich punktów osobliwych funkcji $f^*(s)$ na płaszczyźnie zespolonej.

Stosując metodę trapezów o kroku h do przybliżonego obliczania całki w (5.4), otrzymujemy

$$f(t) \approx f_h(t) = \frac{he^{bt}}{2\pi} f^*(b) + \frac{he^{bt}}{\pi} \sum_{k=1}^{\infty} \operatorname{Re}(e^{ikh t} f^*(b + ikht)). \quad (5.5)$$

Dobierając następnie $h = \pi/lt$ oraz $b = A/2lt$, gdzie l jest liczbą naturalną, zaś A – rzeczywistą dodatnią, dostaniemy:

$$f_h(t) = f_{A,l}(t) = \frac{e^{A/2l}}{2lt} \left[f^* \left(\frac{A}{2lt} \right) + 2 \sum_{k=1}^{\infty} \operatorname{Re} \left(e^{ik\pi/l} f^* \left(\frac{A}{2lt} + \frac{ik\pi}{lt} \right) \right) \right]. \quad (5.6)$$

Stałe A i l zostały wprowadzone, aby umożliwić kontrolę błędu całkowania metodą trapezów i błędu zaokrąglenia (szczegóły zobacz w [2]; zwykle stosuje się wartości $A = 19$, $l = 1$).

Proste przekształcenia algebraiczne (5.6) dają następnie

$$f_{A,l}(t) = \sum_{k=0}^{\infty} (-1)^k a_k(t), \quad (5.7)$$

¹Warto pamiętać, że nie jest to jedyny wzór tego typu; znane są też wzory oparte na różniczkowaniu, a nie całkowaniu (wzór Posta-Widdera [2]) czy wzór oparty na funkcjach Laguerre [1].

gdzie

$$a_k(t) = \frac{e^{A/2t}}{2lt} b_k(t), \quad k \geq 0, \quad (5.8)$$

$$b_0(t) = f^* \left(\frac{A}{2lt} \right) + 2 \sum_{j=1}^l \operatorname{Re} \left[f^* \left(\frac{A}{2lt} + \frac{ij\pi}{lt} \right) e^{ij\pi/t} \right],$$

$$b_k(t) = 2 \sum_{j=1}^l \operatorname{Re} \left[f^* \left(\frac{A}{2lt} + \frac{ij\pi}{lt} + \frac{ik\pi}{t} \right) e^{ij\pi/t} \right], \quad k \geq 1.$$

W zasadzie wzoru (5.7) można już używać do odwracania $f^*(s)$, jednak zbieżność występującego tam szeregu może być wolna, co powoduje konieczność obliczenia dużej liczby współczynników $a_k(t)$. Dlatego też warto się posłużyć wzorem Eulera służącym do obliczania sumy szeregu o naprzemiennych znakach. Wzór ten ma postać

$$\sum_{k=0}^{\infty} (-1)^k a_k \approx \sum_{k=0}^m \binom{m}{k} 2^{-m} \sum_{j=0}^{n+k} (-1)^j a_j,$$

gdzie m i n są liczbami naturalnymi, zwykle n rzędu kilkudziesięciu, zaś m rzędu kilkunastu (np. $n = 38$, $m = 11$ wg [2]).

Jak widać, wzór Eulera oblicza sumę szeregu przez odpowiednio ważoną średnią sum częściowych od n do $n + m$. Jak się okazuje, pozwala on znacznie przyspieszyć obliczanie sumy w (5.7). Łatwo się o tym przekonać, wykonując obliczenia dla prostych przykładów ciągów a_k . Na przykład, przy $a_k = k^{-1}$ wystarczy około 50 składników, by otrzymać dokładność sumy rzędu 10^{-8} , podczas gdy bezpośrednie obliczenia wymagają sumowania aż 10^8 składników szeregu w celu otrzymania takiej samej dokładności.

A zatem, podsumowując powyższe wyniki, otrzymujemy co następuje.

Oryginał $f(t)$ dla transformaty $f^(s)$ obliczamy*

$$f(t) \approx \sum_{k=0}^m \sum_{j=0}^{n+k} \binom{m}{k} 2^{-m} (-1)^j a_j(t),$$

gdzie współczynniki $a_k(t)$ są określone w (5.8), zaś typowe wartości pozostałych parametrów to $n = 38$, $m = 11$, $A = 19$, $l = 1$.

Największą zaletą zaprezentowanej metody jest to, że jest ona szybka i bardzo dobrze sprawdza się w praktyce. Za jej wadę można uznać fakt, że nie jest znane

oszacowanie błędu wyniku, głównie z uwagi na brak oszacowania błędu dla sumy Eulera. Jak pokazuje praktyka, dobre oszacowanie błędu otrzymujemy, wykonując dwukrotnie obliczenia i zmieniając przy tym o 1 któryś z parametrów: m , A lub l . Innymi słowy, różnica wyników dla $m = 10$ i dla $m = 11$ stanowi dobre oszacowanie błędu, podobnie dla A i l . W przypadku otrzymania niezadowolających wyników zwiększamy wartość n .

Ponadto należy zauważyć, że pokazana metoda nie powinna być stosowana dla bardzo małych t , gdyż czynnik

$$\frac{e^{A/2l}}{2lt}$$

w (5.8) dąży do nieskończoności dla $t \rightarrow 0$. Na szczęście fakt ten ma małe znaczenie praktyczne, gdyż zachowanie oryginału dla małych t można łatwo zbadać, wykorzystując własność transformaty Laplace'a, która mówi, że

$$sf^*(s) \rightarrow f(0) \text{ dla } s \rightarrow \infty,$$

jeśli tylko

$$f(t) \rightarrow f(0) \text{ dla } t \rightarrow 0.$$

5.1.2. Odwracanie dwuwymiarowej transformaty Laplace'a

Będziemy teraz chcieli znaleźć oryginał dla transformaty postaci

$$f^*(s_1, s_2) = \int_0^\infty \int_0^\infty e^{-s_1 t_1 - s_2 t_2} f(t_1, t_2) dt_1 dt_2,$$

gdzie

$$t_1 > 0, \quad t_2 > 0, \quad \operatorname{Re}(s_1) > 0, \quad \operatorname{Re}(s_2) > 0,$$

oraz funkcja $f(t_1, t_2)$ przyjmuje wartości rzeczywiste.

W [45] zaproponowano wzór oparty na sumowaniu Poissona. Wygląda on następująco.

Oryginał $f(t_1, t_2)$ dla transformaty $f^*(s_1, s_2)$ obliczamy

$$\begin{aligned}
& f(t_1, t_2) \\
& \approx \frac{\exp(A_1/(2l_1) + A_2/(2l_2))}{4t_1l_1t_2l_2} \\
& \times \left\{ 2 \sum_{k_1=1}^{l_2} \sum_{k=0}^{\infty} (-1)^k \operatorname{Re} \left[\exp\left(-\frac{ik_1\pi}{l_2}\right) f^*\left(\frac{A_1}{2t_1l_1}, \frac{A_2}{2t_2l_2} - \frac{ik_1\pi}{t_2l_2} - \frac{ik\pi}{t_2}\right) \right] \right. \\
& + 2 \sum_{j_1=1}^{l_1} \sum_{j=0}^{\infty} (-1)^j \operatorname{Re} \left[\sum_{k_1=1}^{l_2} \sum_{k=0}^{\infty} (-1)^k \exp\left(-\frac{ij_1\pi}{l_1} - \frac{ik_1\pi}{l_2}\right) \right. \\
& \quad \times \left. f^*\left(\frac{A_1}{2t_1l_1} - \frac{ij_1\pi}{t_1l_1} - \frac{ij\pi}{t_1}, \frac{A_2}{2t_2l_2} - \frac{ik_1\pi}{t_2l_2} - \frac{ik\pi}{t_2}\right) \right] \\
& + 2 \sum_{j_1=1}^{l_1} \sum_{j=0}^{\infty} (-1)^j \operatorname{Re} \left[\exp\left(-\frac{ij_1\pi}{l_1}\right) f^*\left(\frac{A_1}{2t_1l_1} - \frac{ij_1\pi}{t_1l_1} - \frac{ij\pi}{t_1}, \frac{A_2}{2t_2l_2}\right) \right. \\
& \quad + \sum_{k_1=1}^{l_2} \sum_{k=0}^{\infty} (-1)^k \exp\left(-\frac{ij_1\pi}{l_1} - \frac{ik_1\pi}{l_2}\right) \\
& \quad \times \left. f^*\left(\frac{A_1}{2t_1l_1} - \frac{ij_1\pi}{t_1l_1} - \frac{ij\pi}{t_1}, \frac{A_2}{2t_2l_2} + \frac{ik_1\pi}{t_2l_2} + \frac{ik\pi}{t_2}\right) \right] \\
& \left. + f^*\left(\frac{A_1}{2t_1l_1}, \frac{A_2}{2t_2l_2}\right) \right\},
\end{aligned}$$

gdzie typowe wartości parametrów to $A_1 = A_2 = 19$, $l_1 = l_2 = 2$.

Ponadto sumy postaci

$$\sum_{k=0}^{\infty} (-1)^k a_k$$

w powyższym wzorze również mogą być obliczane przy pomocy wzoru Eulera. Podobnie jak poprzednio, można użyć parametrów $n = 38$, $m = 11$.

Błąd tej metody związany z przybliżonym całkowaniem jest mniejszy niż

$$\frac{C(e^{-A_1} + e^{-A_2} - e^{-A_1-A_2})}{(1 - e^{-A_1})(1 - e^{-A_2})}, \quad (5.9)$$

jeżeli $f(t_1, t_2) \leq C$ dla wszystkich t_1, t_2 . Jak widać, manipulując parametrami A_1 i A_2 , możemy dowolnie ten błąd obniżyć. Dla $C = 1$ i $A_1 = A_2 = 19.1$ wynosi on 10^{-8} , zaś dla $A_1 = A_2 = 28.3$ jest równy 10^{-12} .

Niestety, wzór (5.9) nie daje oszacowania całkowitego błędu, gdyż pojawia się także błąd zaokrąglenia, związany z mnożeniem małych liczb przez bardzo duże. W szczególności wyrażenie

$$\frac{\exp(A_1/(2l_1) + A_2/(2l_2))}{4t_1l_1t_2l_2}$$

może przyjmować bardzo duże wartości. Do kontroli błędu zaokrągleń używamy parametrów l_1 i l_2 . Zwykle $l_1 = l_2 = 1$ wystarcza do otrzymania wyniku z dokładnością do 5-6 cyfr, zaś $l_1 = l_2 = 2$ z dokładnością do 10 cyfr.

Wreszcie sumowanie szeregów przy pomocy sumy Eulera jest źródłem dodatkowego błędu. Błąd ten szacujemy w praktyce, obliczając sumę dla $m = 10$ oraz $m = 11$, a następnie obliczając różnicę wyników.

5.1.3. Odwracanie funkcji tworzących

Kolejną bardzo ważną umiejętnością jest obliczanie wyrazów ciągu liczb rzeczywistych $\{q_k\}_{k=0}^{\infty}$, przy założeniu że znamy jego funkcję tworzącą

$$q(z) = \sum_{k=0}^{\infty} q_k z^k. \quad (5.10)$$

Czasami do tego celu wystarczy się posłużyć wielokrotnym różniczkowaniem funkcji $q(z)$ i obliczaniem wartości uzyskanych pochodnych w punkcie $z = 0$. Bardzo często jednak takie podejście nie sprawdza się, gdyż pojawiają się trudności związane z symbolicznym czy numerycznym różniczkowaniem. Dlatego też w ogólnym wypadku lepiej się posłużyć bardziej zaawansowaną techniką. W [2] zaproponowano następującą metodę odwracania funkcji tworzącej opartą na wzorze całkowym Cauchy'ego i całkowaniu numerycznym metodą trapezów.

Element q_k ciągu o funkcji tworzącej $q(z)$ obliczamy

$$q_k \approx \frac{1}{2klr^k} \left(a_0(k, l, r) + (-1)^k a_k(k, l, r) + 2 \sum_{j=1}^{k-1} (-1)^j \operatorname{Re}(a_j(k, l, r)) \right), \quad (5.11)$$

gdzie

$$a_j(k, l, r) = \sum_{n=0}^{l-1} e^{-\pi i n/l} q(re^{\pi i(n+l_j)/lk}),$$

zaś typowe wartości parametrów to $l = 1$, $r = 10^{-4/k}$.

Tym razem nie jest konieczne stosowanie sumy Eulera, gdyż w powyższym wzorze nie występują nieskończone szeregi. Dlatego też błąd metody jest związany tylko z przybliżonym całkowaniem oraz z zaokrągleniami. W szczególności, błąd całkowania metodą trapezów jest nie większy od

$$\frac{Cr^{2kl}}{1-r^{2kl}},$$

jeżeli tylko $|q_k| \leq C$ dla wszystkich k . Dla $l = 1$, $r = 10^{-4/k}$ daje to błąd rzędu $C10^{-8}$. Z kolei błąd zaokrąglenia wynosi w przybliżeniu

$$2klr^k 10^{-m},$$

jeżeli suma w nawiasach w (5.11) jest obliczona z dokładnością do 10^{-m} .

Czasami zachodzi również potrzeba odwrócenia dwuwymiarowej funkcji tworzącej. Dla ciągu $p_{n_1 n_2}$, $n_1 = 0, 1, \dots$, $n_2 = 0, 1, \dots$ funkcja ta jest zdefiniowana jako

$$p(z_1, z_2) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} p_{n_1 n_2} z_1^{n_1} z_2^{n_2}.$$

Zaproponowana w [45] metoda jest oparta na (dyskretnej) sumie Poissona i ma następującą postać.

Element $p_{n_1 n_2}$ ciągu o funkcji tworzącej $p(z_1, z_2)$ obliczamy jako

$$p_{n_1 n_2} \approx \frac{1}{2l_1 n_1 r_1^{n_1}} \sum_{j_1=0}^{l_1-1} \sum_{j=-n_1}^{n_1-1} (-1)^j \exp\left(-\frac{\pi i j_1}{l_1}\right) \\ \times \left\{ \frac{1}{2l_2 n_2 r_2^{n_2}} \sum_{k_1=0}^{l_2-1} \sum_{k=-n_2}^{n_2-1} (-1)^k \exp\left(-\frac{\pi i k_1}{l_2}\right) \right. \\ \left. \times p\left(r_1 \exp\left(\frac{\pi i (j_1 + l_1 j)}{l_1 n_1}\right), r_2 \exp\left(\frac{\pi i (k_1 + l_2 k)}{l_2 n_2}\right)\right) \right\}.$$

Wreszcie możemy mieć też do czynienia z transformatą mieszaną postaci

$$f^*(s, z) = \int_0^{\infty} \sum_{n=0}^{\infty} f(t, n) e^{-st} z^n dt,$$

gdzie $t > 0$, $Re(s) > 0$, $n = 0, 1, \dots$. Zgodnie z [45] transformatę tę odwracamy jak następuje.

Oryginał $f(t, n)$ dla transformaty $f^*(s, z)$ obliczamy

$$f(t, n) \approx \frac{\exp(A/(2l_1))}{2l_1 t} \sum_{j_1=1}^{l_1} \sum_{j=-\infty}^{\infty} (-1)^j \exp\left(-\frac{\pi i j_1}{l_1}\right) \\ \times \left\{ \frac{1}{2l_2 n_2 r_2^n} \sum_{k_1=0}^{l_2-1} \sum_{k=-n}^{n-1} (-1)^k \exp\left(-\frac{\pi i k_1}{l_2}\right) \right. \\ \left. \times f^*\left(\frac{A}{2l_1 t} - \frac{i j_1 \pi}{l_1 t} - \frac{i j \pi}{t}, r \exp\left(\frac{\pi i (k_1 + l_2 k)}{l_2 n}\right)\right) \right\}.$$

Występującą powyżej sumę od $j = -\infty$ do $j = \infty$ obliczamy, rozbijając ją na dwie sumy od $j = 0$ do $j = \infty$ i wykorzystując sumę Eulera do każdej z nich.

5.2. Metoda uniformizacji

Metoda uniformizacji (zobacz np. [137, 78]), która ma wiele zastosowań w obliczeniach numerycznych związanych z procesami markowskimi, opiera się na następującym fakcie.

Jeżeli dany jest łańcuch Markowa z ciągłym czasem o macierzy intensywności Q , to łańcuch ten jest równoważny procesowi, w którym odstępy czasowe pomiędzy zdarzeniami tworzą proces Poissona o intensywności

$$\theta = \max_i (-Q_{ii}),$$

zaś w chwili pojawienia się zdarzenia następuje zmiana stanu łańcucha tak jak w łańcuchu Markowa z dyskretnym czasem o macierzy przejścia

$$L = I + \theta^{-1}Q.$$

Rozkład oryginalnego łańcucha w chwili t można wyrazić wtedy w następujący sposób:

$$e^{Qt} = e^{\theta t(L-I)} = e^{-\theta t} e^{\theta t L} = \sum_{n=0}^{\infty} e^{-\theta t} \frac{(\theta t)^n}{n!} L^n.$$

Wykorzystamy teraz metodę uniformizacji do efektywnego obliczania funkcjonalów całkowitych funkcji liczącej dla BMAP, w szczególności takich jak we wzorach (4.5)-(4.10) (analogiczne funkcjonały dla MMPP otrzymamy, podstawiając jedynie $D_0 = Q - \Lambda$ i $D_1 = \Lambda$).

Rozpocznijmy od samej funkcji liczącej dla BMAP, tj.

$$P_{i,j}(n, t) = \mathbf{P}(N(t) = n, J(t) = j | N(0) = 0, J(0) = i),$$

$$P(n, t) = [P_{i,j}(n, t)]_{i,j}.$$

Wykorzystując uniformizację, można pokazać, że

$$P(n, t) = \sum_{j=0}^{\infty} e^{-\theta t} \frac{(\theta t)^j}{j!} K_{n,j},$$

gdzie

$$\theta = \max_i \{(-D_0)_{ii}\},$$

zaś $K_{n,j}$ oznacza ciąg macierzy $m \times m$ zdefiniowany rekurencyjnie jako:

$$\begin{aligned} K_{0,0} &= I, \\ K_{n,0} &= \mathbf{0}, \quad n \geq 1, \\ K_{0,j+1} &= K_{0,j}(I + \theta^{-1}D_0), \\ K_{n,j+1} &= \theta^{-1} \sum_{i=0}^{n-1} K_{i,j} D_{n-i} + K_{n,j}(I + \theta^{-1}D_0). \end{aligned}$$

Jeżeli teraz interesuje nas funkcjonał postaci:

$$L_n = \int_0^{\infty} g(t) P(n, t) dt,$$

to możemy go przedstawić jako

$$L_n = \sum_{j=0}^{\infty} \gamma_j K_{n,j},$$

gdzie

$$\gamma_j = \int_0^{\infty} e^{-\theta t} \frac{(\theta t)^j}{j!} g(t) dt. \quad (5.12)$$

A zatem cały problem został sprowadzony do obliczania zwykłych całek w postaci (5.12). Całki te można często obliczać symbolicznie, w przeciwnym wypadku zwykle bardzo dobrze całkują się numerycznie.

Możemy teraz podać konkretne wzory dla funkcjonałów $A_k(s)$, $\bar{D}_k(s)$, $C_k(s, \sigma)$ określonych w (4.5) - (4.10). Mamy zatem:

$$A_k(s) = \sum_{j=0}^{\infty} \gamma_j(s) K_{k,j}, \quad \text{dla} \quad \gamma_j(s) = \int_0^{\infty} \frac{e^{-(\theta+s)t} (\theta t)^j}{j!} dF(t),$$

$$\overline{D}_k(s) = \sum_{j=0}^{\infty} \delta_j(s) K_{k,j} \quad \text{dla} \quad \delta_j(s) = \int_0^{\infty} \frac{e^{-(\theta+s)t} (\theta t)^j}{j!} (1 - F(t)) dt,$$

$$C_k(s, \sigma) = \sum_{j=0}^{\infty} \kappa_j(s, \sigma) K_{k,j} \quad \text{dla}$$

$$\kappa_j(s, \sigma) = \int_0^{\infty} \frac{e^{-(\theta+\sigma)t} (\theta t)^j}{j!} dt \int_0^{\infty} e^{-sx} d_x F(x+t).$$

Jeżeli czas obsługi jest stały i równy d , to współczynniki $\gamma_j(s)$ i $\delta_j(s)$ można przedstawić w jawnej postaci:

$$\begin{aligned} \gamma_j(s) &= \frac{e^{-(\theta+s)d} (\theta d)^j}{j!}, \\ \delta_j(s) &= \theta^j \frac{\Gamma(j+1, 0) - \Gamma(j+1, d(s+\theta))}{j! (s+\theta)^{j+1}}, \end{aligned}$$

gdzie $\Gamma(j, x)$ oznacza niekompletną funkcję gamma. Można łatwo otrzymać analogiczne wzory dla kilku innych prostych postaci rozkładu F (np. wykładniczej).

5.3. Estymacja parametrów MMPP i BMAP

Aby móc wiarygodnie przewidywać charakterystyki kolejkowania, należy sparametryzować model ruchu (MMPP lub BMAP) w taki sposób, by posiadał on taką samą probabilistyczną strukturę jak ruch, którego charakterystyki kolejkowania nas interesują. W praktyce robi się to na podstawie zarejestrowanych śladów ruchu (*ang.* trace files).

Parametry prostych i złożonych procesów Poissona dopasowuje się łatwo. Intensywność λ przyjmujemy, mierząc średnią obserwowaną intensywność ruchu, zaś rozkład rozmiaru grupy p_i w złożonym procesie Poissona otrzymujemy na podstawie odpowiedniego histogramu.

Znaczniej trudniej dopasować parametry procesów MMPP i BMAP. Opracowano wiele algorytmów dopasowujących MMPP do śladów ruchu (zobacz np. [81, 179, 98, 209, 188, 181, 187]). Część z nich jest przeznaczona dla MMPP szczególnego typu, jak np. MMPP z tylko dwoma stanami [112, 144, 159]. Algorytmów dla BMAP jest nieco mniej (np. [182, 120, 34, 35]).

Poniżej przedstawiona zostanie metoda EM z [120]. Wprawdzie została ona zaproponowana dla BMAP, ale, jak zostanie pokazane na końcu tego podrozdziału, można ją również z powodzeniem stosować do parametryzacji MMPP.

Metoda EM (*ang.* expectation maximization, zobacz np. [80, 4, 202, 34]) jest wariantem metody największej wiarygodności stosowanym w sytuacji, gdy nie wszystkie zmienne jakiegoś procesu możemy obserwować i zapisywać. Przykładowo, modelując przy pomocy BMAP ruch pakietów IP, możemy obserwować czasy pomiędzy kolejnymi pakietami oraz rozmiary pakietów, nie mamy natomiast informacji o chwilach zmiany stanu łańcucha modulującego i stanach łańcucha (innymi słowy, możemy obserwować jedynie składową $N(t)$ z pary $(J(t), N(t))$ tworzącej realizację BMAP).

Idea metody EM wygląda następująco. Załóżmy, że y oznacza dane dające się zaobserwować. W naszym wypadku $y = \{(t_1, \alpha_1), (t_2, \alpha_2), \dots, (t_M, \alpha_M)\}$, gdzie t_i oznaczają czasy pojawiania się kolejnych pakietów, α_i ich rozmiary w bajtach, zaś M jest liczbą zarejestrowanych pakietów. Niech ϕ oznacza parametryzację procesu. W naszym wypadku ϕ odpowiada macierzom D_i dla BMAP oraz rozkładowi początkowemu stanu łańcucha modulującego (wektorowi ν), tzn.

$$\phi = (\nu, D_1, \dots, D_n),$$

gdzie n jest maksymalnym rozmiarem pakietu, jaki bierzemy pod uwagę.

Przez $\mathcal{L}(\phi, y)$ będziemy oznaczali wiarygodność parametryzacji ϕ dla danych y , zaś przez $\mathcal{L}^c(\phi, x, y)$ kompletną wiarygodność ϕ , z uwzględnieniem niedostępnych danych x . W metodzie EM szukana parametryzacja ϕ jest wyznaczana iteracyjnie według następującego schematu:

$$\phi(r+1) = \arg \max_{\phi} \{\mathbf{E}_{\phi(r)}(\log \mathcal{L}^c(\phi, x, y)|y)\}, \quad (5.13)$$

gdzie $\phi(r)$ oznacza parametryzację otrzymaną w r -tej iteracji, zaś $\mathbf{E}_{\phi(r)}$ warunkową wartość oczekiwaną dla parametryzacji $\phi(r)$. Obliczanie wartości oczekiwanej w wzorze (5.13) nazywamy częścią E algorytmu, natomiast znajdowanie $\arg \max$ – częścią M algorytmu. Można udowodnić, że postępowanie takie daje w kolejnych krokach parametryzacje co najmniej nie gorsze od poprzednich, tzn.

$$\mathcal{L}(\phi(r+1), y) \geq \mathcal{L}(\phi(r), y).$$

Algorytm kończy działanie, gdy osiągnięta zostanie zadowalająca nas zbieżność parametryzacji (czyli wtedy, gdy w kolejnym kroku otrzymamy parametryzację różniącą się co najwyżej o ϵ od poprzedniej) lub po z góry zadanej liczbie iteracji. Początkową parametryzację $\phi(0)$ można przyjąć dowolnie (np. losowo).

Przy estymacji parametrów BMAP mamy

$$\mathcal{L}(\phi, y) = \nu \prod_{k=1}^M f_{\alpha_k}(\Delta t_k) \cdot \mathbf{1},$$

gdzie

$$t_0 = 0, \quad t_M = T, \quad \Delta t_k = t_k - t_{k-1},$$

$$f_j(t) = F^c(t)D_j, \quad F^c(t) = e^{D_0 t}$$

oraz

$$\begin{aligned} \mathcal{L}^c(\phi, x, y) &= \nu_{i_0(1)} \prod_{k=1}^M \prod_{l=0}^{z(k)-1} \mu_{i_l(k)} \exp(-\mu_{i_l(k)} s_l(k)) p_{i_l(k)}(0, i_{l+1}(k)) \\ &\times \mu_{i_{z(k)}(k)} \exp(-\mu_{i_{z(k)}(k)} s_{z(k)}(k)) p_{i_{z(k)}(k)}(\alpha_k, i_0(k+1)), \end{aligned} \quad (5.14)$$

gdzie $z(k)$ oznacza liczbę stanów, którą przyjął łańcuch modulujący w k -tym odstępie pomiędzy pakietami, $i_l(k)$ oznacza l -ty stan łańcucha w k -tym odstępie pomiędzy pakietami, $s_l(k)$ oznacza czas przebywania łańcucha w l -tym stanie w k -tym odstępie pomiędzy pakietami, natomiast prawdopodobieństwa $p_i(k, j)$ są zdefiniowane w (4.2) i (4.3).

Wzór (5.14) można zapisać w prostszej postaci, używając statystyk $T_{i,j}$, s_i oraz $A(1), \dots, A(n)$. $T_{i,j}$ oznacza liczbę zmian stanu bez pojawienia się pakietu:

$$T_{i,j} = \#\{t | 0 \leq t \leq T, J(t-) = i, J(t) = j, N(t) = N(t-)\}, \quad i \neq j,$$

natomiast $[A(k)]_{i,j}$ oznacza liczbę zmian stanu z i na j łącznie z pojawieniem się pakietu o rozmiarze k :

$$[A(k)]_{i,j} = \#\{t_l | 1 \leq l \leq M, J(t_l-) = i, J(t_l) = j, N(t_l) = N(t_l-) + k\}.$$

Wreszcie s_i oznacza całkowity czas, jaki spędził łańcuch modulujący w stanie i :

$$s_i = \int_0^T I(J(t) = i) dt.$$

Stosując te oznaczenia, możemy zapisać (5.14) w bardziej intuicyjny sposób:

$$\begin{aligned} \mathcal{L}^c(\phi, x, y) &= \prod_{i=1}^m \nu_i^{I(J(0)=i)} \prod_{i=1}^m \exp([D_0]_{i,i} s_i) \\ &\times \prod_{i=1}^m \prod_{j=1, j \neq i}^m [D(0)]_{i,j}^{T_{i,j}} \prod_{k=1}^n \prod_{i=1}^m \prod_{j=1}^m [D_k]_{i,j}^{A(k)_{i,j}}. \end{aligned} \quad (5.15)$$

Drugi iloczyn w (5.15) reprezentuje czas przebywania łańcucha modulującego w każdym ze stanów, trzeci iloczyn odpowiada zmianom stanu łańcucha bez pojawienia

się pakietu, zaś ostatni iloczyn odpowiada zmianom stanu łańcucha, którym towarzyszy pakiet.

Z (5.15) otrzymujemy następnie

$$\begin{aligned} & \mathbf{E}_{\phi(r)}\{\log \mathcal{L}^c(\phi, x, y) | N(u), 0 \leq u \leq T\} \\ &= \sum_{i=1}^m \hat{\nu}_i \log(\nu_i) + \sum_{i=1}^m [D_0]_{i,i} \hat{s}_i + \sum_{i=1}^m \sum_{j=1, j \neq i}^m \hat{T}_{i,j} \log[D_0]_{i,j} \\ & \quad + \sum_{k=1}^n \sum_{i=1}^m \sum_{j=1}^m \hat{A}(k)_{i,j} \log[D_k]_{i,j}, \end{aligned} \quad (5.16)$$

gdzie

$$\hat{\nu}_i = \mathbf{P}_{\phi(r)}\{J(0) = i | N(u), 0 \leq u \leq T\}, \quad (5.17)$$

$$\begin{aligned} \hat{s}_i &= \mathbf{E}_{\phi(r)}\{s_i | N(u), 0 \leq u \leq T\} \\ &= \int_0^T \mathbf{P}_{\phi(r)}\{J(t) = i | N(u), 0 \leq u \leq T\} dt, \end{aligned} \quad (5.18)$$

$$\begin{aligned} \hat{T}_{i,j} &= \int_0^T \mathbf{P}_{\phi(r)}\{J(t-) = i, J(t) = j, N(t) = N(t-) \\ & \quad | N(u), 0 \leq u \leq T\} dt, \end{aligned} \quad (5.19)$$

$$\begin{aligned} \hat{A}(k)_{i,j} &= \sum_{l=1}^M \mathbf{P}_{\phi(r)}\{J(t_l-) = i, J(t_l) = j, \\ & \quad N(t_l) = N(t_l-) + k | N(u), 0 \leq u \leq T\}. \end{aligned} \quad (5.20)$$

Wzór (5.16) przedstawia część E algorytmu EM. Aby możliwa była implementacja tej części, należy jeszcze przedstawić wyrażenia (5.17) – (5.20) w postaci umożliwiającej obliczenia numeryczne. Mamy kolejno:

$$\begin{aligned} \hat{\nu}_i &= \mathbf{P}_{\phi(r)}\{J(0) = i | N(u), 0 \leq u \leq T\} \\ &= \frac{\mathbf{P}_{\phi(r)}\{J(0) = i, (N(u), 0 \leq u \leq T)\}}{\mathbf{P}_{\phi(r)}\{N(u), 0 \leq u \leq T\}} \\ &= \frac{\nu_i(r)}{\mathcal{L}(\phi(r), y)} \left(\prod_{k=1}^M f_{\alpha_k}(\Delta t_k) \cdot \mathbf{1} \right)_i, \end{aligned} \quad (5.21)$$

$$\begin{aligned}
\hat{s}_i &= \int_0^T \mathbf{P}_{\phi(r)} \{J(t) = i | N(u), 0 \leq u \leq T\} dt, \\
&= \int_0^T \frac{\mathbf{P}_{\phi(r)} \{J(0) = i, (N(u), 0 \leq u \leq T)\}}{\mathbf{P}_{\phi(r)} \{N(u), 0 \leq u \leq T\}} dt \\
&= \frac{1}{\mathcal{L}(\phi(r), y)} \int_0^T \nu(r) \prod_{k=1}^{\bar{N}(t)} f_{\alpha_k}(\Delta t_k) F^c(t - t_{\bar{N}(t)}) \cdot \mathbf{1}_i \\
&\quad \cdot \mathbf{1}_i^T \cdot f_{\alpha_{\bar{N}(t)+1}}(t_{\bar{N}(t)+1} - t) \cdot \prod_{k=\bar{N}(t)+2}^M f_{\alpha_k}(\Delta t_k) \cdot \mathbf{1} dt, \tag{5.22}
\end{aligned}$$

gdzie $\mathbf{1}_i$ oznacza wektor kolumnowy mający na i -tym miejscu jedynkę i zera na pozostałych miejscach, natomiast $\bar{N}(t)$ liczbę pakietów² w przedziale $(0, t]$. Dalej,

$$\begin{aligned}
\hat{T}_{i,j} &= \int_0^T \mathbf{P}_{\phi(r)} \{J(t-) = i, J(t) = j, N(t) = N(t-) \\
&\quad | N(u), 0 \leq u \leq T\} dt \\
&= \int_0^T \mathbf{P}_{\phi(r)} \{J(t-) = i, J(t) = j, N(t) = N(t-), \\
&\quad (N(u), 0 \leq u \leq T)\} / \mathbf{P}_{\phi(r)} \{N(u), 0 \leq u \leq T\} dt, \\
&= \frac{1}{\mathcal{L}(\phi(r), y)} \int_0^T \nu(r) \prod_{k=1}^{\bar{N}(t)} f_{\alpha_k}(\Delta t_k) F^c(t - t_{\bar{N}(t)}) \cdot \mathbf{1}_i \cdot [D_0]_{i,j} \\
&\quad \cdot \mathbf{1}_j^T \cdot f_{\alpha_{\bar{N}(t)+1}}(t_{\bar{N}(t)+1} - t) \\
&\quad \cdot \prod_{k=\bar{N}(t)+2}^M f_{\alpha_k}(\Delta t_k) \cdot \mathbf{1} \cdot I(N(t) = N(t-)) dt, \tag{5.23}
\end{aligned}$$

$$\begin{aligned}
\hat{A}(k)_{i,j} &= \sum_{l=1}^M \mathbf{P}_{\phi(r)} \{J(t_l-) = i, J(t_l) = j, \\
&\quad N(t_l) = N(t_l-) + k | N(u), 0 \leq u \leq T\}
\end{aligned}$$

²W odróżnieniu od $N(t)$, które oznacza liczbę bajtów w przedziale $[0, t)$, tzn. pakietów z rozmiarami.

$$\begin{aligned}
&= \sum_{l=1}^M \mathbf{P}_{\phi(r)} \{J(t_{l-}) = i, J(t_l) = j, N(t_l) = N(t_{l-}) + k, \\
&\quad (N(u), 0 \leq u \leq T)\} / \mathbf{P}_{\phi(r)} \{N(u), 0 \leq u \leq T\} \\
&= \frac{1}{\mathcal{L}(\phi(r), y)} \sum_{l=1}^M \nu(r) \prod_{u=1}^{l-1} f_{\alpha_u}(\Delta t_u) F^c(\Delta t_l) \cdot \mathbf{1}_i \cdot [D_k]_{i,j} \\
&\quad \cdot \mathbf{1}_j^T \cdot \prod_{u=l+1}^M f_{\alpha_u}(\Delta t_u) \cdot \mathbf{1} \cdot I(N(t_l) = N(t_{l-}) + k) dt. \quad (5.24)
\end{aligned}$$

Można sprawdzić, że (5.16) osiąga maksimum dla

$$\hat{\nu}_i = \nu_i(r) \cdot \left(\prod_{k=1}^M f_{\alpha_k}(\Delta t_k) \cdot \mathbf{1} \right)_i / \mathcal{L}(\phi(r), y), \quad i = 1, \dots, m, \quad (5.25)$$

$$[\hat{D}_0]_{i,j} := \hat{T}_{i,j} / \hat{s}_i, \quad i, j = 1, \dots, m, \quad (5.26)$$

$$[\hat{D}_k]_{i,j} := [\hat{A}_k]_{i,j} / \hat{s}_i, \quad i, j = 1, \dots, m, \quad (5.27)$$

$$[\hat{D}_0]_{i,i} := - \sum_{j=1, j \neq i}^m [\hat{D}_0]_{i,j} - \sum_{u=1}^n \sum_{v=1}^m [\hat{D}_u]_{i,v}, \quad i = 1, \dots, m, \quad (5.28)$$

A zatem wzory (5.25) – (5.28) przedstawiają część M algorytmu.

Końcowy algorytm zostanie zapisany przy pomocy wektorów $L(k)$ i wektorów kolumnowych $R(k)$ zdefiniowanych rekurencyjnie:

$$L(0) = \nu, \quad L(k) = L(k-1) f_{\alpha_k}(\Delta t_k), \quad k = 1, \dots, M,$$

$$R(M+1) = \mathbf{1}, \quad R(k) = f_{\alpha_k}(\Delta t_k) R(k+1), \quad k = M, \dots, 1.$$

Szczegółowa realizacja algorytmu EM dla BMAP zapisana w pseudokodzie została przedstawiona na rys. 5.1. Znaczenie poszczególnych symboli jest następujące: m – liczba stanów łańcucha modulującego w BMAP, n – maksymalny rozmiar grupy (pakietu), M – liczba pakietów w próbkę. Wymiary poszczególnych obiektów są następujące: T , A_k oraz D_0, \dots, D_n są macierzami $m \times m$; ν , $L(k)$ i $\mathbf{1}_i^T$ są wektorami o rozmiarze m , zaś $R(k)$ i $\mathbf{1}_j$ są wektorami kolumnowymi o rozmiarze m . Występującą w algorytmie całkę W można efektywnie obliczać, wykorzystując następujący wzór (otrzymany dzięki metodzie uniformizacji):

$$W = W(i, j, k) = \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} a_k(u, i) \alpha_k(v, j) e^{-q \Delta t_k} \frac{q^{u+v}}{(u+v)!} \cdot \frac{(\Delta t_k)^{u+v+1}}{u+v+1},$$

gdzie

$$\begin{aligned} a_k(u, i) &= L(k-1) \cdot A^u \cdot \mathbf{1}_i, \\ \alpha_k(v, j) &= \mathbf{1}_j^T \cdot A^v \cdot D_{\alpha_k} \cdot R(k+1), \\ A &= \frac{1}{q}D + I, \quad q = 1.02 \cdot \max_i |[D]_{i,i}|. \end{aligned}$$

Wyjaśnijmy, że przeskalowanie 1.02 wprowadza się, by dyskretny łańcuch Markowa z macierzą przejścia A nie był okresowy.

Przedstawiony algorytm posiada następujące własności.

- Kolejne iteracje zachowują zera w macierzach parametryzujących BMAP, co umożliwia dopasowanie parametrów MMPP przy użyciu dokładnie tego samego algorytmu. Wystarczy przyjąć $n = 1$ oraz D_1 zainicjować tak, by poza przekątną znajdowały się zera. Otrzymane parametry MMPP to $Q = D_0 + D_1$ i $\Lambda = D_1$.
- W typowych sytuacjach zbieżność uzyskuje się po kilkuset iteracjach, choć już po 50-100 iteracjach występują zwykle małe zmiany parametrów i funkcji wiarygodności.
- Algorytm jest słabo wrażliwy na początkowy zestaw parametrów, tzn. dla różnych początkowych parametryzacji liczba iteracji oraz jakość otrzymanych wyników są bardzo podobne. Dlatego też w praktyce wystarczy stosować losową inicjalizację.
- Złożoność czasowa algorytmu jest rzędu $O(Mm^2)$. We wzorze tym pominięty został fragment nm^2 związany z pierwszą pętlą na rys. 5.1, gdyż w praktyce M jest zwykle o co najmniej trzy rzędy wielkości większe niż n .

5.4. Prawdopodobieństwo całkowite

Ponieważ w dowodach twierdzeń wielokrotnie wykorzystywana jest ciągła wersja twierdzenia o prawdopodobieństwie całkowitym, dla wygody Czytelnika przypomnimy teraz to twierdzenie.

Na elementarnym kursie rachunku prawdopodobieństwa uczy się podstawowej (dyskretnej) wersji twierdzenia o prawdopodobieństwie całkowitym, które mówi, że jeżeli B_1, B_2, \dots, B_n są zdarzeniami parami wyłączającymi się (tzn. $B_i B_j = \emptyset, i \neq j$) oraz $\mathbf{P}(B_i) > 0, i = 1, \dots, n$, to prawdopodobieństwo dowolnego zdarzenia A zawartego w

$$\bigcup_{j=1}^n B_j$$

$r := 0$ oraz inicjalizacja ν, D_0, \dots, D_n (np. losowa)

do

Część E:

for $i, j = 1, \dots, m, k = 1, \dots, n$

$$T_{i,j} := 0$$

$$[A(k)]_{i,j} := 0$$

$$L(0) = \nu$$

for $k = 1, \dots, M$ $L(k) := L(k-1)f_{\alpha_k}(\Delta t_k)$

$$R(M+1) = \mathbf{1}$$

for $k = M, \dots, 1$ $R(k) = f_{\alpha_k}(\Delta t_k)R(k+1)$

for $i, j = 1, \dots, m, k = 1, \dots, M$

$$W = L(k-1) \int_{t_{k-1}}^{t_k} F^c(t-t_{k-1}) \cdot \mathbf{1}_i \cdot \mathbf{1}_j^T \cdot f_{\alpha_k}(t_k-t) dt \cdot R(k+1)$$

$$T_{i,j} := T_{i,j} + W$$

$$[A(\alpha_k)]_{i,j} := [A(\alpha_k)]_{i,j} + L(k-1) \cdot F^c(\Delta t_k) \cdot \mathbf{1}_i \cdot \mathbf{1}_j^T \cdot R(k+1)$$

Obliczenie wiarygodności: $\mathcal{L} = \nu R(1)$

Część M:

for $i, j = 1, \dots, m, k = 1, \dots, n$

$$\nu_i := \nu_i \cdot \mathbf{1}_i^T \cdot R(1) / \mathcal{L}$$

$$[D_0]_{i,j} := T_{i,j} \cdot [D_0]_{i,j} / T_{i,i}$$

$$[D_k]_{i,j} := [A(k)]_{i,j} \cdot [D_k]_{i,j} / T_{i,i}$$

$$[D_0]_{i,i} := -\sum_{j=1, j \neq i}^m [D_0]_{i,j} - \sum_{u=1}^n \sum_{v=1}^m [D_u]_{i,v}$$

$r := r + 1$

until (zbieżność lub $r = r_{max}$)

Rys. 5.1. Algorytm EM dla BMAP

Fig. 5.1. EM algorithm for the BMAP

możemy obliczyć jako

$$\mathbf{P}(A) = \sum_{j=1}^n \mathbf{P}(A|B_j)\mathbf{P}(B_j). \quad (5.29)$$

Dowód tego faktu jest elementarny.

Intuicyjnie jest jasne, że jeżeli zamiast zdarzeń B_i użyjemy zdarzeń postaci $\eta = x$, gdzie x jest liczbą, zaś η zmienną losową, to twierdzenie pozostanie prawdziwe. Udowodnienie tego wymaga jednak pewnych zabiegów technicznych i odpowiedniego sformułowania, gdyż zdarzenia $\eta = x$ mogą mieć teraz (i często mają) zerowe prawdopodobieństwa.

Twierdzenie 5.1 (o prawdopodobieństwie całkowitym). *Jeżeli zmienna losowa η ma dystrybuantę $F_\eta(y)$, to*

$$F_\xi(x) = \mathbf{P}(\xi < x) = \int_{-\infty}^{\infty} \mathbf{P}(\xi < x|\eta = y)dF_\eta(y) \quad (5.30)$$

zachodzi dla dowolnej zmiennej losowej ξ .

Jak widać, sumę we wzorze (5.29) względem prawdopodobieństw $\mathbf{P}(B_j)$ zastąpiła całka względem rozkładu η .

Dowód Twierdzenia 5.1. Dowód przeprowadzimy, przy założeniu że zmienne losowe ξ i η mają łączną gęstość rozkładu prawdopodobieństwa $f(x, y)$. W ogólnym przypadku dowód wymaga zastosowania bardziej zaawansowanych narzędzi matematycznych.

Oznaczmy gęstości brzegowe:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x, y)dy,$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Jeżeli dla pewnego y zachodzi $f_\eta(y) > 0$, to korzystając z definicji prawdopodobieństwa warunkowego dla zmiennych losowych (zobacz np. [32], str. 140), otrzymujemy:

$$f_\eta(y)\mathbf{P}(\xi < x|\eta = y) = \int_{-\infty}^x f(u, y)du. \quad (5.31)$$

Jeżeli zaś $f_\eta(y) = 0$, to

$$\int_{-\infty}^x f(u, y) du = 0$$

dla wszystkich x . Istotnie, mamy

$$\int_{-\infty}^x f(u, y) du \leq \int_{-\infty}^{\infty} f(u, y) du = f_\eta(y) = 0.$$

Jeśli więc $f_\eta(y) = 0$, to równość (5.31) jest spełniona automatycznie. A zatem jest ona spełniona dla wszystkich y .

Teraz, całkując (5.31) względem y od $-\infty$ do $+\infty$, otrzymujemy

$$\begin{aligned} \int_{-\infty}^{\infty} f_\eta(y) \mathbf{P}(\xi < x | \eta = y) dy &= \int_{-\infty}^{\infty} \int_{-\infty}^x f(u, y) du dy \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, y) dy du \\ &= \int_{-\infty}^x f_\xi(u) du = \mathbf{P}(\xi < x), \end{aligned}$$

co kończy dowód, ponieważ

$$\int_{-\infty}^{\infty} f_\eta(y) \mathbf{P}(\xi < x | \eta = y) dy = \int_{-\infty}^{\infty} \mathbf{P}(\xi < x | \eta = y) dF_\eta(y).$$

■

Jak widać, jeżeli zmienna losowa η ma gęstość $f_\eta(y)$, to wzór (5.30) upraszcza się do postaci

$$\mathbf{P}(\xi < x) = \int_{-\infty}^{\infty} \mathbf{P}(\xi < x | \eta = y) f_\eta(y) dy. \quad (5.32)$$

5.5. Symulacja BMAP

Proces BMAP można bardzo łatwo symulować, wykorzystując do tego celu jego konstruktywną definicję przedstawioną na stronie 109. Niemal bezpośrednio zapisując tę definicję w języku C, otrzymujemy następujący kod do symulacji BMAP:

```
double sum1,sum2=0,los;
int j,k;
do
{
    sum1=0;
    los=uniform(0,1);
    for(j=0;j<=N;j++)
    {
        for(k=1;k<=m;k++)
        {
            sum1+=p[J][j][k];
            if (sum1>los) break;
        }
        if (sum1>los) break;
    }
    sum2+=-log(1-uniform(0,1))/mi[J];
    J=k;
}while (j==0);
size=j;
interarrival_t=sum2;
```

Zmienna J zawiera stan łańcucha modulującego, N – maksymalny rozmiar grupy zdarzeń, tablice mi i p zawierają odpowiednio intensywności μ_i oraz prawdopodobieństwa $p_i(j, k)$, zaś $uniform(0, 1)$ oznacza generator liczb pseudolosowych o rozkładzie jednostajnym na przedziale $(0, 1)$.

Po jednorazowym wykonaniu tego kodu otrzymujemy czas do następnej grupy w BMAP (zmienna `interarrival_t`) oraz rozmiar następnej grupy (zmienna `size`). Wykonując kod wielokrotnie możemy otrzymywać dowolnie długie realizacje procesu BMAP.

PODSUMOWANIE

Praca ta podsumowuje wyniki wieloletnich badań autora nad charakterystykami kolejowania strumieni pakietów w sieciach pakietowych. W badaniach tych do modelowania ruchu pakietów używane były procesy markowskie, wybrane ze względu na duże możliwości naśladowania własności statystycznych ruchu, względną prostotę i dostępność ogólnych narzędzi analitycznych do ich badania.

Do wyznaczania charakterystyk kolejowania wykorzystane zostało nowatorskie podejście. W podejściu tym najpierw budowany jest układ równań całkowych dla interesującej nas charakterystyki, później układ ten jest przekształcany do prostszej postaci przy wykorzystaniu transformaty Laplace'a, a następnie jego rozwiązanie zostaje przedstawione przy pomocy pewnego ciągu rekurencyjnego.

Głównymi zaletami tego podejścia są:

- możliwość wyznaczania charakterystyk kolejek dla skończonych rozmiarów buforów – kluczowa z uwagi na skończone rozmiary buforów w urządzeniach sieciowych;
- możliwość wyznaczania charakterystyk zarówno w stanie ustalonym, jak i nieustalonym, co umożliwia pełniejszy opis pracy mechanizmu kolejowania;
- możliwość wyznaczania praktycznie wszystkich istotnych charakterystyk kolejowania;
- możliwość przedstawienia rozwiązań w zwartej postaci, która jest wygodna do implementacji i redukuje nakład obliczeniowy.

Wykorzystując to podejście, udało się otrzymać wiele całkowicie nowych wyników dla kolejek z markowskimi modelami ruchu, w tym rozkłady długości kolejki, rozkłady opóźnienia, prawdopodobieństwa przepełnienia bufora, współczynniki strat pakietów, czasy do przepełnienia bufora, długości okresu przepełnienia bufora czy statystyczną strukturę strat pakietów.

Przedyskutowane zostały też zagadnienia związane z praktycznym wykorzystaniem otrzymanych wyników teoretycznych do celów obliczeniowych, w tym omó-

wiono numeryczne odwracanie transformat Laplace'a i funkcji tworzących, pokazano metodę uniformizacji pomocną przy wyznaczaniu macierzy współczynników występujących w twierdzeniach, zbadano złożoność obliczeniową proponowanych rozwiązań, oraz pokazano, jak odpowiednio parametryzować markowskie modele ruchu.

Wreszcie otrzymane wyniki teoretyczne zostały zilustrowane licznymi przykładami obliczeniowymi, w których wykorzystano m. in. parametryzacje modeli oparte na śladach ruchu IP.

BIBLIOGRAFIA

1. Abate J., Choudhury G. L. and Whitt W. – On the Laguerre method for numerically inverting Laplace transforms. *INFORMS Journal on Computing* 8, pp. 413–427, (1996).
2. Abate J., Choudhury G. L. and Whitt W. – An introduction to numerical transform inversion and its application to probability models. Chapter in *Computational Probability*, pp. 257–323, W. Grassman (ed.), Kluwer, Boston, (2000).
3. Abbas M. and Ahmad Z. A. – Performance analysis of an ATM statistical multiplexer with batch arrivals. *IEE Proceedings: Communications*, 141, pp. 190–195, (1994).
4. El Abdouni Khayari R., Sadre R. and Haverkort B. R. – Fitting world-wide web request traces with the EM-algorithm. *Performance Evaluation* 52, pp. 175–191, (2003).
5. Adas A. – Traffic Models in Broadband Networks. *IEEE Communications Magazine* (7), pp. 82–89, (1997).
6. Akar N. and Sohraby K. – Matrix geometric solutions in M/G/1 type Markov chains with multiple boundaries: A generalized state-space approach. In *Teletraffic contributions for the information age*. Eds., V. Ramaswami and P. E. Wirth, pp. 487–496, (1997).
7. Akar N. and Sohraby K. – An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Stochastic Models*, 13, pp. 381–416, (1997).
8. Alfa A. S. and Chakravarthy S. – A discrete queue with Markovian arrival process and phase type primary and secondary services. *Stochastic Models*, 10, pp. 437–451, (1994).
9. Alfa A. S. and Neuts M. F. – Modelling vehicular traffic using the discrete time Markovian arrival process. *Transportation Science*, 29, pp. 109–117, (1995).
10. Alfa A. S. and Frigui I. – Discrete NT-policy single server queue with Markovian arrival process and phase type service. *European Journal of Operational Research*, 88, pp. 599–613, (1996).
11. Andersen A. T., Jensen A. and Nielsen B. F. – Modelling and performance study of packet traffic with self-similar characteristics over several time-scales with Markovian Arrival Processes. *VTT Symposium (Valtion Teknillinen Tutkimuskeskus)*, pp. 154–269, (1995).

12. Andersen A. T. and Nielsen B. F. – Application of superpositions of two state Markovian sources to the modelling of self-similar behaviour. Proceedings IEEE INFOCOM'97, pp. 196–204, (1997).
13. Andersen A. and Nielsen B. – A Markovian approach for modeling packet traffic with long-range dependence, IEEE Journal on Selected Areas in Communications 16(5), pp. 719-732, (1998).
14. Appenzeller G., Keslasy I. and McKeown N. – Sizing Router Buffers. Proc. ACM SIGCOMM'04, pp. 281–292, September, (2004).
15. Asmussen S. and O'Conneide C. A. – On the tail of the waiting time in a Markov-modulated M/G/1 queue. Operations Research, vol. 50, Issue 3, pp. 559–565, (2002).
16. Asmussen S. and Moller J. R. – Calculation of steady state waiting time distribution in GI/PH/c and MAP/PH/c queues. Technical report, Centre for Mathematical Sciences, Lund University, Sweden, (1999).
17. Asmussen S., Jobmann M. and Schwefel H. P. – Exact Buffer Overflow Calculations for Queues via Martingales. Queueing Systems 42(1): pp. 63–90, (2002).
18. Baiocchi A., Melazzi N., Listanti M., Roveri A. and Winkler R. – Loss performance analysis of an ATM multiplexer loaded with high-speed ON-OFF sources. IEEE Journal on Selected Areas in Communications, 9, pp. 388–393, (1991).
19. Baiocchi A. and Blefari-Melazzi N. – Steady-state analysis of the MMPP/G/1/K queue. IEEE Trans. Commun. 41, No.4, pp. 531–534, (1992).
20. Balay R. I. and Nilsson A. A. – Performance modeling of an ATM multiplexer with heterogeneous speed links. IEEE International Conference on Communications, 2, pp. 778–782, (1995).
21. Barman D., Smaragdakis G. and Matta I. – The effect of router buffer size on high-speed TCP performance. Proceedings of IEEE GLOBECOM'04, pp. 1617–1621, (2004).
22. Bean N., Green D. and Taylor P. – Approximations to the output process of MAP/PH/1 queues. In Advances in Matrix-analytic Methods in Stochastic Models, Ed., Alfa A. S. and Chakravarthy S. R. Notable Publications, pp. 151–169, (1998).
23. Bin L. and Chakravarthy S. – Performance analysis of a resequencing model in telecommunications network. Fifth International Conference on Telecommunication Systems, pp. 633–652, (1997).
24. Bini D. and Meini B. – On the solution of nonlinear matrix equation arising in queueing problems. SIAM J. Matrix analysis and applications, 17, pp. 906–926, (1996).
25. Bini D. and Meini B. – On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems. In Computations with Markov chains. Ed., W.J. Stewart, Kluwer Academic publishers, pp. 21–38, (1996).

26. Bini D. and Meini B. – Improved cyclic reduction for solving queueing problems. *Numerical Algorithms*, 15, pp. 57–74, (1997).
27. Bini D. and Meini B. – Inverting block Toeplitz matrices in block Hessenberg form by means of displacement operators: application to queueing problems. *Linear Algebra and Applications*, 272, pp. 1–16, (1998).
28. Blondia C. – A Discrete-Time Batch Markovian Arrival Process as B-ISDN Traffic Model. *Belgian Journal of Operations Research, Statistics and Computer Science*, 32, pp. 3–23, (1992).
29. Blondia C. and Casals O. – Performance analysis of statistical multiplexing of VBR sources. *Proceedings-IEEE INFOCOM*, 2, pp. 828–838, (1992).
30. Blondia C. – The finite Capacity N/G/1 Queue. *Communications in Statistics: Stochastic Models*, 5(2), pp. 273–294, (1989).
31. de Boer P. T., Nicola V. F. and van Ommeren J. C. W. – The Remaining Service Time Upon Reaching a High Level in M/G/1 Queues. *Queueing Systems* 39, pp. 55-78, (2001).
32. Bratijczuk M., Chydzinski A. – *Rachunek prawdopodobieństwa*. Wydawnictwo Politechniki Śląskiej, Gliwice, (2000).
33. Bratiychuk M. and Chydzinski A. – On the ergodic distribution of oscillating queueing systems. *Journal of Applied Mathematics and Stochastic Analysis* 16:3, pp. 311–326, (2003).
34. Breuer L. – An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, vol. 112, pp. 123–138, (2002).
35. Buchholz P. – An EM-Algorithm for MAP Fitting from Real Traffic Data. In: P. Kemper, W. H. Sanders (eds.), *Computer Performance Evaluation Modelling Techniques and Tools*, Springer LNCS 2794, pp. 218–236, (2003).
36. Carle G. – Adaptable error control for efficient provision of reliable services in ATM networks. *Proc. 1st Workshop ATM Traffic Management, IFIP, WG.6.2 Broadband Communication*, Paris, pp. 399-406, (1995).
37. Carmo R. M. L. R., de Souza e Silva E. and Marie R. – Efficient solutions for an approximation technique for the transient analysis of Markovian models. Technical report, IRISA Publication Interne. N 1067, (1996).
38. Chakravarthy S. R. – Analysis of a multi-server queue with batch Markovian arrivals and group services. *Engineering Simulation*, 18, pp. 51–66, (2000).
39. Chakravarthy S. and Dudin A. N. – A Multi-Server Retrial Queue with BMAP Arrivals and Group Services. *Queueing Systems*, volume 42, Number 1, pp. 5–31, (2002),

40. Chakravarthy S. and Dimitrova T. – Service control of queueing system with nonrenewal arrivals and phase type services. In *Mathematics with vision, Proceedings of the First International Mathematica Symposium*. Eds. V. Keranen, V and P. Mitic, Computational Mechanics Publication, pp. 93–102, (1995).
41. Chakravarthy S. – Analysis of the MAP/PH/1/K queue with service control. *Applied Stochastic Models and Data Analysis* 12, pp. 179–181, (1996).
42. Chakravarthy S., Chukova S. and Dimitrov B. – Analysis of MAP/M/2/K queueing model with infinite resequencing buffer. *Performance Evaluation*, 31, pp. 211–228, (1998).
43. Chakravarthy S. – The Batch Markovian Arrival Process: A Review and Future Work. *Advances in Probability and Stochastic Processes*. Eds., A. Krishnamoorthy, et al., pp. 21–49, (2001).
44. Choudhury G. L. and Lucantoni D. M. – Numerical computation of the moments of a probability distribution from its transform. *Operations Research*, 44, pp. 368–381, (1996).
45. Choudhury G. L., Lucantoni D. M. and Whitt W. – Multi-Dimensional Transform Inversion with Applications to the Transient M/G/1 Queue. *Annals of Applied Probability* vol. 4, pp. 719–740, (1994).
46. Chydzinski A. – The oscillating queue with input stream intensity switching. *Applied Stochastic Models and Data Analysis* 10, pp. 327–331, (2001).
47. Chydzinski A. – The M/G-G/1 oscillating queueing system. *Queueing Systems*, vol. 42:3, pp. 255–268, (2002).
48. Chydzinski A. – The M-M/G/1-type oscillating systems. *Cybernetics and Systems Analysis*, vol. 39(2), pp. 316–324, (2003).
49. Chydzinski A. – Diffusion approximation in analysis of ATM traffic control mechanisms. *Archiwum Informatyki Teoretycznej i Stosowanej*, vol. 15(3), pp. 207–222, (2003).
50. Chydzinski A. – The oscillating queue with finite buffer. *Performance Evaluation*, vol. 57(3), pp. 341–355, (2004).
51. Chydzinski A. – On transient analysis of queues with Poisson input stream. *Proceedings of International Workshop in Applied Probability*, Piraeus, Greece, pp. 85–89, (2004).
52. Chydzinski A. – On the remaining service time upon reaching a target level in M/G/1 queues. *Queueing Systems*, vol. 47, issue 1/2, pp. 71–80, (2004).
53. Chydzinski A. – On the Transient Remaining Service Time in a Large Buffer Queue. *WSEAS Transactions on Communications*, vol. 3, issue 4, pp. 958–962, (2004).
54. Chydzinski A. – On the Statistical Structure of Losses Caused by the Buffer Overflow. *Proceedings of TELE-INFO'05*, pp. 165–171, Prague, March (2005).

55. Chydzinski A. – On the Distribution of Consecutive Losses in a Finite Capacity Queue. WSEAS Transactions on Circuits and Systems, vol. 4, issue 3, pp. 117–124, (2005).
56. Chydzinski A. and Winiarczyk R. – Distribution of the First Buffer Overflow Time in a Deterministic Service Time Queue. Proceedings Of IEEE Symposium on Computers and Communications. La Manga, Spain, June (2005).
57. Chydzinski A. – Transient Analysis of the MMPP/G/1/K Queue. Telecommunication Systems, vol. 32, n. 4, pp. 247–262, (2006).
58. Chydzinski A. – Duration of the buffer overflow period in a batch arrival queue. Performance Evaluation, vol. 63:4-5, pp. 493–508, (2006).
59. Chydzinski A. – On the Overflow Interval in a Poisson Arrival Queue. WSEAS Transactions on Computers, vol. 5, issue 7, pp. 1421–1428, (2006).
60. Chydzinski A. and Winiarczyk R. – Blocking probability in a BMAP queue. Proceedings of 11-th IEEE Symposium on Computers and Communications, pp. 547–552. Pula, Italy, July (2006). (The best paper award in modeling and simulation).
61. Chydzinski A. – Queue Size in a BMAP Queue with Finite Buffer. Proceedings of Next Generation Teletraffic and Wired/Wireless Advanced Networking '06. Lecture Notes in Computer Science, vol. 4003, pp. 200–210, (2006).
62. Chydzinski A. – Delay analysis in a BMAP queue. Proceedings of 13-th International Conference on Telecommunications. Funchal, Portugal, May (2006).
63. Chydzinski A. – Buffer Overflow Period in a Constant Service Rate Queue. Proceedings of TELE-INFO '06, pp. 186–191, Istanbul, May (2006).
64. Chydzinski A. – On the Queue Length Distribution in BMAP Systems. Journal of Communications Software and Systems, vol. 2, no. 4, pp. 275–282, (2006).
65. Chydzinski A. – Buffer overflow calculations in a batch arrival queue. ACM SIGMETRICS Performance Evaluation Review, vol. 34, issue 2, pp. 19–21, (2006).
66. Chydzinski A., Wojcicki R. and Hryn G. – On the Number of Losses in an MMPP Queue. Proceedings of Next Generation Teletraffic and Wired/Wireless Advanced Networking '07. Lecture Notes in Computer Science, vol. 4712, pp. 38–48, (2007).
67. Chydzinski A. – Time to reach buffer capacity in a BMAP queue. Stochastic Models, vol. 23, pp. 195–209, (2007).
68. Chydzinski A. – Buffer Overflow Period in an MAP Queue. Mathematical Problems in Engineering, vol. 2007, pp. 1–18, (2007).
69. Chydzinski A. – Time to Buffer Overflow in an MMPP Queue. Proceedings of Networking '07, Atlanta, May 2007. Lecture Notes in Computer Science, vol. 4479, pp. 879–889, (2007).
70. Chydzinski A. and Winiarczyk R. – On the Blocking Probability in Batch Markovian Arrival Queues. Microprocessors and Microsystems, przyjęto do druku, (2007).

71. Chydzinski A. – Solving Finite-Buffer Queues with Markovian Arrivals. ACM SIGMETRICS Performance Evaluation Review, przyjęto do druku, (2007).
72. Chydzinski A. – A new approach to analyzing Markovian queues. Zgłoszono do publikacji, (2007).
73. Chydzinski A. – Packet loss process in a queue with Markovian arrivals. Zgłoszono do publikacji, (2007).
74. Cohen J. W. – *The Single Server Queue*. Revised edition North-Holland Publishing Company, Amsterdam, (1982).
75. Conolly B. – *Lecture Notes on Queueing Systems*. Ellis Horwood Limited, Susses, England, (1975).
76. Cooper R. – *Introduction to Queueing Theory*. 2nd Edition North Holland, (1981).
77. Crovella M. and Bestavros A. – Self-similarity in World Wide Web traffic: Evidence and possible causes. IEEE/ACM Transactions on Networking 5(6): pp. 835-846, (1997).
78. Czachórski T. – *Modele kolejkowe w ocenie efektywności sieci i systemów komputerowych*. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice, (1999).
79. Davies B. and Martin B. L. – Numerical inversion of the Laplace transform: A survey and comparison of methods. J. Comp. Phys. 33, pp. 1–32, (1979).
80. Dempster A., Laird N. and Rubin D. – Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B, Volume: 39, pp. 1–38, (1997)
81. Deng L. and Mark J. – Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization, Telecommunication Systems 1, pp. 321-338, (1993).
82. Dhamdhere A. and Dovrolis C. – Open issues in router buffer sizing. ACM SIGCOMM Computer Communication Review 36, pp. 87–92, (2006).
83. Dhamdhere A., Jiang H. and Dovrolis C. – Buffer Sizing for Congested Internet Links. Proc. of IEEE INFOCOM, Miami FL, pp. 1072–1083, March, (2005).
84. Dudin A. N. – Optimal multi-threshold control for a BMAP/G/1 queue with N service modes. Queueing Systems: 30, pp. 273–287, (1998).
85. Dudin A. N. and Nishimura S. – Optimal control for a BMAP/G/1 queue with two service modes. Mathematical Problems in Engineering, 5, pp. 255–273, (1999).
86. Enachescu M., Ganjali Y., Goel A., McKeown N. and Roughgarden T. – Part III: routers with very small buffers. ACM SIGCOMM Computer Communication Review 35, pp. 83–90, (2005).
87. Erramilli A., Singh R. P. and Pruthi P. – An application of deterministic chaotic maps to model packet traffic. Queueing Systems, vol. 20, pp. 171-206, (1995).

88. Fakinos D. – The expected remaining service time in a single server queue. *Operations Research*, 30, pp. 1014–1018, (1982).
89. Fiems D., Steyaert B. and Bruneel H. – Performance of a Partially Shared Buffer with Correlated Arrivals. *Lecture Notes in Computer Science* 4479, pp. 1196–1200, (2007).
90. Filipowicz B. – *Analiza i synteza systemów obsługi i sieci kolejkowych. Modele stochastyczne w badaniach operacyjnych*. Wydawnictwa Naukowo-Techniczne, Warszawa, (1996).
91. Fischer W. and Meier-Hellstern K. – The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation* 18, No.2, 149–171, (1992).
92. Fonseca N. L. S. and Silvester J. A. – Modelling the output process of an ATM multiplexer with correlated priorities. *IEEE International Conference on Communications*, 2, pp. 816–821, (1997).
93. <ftp://pma.nlanr.net/traces/>
94. Gail H. R., Hantler L. and Taylor B. A. – Solutions of the basic matrix equation for M/G/1 and G/M/1 type Markov chains. *Stochastic Models*, 10, pp. 1–43, (1994).
95. Gail H. R., Hantler L. and Taylor B. A. – Spectral analysis of M/G/1 and G/M/1 type Markov chains. *Adv. Appl. Prob.*, 28, pp. 114–165, (1996).
96. Garcia J. and Casals O. – A discrete time queueing model to study the cell delay variation in an ATM queueing network. *Performance Evaluation*, 21, pp. 3–22, (1994).
97. Gelenbe E., Mang X. and Önvural R. – Diffusion based statistical call admission control in ATM. *Performance Evaluation*, 27&28, pp. 411–436, (1996).
98. Ge H., Harder U. and Harrison P. G. – Parameter estimation for MMPPs using the ME algorithm. *Proc. UKPEW*, July (2003).
99. Gorinsky S., Kantawala A. and Turner J. – Link Buffer Sizing: A New Look at the Old Problem, *Proc. Of 10-th IEEE Symposium on Computers and Communications*. La Manga, Spain, June, (2005).
100. Gouweleuw F. N. – Calculating the loss probability in a BMAP/G/1/N+1 queue. *Commun. Stat., Stochastic Models* 12, No.3, pp. 473–492, (1996).
101. Grassman W. K. and Heyman D. P. – Equilibrium distribution of block-structured Markov chains with repeating rows. *J. Appl. Prob.*, 27, pp. 557–576, (1993).
102. Grassman W. K. and Heyman D. P. – Computation of steady-state probabilities for infinite state Markov chains with repeating rows. *ORSA J. on Computing*, 5, pp. 292–303, (1993).
103. Gnrenfelder R. and Robert S. – Which arrival law parameters are decisive for queueing system performance. *Proc. of ITC* 14, (1994).

104. He Q. and Jewkes E. M. – A level crossing analysis of the MAP/G/1 queue. In *Matrix-analytic Methods in Stochastic Models*, Eds., S. R. Chakravathy and A. S. Alfa, Marcel Dekker, pp. 107–116, (1996).
105. Heffes H. and Lucantoni D. M. – A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communication*, special issue on Network Performance Evaluation, SAC-4, 6, pp. 856–868, (1986).
106. Herrmann C. – Correlation effect on per-stream QOS parameters of ATM traffic superpositions relevant to Connection Admission Control. *IEEE International Conference on Communications*, pp. 1027–1031, (1993).
107. Hofkens T., Spaey K. and Blondia C. – Transient Analysis of the D-BMAP/G/1 Queue with an Application to the Dimensioning of a Playout Buffer for VBR Video. *Proc. of Networking'04. Lecture Notes in Computer Science*, Vol. 3042, pp. 1338–1343, (2004).
108. Hryn G., Jerzak Z. and Chydzinski A. – MMPP-based HTTP traffic generation with multiple emulated sources, *Archiwum Informatyki Teoretycznej i Stosowanej*, 16(4), pp. 321–335, (2004).
109. Iyer S., Kompella R. R. and McKeown N. – Analysis of a memory architecture for fast packet buffers. In *Proc. of IEEE High Performance Switching and Routing*, Dallas, Texas, May, (2001).
110. Jean-Marie A., Calasb Y. and Alemua T. – On the compromise between burstiness and frequency of events. *Performance Evaluation*, Vol. 62, Issues 1–4, pp. 382–399, (2005).
111. Jedrus S. – The design of multiple correlated multifractal traffic models. *Archiwum Informatyki Teoretycznej i Stosowanej*, vol. 12, No. 4, (2000).
112. Kang S. and Sung D. – Two-state MMPP modelling of ATM superposed traffic streams based on the characterisation of correlated interarrival times, in: *Proc. of IEEE GLOBECOM '95*, pp. 1422-1426, (1995).
113. Kang S. H., Kim Y. H., Sung D. K. and Choi B. D. – An application of Markovian arrival process (MAP) to modeling superposed ATM cell streams. *IEEE Transactions on Communications* 50(4): pp. 633–642, (2002).
114. Kang K. and Kim C. – Performance analysis of statistical multiplexing of heterogeneous discrete-time Markovian arrival processes in an ATM network. *Computer Communications*, 20, pp. 970–978, (1997).
115. Kendall D. G. – Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chain. *American Mathematical Statistics* 24, pp. 338–354, (1953).

116. Kim Y. H. and Un C. K. – Performance analysis of statistical multiplexing for heterogeneous bursty traffic in ATM network. *IEEE Trans. Commun.* 42(2-4) pp. 745–753, (1994).
117. Kleinrock L. – *Queueing Systems: Theory*. John Wiley and Sons New York, (1975).
118. Klemm A., Lindemann C. and Lohmann M. – Traffic Modeling of IP Networks Using the Batch Markovian Arrival Process. *Proc. International Conference on Computer Performance Evaluation, Modelling Techniques and Tools, Lecture Notes In Computer Science 2324*, pp. 92–110, (2002).
119. Klemm A., Lindemann C. and Lohmann M. – Traffic Modeling and Characterization for UMTS Networks, *Proc. GLOBECOM 2001*, pp. 1741–1746, San Antonio TX, November, (2001).
120. Klemm A., Lindemann C. and Lohmann M. – Modeling IP traffic using the batch Markovian arrival process. *Performance Evaluation*, Vol. 54, Issue 2, (2003).
121. Kofman D. and Korezlioglu H. – Loss probabilities and delay and jitter distributions in a finite buffer queue with heterogeneous batch Markovian arrival processes. *IEEE Global Telecommunications Conference*, 2, pp. 830–834, (1993).
122. Korolyuk S. – *Granicznye zadaczi dla sloznych puassonowskich processow*. Naukowa Dumka, Kijów, (1975).
123. Koutsonikolas D. and Hu Y. C. – The Case for FEC-Based Reliable Multicast in Wireless Mesh Networks. *Proc. of the 37th IEEE/IFIP International Conference on Dependable Systems and Networks*, pp. 491–501, (2007).
124. Kulkarni L. and Li S. Q. – Transient behaviour of queueing systems with correlated traffic. *Perform. Eval.* 27&28, pp. 117–145, (1996).
125. Landman J. and Kritzing P. – Delay analysis of downlink IP traffic on UMTS mobile networks. *Performance Evaluation* 62, pp. 68–82, (2005).
126. Latouche G. and Ramaswami V. – A logarithmic reduction algorithm for quasi-birth-death processes. *J. Appl. Prob.*, 30, pp. 650–674, (1993).
127. Latouche G. – Algorithms for infinite Markov chains with repeating columns. In *Linear algebra, queueing models and Markov chains*. Eds., Meyer, C. D. and R. J. Plemmons, pp. 231–265, (1993).
128. Latouche G. and Stewart G. W. – Numerical methods for M/G/1 type queues. In *Computations with Markov chains*. Ed., W. J. Stewart, Kluwer Academic publishers, pp. 571–581, (1996).
129. Lee H. W., Park N. I. and Jeon J. – A new approach to the queue length and waiting time of BMAP/G/1 queues. *Computers & Operations Research*, Vol. 30, pp. 2021–2045, (2003).

130. Lee H. W., Lee S. W. and Jeon J. – Using Factorization in Analyzing D-BMAP/G/1 Queues. *Journal of Applied Mathematics and Stochastic Analysis*, Vol. 2005, No. 2, pp. 119–132, (2005).
131. Lee H. W. and Baek J. W. – BMAP/G/1 queue under D-policy: queue length analysis. *Stochastic Models*, Vol. 21, pp. 484–505, (2005).
132. Lee H. W., Moon J. M., Kim B. K., Park J. G. and Lee S. W. – A simple eigenvalue method for low-order D-BMAP/G/1 queues. *Applied Mathematical Modelling*, Vol. 29, pp. 277–288, (2005).
133. Lee D. S. and Li S. Q. – Transient analysis of multi-server queues with Markov-modulated Poisson arrivals and overload control. *Perform. Eval.* 16, No.1-3, pp. 49–66, (1992).
134. Le Ny L. M. and Sericola B. – Transient Analysis of the BMAP/PH/1 Queue. *International Journal of Simulation: Systems, Science & Technology. Special Issue on Analytical & Stochastic Modeling Techniques*, 3(3-4), (2002).
135. Leland W., Taqqu M., Willinger W. and Wilson D. – On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking* 2(1): pp. 1–15, (1994).
136. Li S. and Hwang C. – On the convergence of traffic measurement and queuing analysis: A statistical match and queuing (SMAQ) tool, *IEEE/ACM Transactions on Networking*, 95-110, (1997).
137. Lucantoni D. M. – New results on the single server queue with a batch Markovian arrival process. *Commun. Stat., Stochastic Models* 7, No.1, pp. 1–46, (1991).
138. Lucantoni D. M., Choudhury G. L. and Whitt W. – The transient BMAP/G/1 queue. *Commun. Stat., Stochastic Models* 10, No. 1, pp. 145-182, (1994).
139. Lucantoni D. – Further transient analysis of the BMAP/G/1 queue. *Commun. Stat., Stochastic Models* 14, No. 1-2, pp. 461–478, (1998).
140. Lucantoni D. M., Meier-Hellstern K. S. and Neuts M. F. – A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability* 22, pp. 676–705, (1990).
141. Lucantoni D. M. and Neuts M. F. – Simpler proofs of some properties of the fundamental period of the MAP/G/1 queue. *J. Appl. Prob.*, 31, pp. 235–243, (1994).
142. Machihara F. – Bridge between preemptive and non-preemptive queueing models. *Performance Evaluation*, 23, pp. 93–106, (1995).
143. Machihara F. and Hatoyama-machi H. – A preemptive priority queue with nonrenewal inputs. In *Matrix-analytic Methods in Stochastic Models*, Eds., S. R. Chakravarthi and A. S. Alfa, Marcel Dekker, pp. 21–42, (1996).
144. Meier-Hellstern K. – A fitting algorithm for Markov-modulated Poisson process having two arrival rates, *European Journal of Operational Research* 29, (1987).

145. Meini B. – New convergence results on functional iteration techniques for the numerical solution of M/G/1 type Markov chains. *Numer. Math.*, 78, pp. 39–58, (1997).
146. Meini B. – An improved FFT-based version of Ramaswamis formula. *Stochastic Models*, 13, pp. 223–238, (1997).
147. Meini B. – Solving M/G/1 type Markov chains: Recent advances and applications. *Stochastic Models*, 14, pp. 479–496, (1998).
148. Murthy G. R. Kim M. and Coyle E. J. – Equilibrium analysis of skip free Markov chains: nonlinear matrix equations. *Stochastic Models*, 7, pp. 547–571, (1991).
149. Naoumov V., Krieger U. R. and Wagner D. – Analysis of a multiserver delay-loss system with a general Markovian arrival process. In *Matrix-analytic Methods in Stochastic Models*, Eds., S. R. Chakravarthy and A. S. Alfa, Marcel Dekker, pp. 43–66, (1996).
150. Neuts M. F. – Renewal processes of phase type. *Naval Res. Logist. Quart.* 25, pp. 445–454, (1978).
151. Neuts M. F. – A Versatile Markovian Point Process *Journal of Applied Probability* 14, pp. 764–779, (1979).
152. Neuts M. F. – *Structured Stochastic Matrices of M/G/1-type and their Applications*, Marcel Dekker, New York, NY, (1989).
153. Neuts M. F. – The fundamental period of the queue with Markov-modulated arrivals. *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin*, Academic Press, pp. 187–200, (1989).
154. Niu Z., Takahashi Y. and Endo N. – Performance evaluation of SVC-based IP-over-ATM networks. *Transactions on Communications*, E81-B, pp. 948–957, (1998).
155. <http://pma.nlanr.net/>
156. Nogueira A., Salvador P., Valadas R. and Pacheco A. – Modeling self-similar traffic through Markov modulated Poisson processes over multiple time scales. 6-th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC'03), *Lecture Notes in Computer Science* 2720, pp. 550–560, July, (2003).
157. Nonnenmacher J., Biersack E. and Towsley D. – Parity-based loss recovery for reliable multicast transmission. *Proc. SIGCOMM 97*, Cannes, France, pp. 289–299, (1997).
158. Norros I. – A storage model with self-similar input. *Queueing Systems* vol. 16, pp. 387–396, (1994).
159. Nunes C. and Pacheco A. – Parametric estimation in MMPP(2) using time discretization. *Proc. of the 2nd Internat. Symposium on Semi-Markov Models: Theory and Applications*, December, (1998).
160. <http://www.omnetpp.org/>

161. Osowski J. – *Zarys rachunku operatorowego*. Wydawnictwa Naukowo Techniczne, Warszawa, (1981).
162. Papir Z. – *Ruch telekomunikacyjny i przeciążenia sieci pakietowych*. Wydawnictwa Komunikacji i Łączności, Warszawa, (2001).
163. Paxson V., Floyd S. – Wide-area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Netw.*, Volume 3, Issue 3, pp. 226–244, (1995).
164. Podolsky M., Romer C. and McCanne S. – Simulation of FEC-Based Error Control for Packet Audio on the Internet, *Proc. of INFOCOM'98*, pp. 505–515, (1998).
165. Puliafito A., Balakrishnan M., Trivedi K. S. and Viniotis I. – Buffer sizing for ABR traffic in an ATM switch. *IEEE International Conference on Communications*, 1, pp. 316–320, (1995).
166. Raina G. and Wischik D. – Buffer sizes for large multiplexers: TCP queueing theory and instability analysis. In *EuroNGI*, (2005).
167. Ramaswami V. – The N/G/1 queue and its detailed analysis. *Adv. Appl. Prob.* 12, pp. 222–261, (1980).
168. Ramaswami V. – Nonlinear matrix equations in applied probability – Solution techniques and open problems. *SIAM Review*, 30, pp. 256–263, (1988).
169. Ramaswami V. and Wang J. L. – Discrete time queueing analysis of ATM systems with phasetype distributed talk spurts traffic sources. *IEEE International Conference on Communications*, 2, pp. 1093–1098, (1996).
170. Ramaswami V. and Wang J. L. – Discrete time queueing analysis of ATM systems with heterogeneous traffic sources. *IEEE Global Telecommunications Conference*, 1, pp. 623–628, (1996).
171. Ramaswami V. – A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Stochastic Models*, 4, pp. 183–188, (1988).
172. Ramaswamy V., Cuellar L., Eidenbenz S., Hengartner N., Ambuhl C. and Weber B. – Light-Weight Control of Non-responsive Traffic with Low Buffer Requirements. *Lecture Notes in Computer Science 4479*, pp. 856–877, (2007).
173. Rananand N. – Approximating a variable bit rate source by Markov processes. *IEEE Global Telecommunications Conference*, 2, pp. 1107–1112, (1994).
174. Rasti S. – Moments of the queue occupancy in an ATM multiplexer loaded with on/off sources. *ICCS-Conference Proceedings*, pp. 754–759, (1994).
175. Riedi R., Crouse M., Ribeiro V. and Baraniuk R. – A multifractal wavelet model with application to network traffic. In *IEEE Trans. on Information Theory*, 45(3), pp. 992–1018, (1999).
176. Riska A. and Smirni E. – Exact aggregate solutions for M/G/1-type Markov processes. *Proc. of ACM SIGMETRICS Conf.*, Marina del Rey, CA, pp. 86-96, (2002).

177. Rizzo L. and Visicano L. – RMDP: an FEC-based reliable multicast protocol for wireless environments. *Mobile Computing and Communications Review*, vol. 2, no. 2, (1998).
178. Robert S. and Le Boudec J. – New models for self-similar traffic. *Performance Evaluation* 30(1/2), (1997).
179. Ryden T. – An EM algorithm for parameter estimation in Markov modulated Poisson processes, *Comput. Stat. Data Anal.* 21, pp. 431-447, (1996).
180. Saito H. – Performance evaluation and dimensioning for AAL2 CLAD. *Proceedings-IEEE INFOCOM*, 1, pp. 153–160, (1999).
181. Salvador P., Valadas R. and Pacheco A. – Multiscale Fitting Procedure Using Markov Modulated Poisson Processes. *Telecommunication Systems* 23:1-2, pp. 123–148, (2003).
182. Salvador P., Pacheco A. and Valadas R. – Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. *Computer Networks* 44, pp. 335-352, (2004).
183. Salvador P., Nogueira A., Valadas R. and Pacheco A. – Multi-time-Scale Traffic Modeling Using Markovian and L-Systems Models. *Proc. ECUMN'04, Lecture Notes on Computer Science* 3262, pp. 297–306, (2004).
184. Sashisekaran T. and Chakravarthy S. – A stochastic model for a hysteresis based priority queueing strategy for ATM networks with batch arrivals – theory. *Proceedings of the third IEEE Symposium on Computers and Communications*, pp. 530–536, (1998).
185. Shah-Heydari S. and Le-Ngoc T. – MMPP models for multimedia traffic. *Telecommunication Systems* 15, No. 3-4, pp. 273–293, (2000).
186. Shu Y., Jin Z., Wang J. and Yang O. W. – Prediction-based admission control using FARIMA models. *Proc. IEEE ICC'00*, vol. 3, New Orleans, pp. 1325–1329, June, (2000).
187. Singh L. N. and Dattatreya G. R. – A novel approach to parameter estimation in Markov-modulated Poisson processes, *2004 IEEE Emerging Technologies Conference (ETC)*, Richardson, TX, Oct., (2004).
188. Skelly P., Schwartz M. and Dixit S. – A histogram-based model for video traffic behavior in an ATM multiplexer. *IEEE/ACM Trans. Netw.* 1(4), pp. 446–459, (1993).
189. Sohraby K. and Sidi M. – On the performance of bursty and modulated sources subject to leaky bucket rate-based access control schemes. *IEEE Transactions on Communications*, 42, pp. 477–487, (1994).
190. Sridhar R., Rosenberg C. and Kumar A. – Revenue maximisation in ATM networks using the CLP capability and buffer priority management. *IEEE Global Telecommunications Conference*, 1, pp. 734–738, (1995).

191. Subramanian V. G. and Srikant R. – Statistical multiplexing with priorities: Tail probabilities of queue lengths and waiting times. *Proceedings of the IEEE Conference on Decision and Control*, 2, pp. 1780–1785, (1997).
192. Takagi H. – *Queueing analysis - Vacation and Priority System*. North-Holland, Amsterdam, (1991).
193. Takagi H. – *Queueing analysis - Finite Systems*. North-Holland Amsterdam, (1993).
194. Takine T., Matsumoto Y., Suda T. and Hasegawa T. – Mean waiting times in nonpreemptive priority queues with Markovian arrival and i.i.d. service processes, *Performance Evaluation*, 20, pp. 131–149, (1994).
195. Takine T., Sengupta B. and Hasegawa T. – Conformance measure for traffic shaping in high speed networks with an application to the leaky bucket. *INFOCOM*, 2, pp. 474–481, (1994).
196. Trivedi K. – *Probability and Statistics with Reliability Queuing and Computer Science Applications (2nd edition)*. John Wiley and Sons, New York, (2001).
197. Tsakiridou S. and Stavrakakis I. – Mean delay analysis of a statistical multiplexer with batch arrival processes – a generalization to Viterbis formula. *Performance Evaluation* 25, pp. 1–15, (1996).
198. Van Houdt B. and Blondia C. – Approximated transient queue length and waiting time distributions via steady state analysis. *Stochastic Models*, 21:2/3, pp. 725-744, (2005).
199. Villamizar C. and Song C. – High Performance TCP in the ANSNET. *ACM SIGCOMM Computer Communication Review*, 24(5), pp. 45-60, November, (1994).
200. Wang S. S. and Silvester J. A. – Discrete-time performance model for integrated service ATM multiplexers. *IEEE Global Telecommunications Conference*, 2, pp. 757–761, (1993).
201. Wang S. S. and Silvester J. A. – Estimate the loss and tail probabilities for multimedia communication systems. *IEEE Global Telecommunications Conference*, 2, pp. 908–912, (1994).
202. Wei W., Wang B. and Towsley D. – Continuous-time hidden Markov models for network performance evaluation. *Performance Evaluation* 49, pp. 129–146, (2002).
203. Willinger W., Taqqu M. S., Sherman R. and Wilson D. V. – Self-Similarity through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level. *Proc. ACM SIGCOMM95*, pp. 100--113, (1995).
204. Wischik D. and McKeown N. – Part I: Buffer Sizes for Core Routers. *ACM SIGCOMM Computer Communication Review* 35, pp. 75–78, (2005).
205. Wong T. C., Mark J. W. and Chua K. C. – Delay performance of voice and MMPP video traffic in a cellular wireless ATM network. *IEE Proceedings Communications*, Vol. 148, 302–309, (2001).

206. Wójcicki R., Hryń G., Chydziański A. – Transmisja danych protokołów TCP/IP w sieciach optycznych. Rozdział w: Sieci komputerowe: aplikacje i zastosowania. Praca zbior. pod redakcją A. Kwietnia, J. Obera, B. Pochopienia i P. Gaja. Wydawnictwa Komunikacji i Łączności, Warszawa (2007).
207. Wu G. L. and Mark J. W. – Computational Methods for Performance Evaluation of a Statistical Multiplexer Supporting Bursty Traffic. IEEE/ACM Transactions on Networking, Vol. 4, No. 3, pp. 386–397, (1996).
208. Ye J. and Li San-qi. – Analysis of multi-media traffic queues with finite buffer and overload control: Algorithm. Proceedings-IEEE INFOCOM, 3, pp. 1464–1474, (1991).
209. Yoshihara T., Kasahara S. and Takahashi Y. – Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process, Telecommunication Systems 17(1/2), pp. 185–211, (2001).

CHARAKTERYSTYKI KOLEJKOWANIA MARKOWSKICH MODELI RUCHU W SIECIACH PAKIETOWYCH

Streszczenie

W monografii tej zaprezentowano zbiór rezultatów analitycznych, umożliwiających obliczanie charakterystyk kolejek pakietów występujących w węzłach sieci pakietowych. Rezultaty te obejmują nie tylko podstawowe charakterystyki kolejkowania, takie jak rozkłady długości kolejki czy rozkłady opóźnienia, ale również parametry dokładnie opisujące procesy przepełniania się bufora i strat pakietów, w tym prawdopodobieństwo przepełnienia bufora, długość okresu przepełnienia bufora, czas do przepełnienia bufora, współczynnik strat pakietów, rozkład liczby pakietów traconych w okresie przepełnienia bufora.

Do modelowania strumieni pakietów użyto markowskich modeli ruchu o rosnącej (w kolejnych rozdziałach książki) złożoności. W szczególności, wykorzystano proces Poissona, złożony proces Poissona, proces Poissona z markowską modulacją (MMPP) oraz markowski proces zdarzeń o grupowej strukturze (BMAP).

Przedstawione wyniki uwzględniają ograniczone rozmiary buforów w urządzeniach sieciowych oraz złożoną statystycznie strukturę ruchu obserwowaną w sieciach pakietowych (szczególnie rezultaty dla MMPP i BMAP).

Książka zawiera wiele przykładów obliczeniowych. Wykorzystują one, w większości wypadków, parametryzacje markowskich modeli ruchu oparte na zarejestrowanych śladach ruchu IP.

Dodatkowo w monografii omówiono kilka narzędzi numerycznych ułatwiających praktyczne wykorzystanie przedstawionych wcześniej wyników analitycznych.

QUEUEING CHARACTERISTICS FOR MARKOVIAN TRAFFIC MODELS IN PACKET-ORIENTED NETWORKS

Abstract

This monograph presents the author's contribution to the field of performance evaluation of packet buffering processes in network nodes. In particular, a detailed characterization of finite-buffer queues fed by Markovian traffic models is presented by means of theorems and formulas. The Markovian traffic models were chosen due to their ability to mimic a very complex statistical behaviour of the traffic, including the self-similarity and long-range dependence.

For analytical purposes, a new powerful method that can be used for all Markovian traffic models is proposed. This method combines the Laplace transform technique with the use of special recurrent sequences to solve large systems of linear equations. The method has several important advantages.

Firstly, it enables an analysis of finite-buffer queueing models. This issue is crucial from the practical point of view as in all real devices (routers, switches) the buffering space is limited. On the other hand, most previous studies covered only infinite-buffer queues, less demanding in analysis. Secondly, it makes possible both transient and steady-state characterization of the queueing process. Thirdly, it allows for finding virtually all important performance characteristics and gives results in a closed, easy to use form.

The following characteristics are studied by means of this method: the queue size distribution, the queueing delay distribution, the blocking probability, the loss ratio, the buffer overflow period, the time to buffer overflow and the statistical structure of packet losses. For each characteristic a new formula describing its transient and steady-state behaviour is obtained.

The analytical results are illustrated by numerical examples, most of them obtained for traffic parameterizations based on IP trace files.

The monograph consists of five chapters organized in the following way.

In Chapter 1 the potential method is presented. This method permits finding closed-form solutions for large systems of linear equations in a special form. The systems of equations in this form appear frequently in the remaining part of the book and the potential method is used to solve them effectively.

Chapters 2, 3 and 4 are the main part of the monograph and they present the queueing characteristics for different Markovian traffic models, with an emphasis on growing complexity of the model in every next chapter. In particular, Chapter 2 is devoted to the simple and compound Poisson processes, Chapter 3 focuses on the Markov-modulated Poisson process (MMPP), while Chapter 4 presents results for the batch Markovian arrival process (BMAP). All these chapters are structured as follows. Firstly, the description and the properties of the traffic model are given. Secondly, the theorems presenting formulas for the queueing characteristics are proven. Thirdly, the numerical examples are presented. Each chapter ends with a bibliographical note.

Chapter 5 presents a set of mathematical and numerical tools that enable or make easier obtaining numerical results from formulas proven in the previous chapters. Namely, a set of algorithms for inverting the Laplace transforms and generating functions is shown, the uniformization method for computing coefficient matrices for MMPP and BMAP is presented, the expectation maximization algorithm for MMPP and BMAP parameter fitting is discussed and the continuous version of the total probability formula is recalled.

INDEX

- algorytm EM, 152
- autokorelacja BMAP, 111
- autokorelacja MMPP, 60
- autokorelacja procesu Poissona, 22
- autokowariancja BMAP, 110
- autokowariancja MMPP, 60

- BMAP, 108

- całkowitoliczbowe błędzenie
 - przypadkowe, 14
- charakterystyki kolejkowania, 9
- czas do przepełnienia bufora, 53, 100, 130
- czas obsługi, 20
- czas przepełnienia bufora, 32
- część E algorytmu EM, 148
- część M algorytmu EM, 150

- dalekosiężność, 9
- dopasowanie parametrów, 145
- dwuwymiarowy łańcuch Markowa, 109
- dyscyplina FCFS, 21
- dyscyplina FIFO, 21
- długość kolejki, 23, 67, 116

- estymacja parametrów, 145

- FEC, 31
- funkcja licząca BMAP, 111
- funkcja licząca MMPP, 60
- funkcja tworząca BMAP, 111
- funkcja tworząca MMPP, 61
- funkcja tworząca procesu Poissona, 22

- funkcja tworząca złożonego procesu Poissona, 44
- funkcjonały funkcji liczącej BMAP, 111
- funkcjonały funkcji liczącej MMPP, 61

- grupowa struktura, 108

- iloczyn Kroneckera, 63
- intensywność procesu Poissona, 21
- intensywność strumienia grup w BMAP, 110

- kolejkowanie ruchu pakietów, 9
- konstruktywna definicja BMAP, 109
- kontrola błędów, 137

- liczba kolejnych strat pakietów, 38, 97
- liczba strat pakietów, 74

- łańcuch modulujący BMAP, 109
- łańcuch modulujący MMPP, 57

- macierz intensywności BMAP, 109
- macierz intensywności MMPP, 57
- macierz korelacji MMPP, 60
- MAP, 112
- markowska struktura, 58, 112
- markowski proces zdarzeń o grupowej strukturze, 108
- metoda EM, 145
- metoda największej wiarygodności, 146
- metoda uniformizacji, 143
- MMPP, 57

- model systemu kolejkowego, 20
modelowanie ruchu IP, 109
modulowany markowsko proces Poissona, 57
moment zwykły, 60
- N-proces, 133
notacja Kendalla, 21
- obciążenie systemu, 41, 49
odwracanie dwuwymiarowej transformaty Laplace'a, 139
odwracanie funkcji tworzących, 141
odwracanie jednowymiarowej transformaty Laplace'a, 136
odwracanie transformaty mieszanej, 142
ogon rozkładu, 31, 85
okres bezczynności systemu, 50
okres przepełnienia bufora, 31, 46, 86
operacje zmiennoprzecinkowe, 106
opóźnienie, 28, 82, 122
- parametry kolejkowania ruchu, 9
parametryzacja BMAP, 110
parametryzacja MMPP, 57
pojedyncze stanowisko obsługi, 20
potencjał, 15
prawdopodobieństwo przepełnienia bufora, 27, 70, 74, 119
proces BMAP, 109
proces MMPP, 57
proces Poissona, 21
procesy odnowy, 9
przestrzeń stanów łańcucha, 57
przykładowe parametryzacje BMAP, 113
przykładowe parametryzacje MMPP, 63
- rozkład graniczny, 38, 52
rozkład rozmiaru pakietu, 108
- rozkład stacjonarny łańcucha modulującego, 58, 110
rozkład typu fazowego, 112
rozkład wykładniczy, 21
rozmiar pakietu, 115
- samopodobieństwo, 9, 58
skończony bufor, 20
spiętrzenia, 9, 64, 108
spłot ciągów, 44
spłot dystrybuant, 30
spłot funkcji, 25
stan nieustalony, 10, 27
stan ustalony, 10, 27
statystyczna struktura strat, 38, 97
suma Kroneckera, 63
superpozycja MMPP, 63
symulacja BMAP, 154
symulator OMNeT++, 11
system BMAP/G/1/b, 116
system $M^X/G/1/b$, 21, 46
system M/G/1/b, 21, 23, 67
system MMPP/G/1/b, 21
system z nieskończonym buforem, 10
- śląd ruchu, 65
śląd ruchu IP, 115
średnia intensywność BMAP, 110
średnia intensywność MMPP, 58
- teoria kolejek, 9
teoria potencjału, 14
transformata Laplace'a, 136
transformata Laplace'a-Stieltjesa, 20, 136
- wariancja BMAP, 110
wariancja MMPP, 60
wirtualny czas oczekiwania, 28
współczynnik strat, 28, 74
wyznaczanie oryginału, 136

- wzór całkowy Bromwicha, 137
- wzór Eulera, 138
- wzór na prawdopodobieństwo całkowite,
151
- własność braku pamięci, 22
- zbieżność do stanu ustalonego, 73
- zbieżność parametryzacji, 146
- zdarzenia parami wyłączające się, 151
- złożoność czasowa, 105
- złożoność mnożenia macierzy, 105
- złożoność obliczeniowa, 105, 133
- złożoność pamięciowa, 105
- złożony proces Poissona, 43

SPIS OZNACZEŃ

$\mathbf{0}$ – macierz kwadratowa z zerowymi wszystkimi elementami

$\mathbf{1}$ – wektor kolumnowy złożony z samych jedynek

b – pojemność systemu (rozmiar bufora + 1)

$f(s) = \int_0^\infty e^{-sx} dF(x)$ – transformata Laplace'a-Stieltjesa rozkładu czasu obsługi

$f_g(x)$ – gęstość rozkładu czasu obsługi

$h(t)$ – gęstość rozkładu okresu przepełnienia bufora dla dużych buforów

$h_n(t)$ – gęstość rozkładu pierwszego okresu przepełnienia bufora β_1

$l_n(s)$ – transformata rozkładu czasu do przepełnienia bufora

m – liczba stanów łańcucha modulującego dla MMPP i BMAP

m_k – k -ty moment zwykły rozkładu prawdopodobieństwa

$m_k(t) = \mathbf{P}(N(t) = k)$ – rozkład liczby pakietów w przedziale czasu $(0, t]$ w procesie Poissona

$p(z)$ – funkcja tworząca dla ciągu p_i

p_i – prawdopodobieństwo pojawienia się grupy o rozmiarze i w złożonym strumieniu Poissona, $p_i = \mathbf{P}(\alpha_j = i)$

$p_i(j, k)$ – prawdopodobieństwa przejścia w BMAP

p_i^{k*} – k -krotny splot ciągu p_i

p – rozkład stacjonarny włożonego łańcucha Markowa (w chwilach nadejścia pakietu) dla MMPP

$q(l)$ – rozkład liczby pakietów traconych w okresie przepełnienia bufora w stanie ustalonym, $q(l) = \lim_{k \rightarrow \infty} \mathbf{P}(\gamma_k = l)$

$q_n(l)$ – rozkład liczby pakietów utraconych w pierwszym okresie przepełnienia bufora, $q_n(l) = \mathbf{P}(\gamma_1 = l | X(0) = n)$

$s_n(t)$ – gęstość prawdopodobieństwa czasu nadejścia n -tego pakietu w procesie Poissona

t_k – czas nadejścia k -tego pakietu

$w_n(s, \sigma)$ – transformata rozkładu opóźnienia

$Corr(k)$ – autokorelacja

$Cov(k)$ – autokowariancja

D_k – parametry BMAP w postaci macierzy $m \times m$

$F(x)$ – dystrybuanta rozkładu czasu obsługi

F^{k*} – k -krotny spłot rozkładu $F(x)$

\hat{F} – macierz przejścia dla markowskiego procesu odnowy w MMPP

$H(t)$ – dystrybuanta rozkładu okresu przepełnienia bufora dla dużych buforów ($\lim_{b \rightarrow \infty} H_{b-1}(t)$)

$H_n(t)$ – dystrybuanta rozkładu pierwszego okresu przepełnienia bufora β_1

I – macierz jednostkowa

$I(x > y)$ – funkcja charakterystyczna ($I(x > y) = 1$ gdy $x > y$ oraz 0 w przeciwnym wypadku)

$J(t)$ – modulujący łańcuch Markowa dla MMPP i BMAP

J_k – stan łańcucha modulującego w chwili nadejścia k -tego pakietu dla MMPP

$L(t)$ – liczba strat pakietów w przedziale $(0, t]$

LR – współczynnik strat; ułamek ogólnej liczby pakietów traconych na skutek przepełnienia bufora w dłuższych okresach czasu

$N(t)$ – liczba zdarzeń w strumieniu wejściowym w przedziale czasu $(0, t]$

$\mathbf{P}(\cdot)$ – prawdopodobieństwo zdarzenia

$P^*(z, t)$ – funkcja tworząca dla funkcji liczącej dla MMPP i BMAP

$P_{i,j}(n, t) = \mathbf{P}(N(t) = n, J(t) = j | N(0) = 0, J(0) = i)$ – funkcja licząca dla MMPP i BMAP

Q – macierz intensywności modulującego łańcucha Markowa dla MMPP

$R(z)$ – funkcja tworząca dla potencjału R_k

R_k – potencjał dla ciągu macierzy lub skalarów

$T_k = t_k - t_{k-1}$ – czas pomiędzy $(k - 1)$ -szym a k -tym pakietem

$V(t)$ – opóźnienie (jak długo przebywa w buforze pakiet przybyły w chwili t)

Var – wariancja

$X(t)$ – długość kolejki w chwili t (łącznie z aktualnie obsługiwanym)

α_j – rozmiar j -tej grupy zdarzeń w złożonym procesie Poissona lub w BMAP

β_k – długość k -tego okresu przepełnienia bufora

$\delta_n(s)$ – transformata liczby strat pakietów w przedziale $(0, t]$

δ_{ij} – symbol Kroneckera ($\delta_{ij} = 1$ jeżeli $i = j$ oraz $\delta_{ij} = 0$ w przeciwnym wypadku)

γ_k – liczba pakietów utraconych w k -tym okresie przepełnienia bufora

$\chi(t)$ – prawdopodobieństwo, że w przedziale czasu $(0, t]$ ani razu nie nastąpi przepełnienie bufora

$\chi_n(t)$ – ogon rozkładu pierwszego okresu przepełnienia bufora

λ – intensywność procesu Poissona lub średnia intensywność dla MMPP i BMAP

$(\lambda_1, \dots, \lambda_m)$ – wektor intensywności dla MMPP

μ_i – intensywności zmiany stanu lub przybycia pakietu w BMAP

ω_k – stan łańcucha modulującego tuż po k -tym okresie przepełnienia (dla MMPP)

$\phi_n(s, l)$ – transformata rozkładu długości kolejki

π – stacjonarna intensywność modulującego łańcucha Markowa dla MMPP i BMAP

ρ – obciążenie oferowane systemowi, tzn. iloczyn intensywności strumienia wejściowego i średniego czasu obsługi

τ_n – czas do przepelnienia bufora dla początkowej długości kolejki n

$\Delta(t)$ – średnia liczba strat pakietów w przedziale $(0, t]$

$\mathbf{E}(\cdot)$ – wartość oczekiwana

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ – macierz diagonalna z elementami $\lambda_1, \dots, \lambda_m$ na przekątnej

$\Phi_n(t, l)$ – rozkład długości kolejki w chwili t , przy założeniu że początkowa długość kolejki wynosiła n

$\Phi(l) = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) = l)$ – rozkład długości kolejki w stanie ustalonym

\mathcal{L} – wiarygodność parametryzacji

SPIS RYSUNKÓW I TABEL

Rysunki

2.1. Przykładowa ewolucja długości kolejki w systemie z buforem o rozmiarze b	32
2.2. Gęstość rozkładu okresu przepełnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.1).	40
2.3. Gęstość rozkładu okresu przepełnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.2).	41
2.4. Gęstość rozkładu okresu przepełnienia bufora w stanie ustalonym w systemach A i B (Przykład 2.3).	43
3.1. Autokorelacja odstępów czasu pomiędzy pakietami w funkcji opóźnienia dla parametryzacji MMPP z Przykładu 3.1.	64
3.2. Autokorelacja odstępów czasu pomiędzy pakietami w funkcji opóźnienia w Przykładzie 3.2.	66
3.3. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.3.	71
3.4. Rozkład długości kolejki w Przykładzie 3.3 w różnych chwilach czasu (stan nieustalony).	72
3.5. Prawdopodobieństwo przepełnienia bufora w funkcji rozmiaru bufora (Przykład 3.3) w różnych chwilach czasu (stan nieustalony).	73
3.6. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.4 dla bufora o rozmiarze 100 KB.	74
3.7. Rozkład długości kolejki w stanie ustalonym w Przykładzie 3.4 dla bufora 1 MB.	75
3.8. Rozkład długości kolejki w Przykładzie 3.4 w różnych chwilach czasu (stan nieustalony).	76
3.9. Średnia długość kolejki w funkcji czasu w Przykładzie 3.4 dla 5 różnych początkowych długości kolejki.	77
3.10. Średnia długość kolejki w funkcji czasu dla skorelowanego (MMPP) i nieskorelowanego (Poisson) ruchu (Przykład 3.4).	78

3.11. Prawdopodobieństwo przepełnienia bufora w funkcji czasu dla skorelowanego (MMPP) i nieskorelowanego (Poisson) ruchu (Przykład 3.4).	79
3.12. Współczynnik strat w funkcji rozmiaru bufora dla dwóch różnych obciążeń systemu ρ w Przykładzie 3.5.	80
3.13. Średnia liczba pakietów traconych w jednostce czasu (ms) w funkcji czasu dla 5 różnych początkowych długości kolejki – 0, 0.25 <i>b</i> , 0.5 <i>b</i> , 0.75 <i>b</i> , <i>b</i>	81
3.14. Średnia liczba pakietów traconych w jednostce czasu w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i początkowej kolejki 0 w Przykładzie 3.5.	82
3.15. Średnia liczba pakietów traconych w jednostce czasu w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i początkowej kolejki 0.5 <i>b</i> w Przykładzie 3.5.	83
3.16. Średnia liczba pakietów traconych w jednostce czasu w funkcji czasu dla różnych początkowych stanów łańcucha modulującego i początkowej kolejki <i>b</i> w Przykładzie 3.5.	84
3.17. Średnie opóźnienie kolejkowania w czasie (Przykład 3.6).	86
3.18. Średnie opóźnienie w funkcji rozmiaru bufora w różnych chwilach czasu (Przykład 3.6).	87
3.19. Średnie opóźnienie kolejkowania w funkcji czasu (Przykład 3.7) dla różnych początkowych zapełnień bufora - 0, 0.25 <i>b</i> , 0.5 <i>b</i> , 0.75 <i>b</i> , <i>b</i>	88
3.20. Gęstość rozkładu prawdopodobieństwa pierwszego okresu przepełnienia bufora ($h_{n,i}(t)$) dla różnych rozmiarów buforów w Przykładzie 3.8.	94
3.21. Średnia długość pierwszego okresu przepełnienia bufora w funkcji rozmiaru bufora w Przykładzie 3.8.	94
3.22. Gęstość rozkładu prawdopodobieństwa pierwszego okresu przepełnienia bufora ($h_{n,i}(t)$) dla różnych <i>n</i> w Przykładzie 3.8.	95
3.23. Gęstości rozkładu prawdopodobieństwa dla kolejnych okresów przepełnienia bufora w Przykładzie 3.8.	95
3.24. Gęstości rozkładu prawdopodobieństwa dla kolejnych okresów przepełnienia bufora w Przykładzie 3.9.	96
3.25. Gęstości rozkładu prawdopodobieństwa okresu przepełnienia bufora w stanie ustalonym dla kolejek ze skorelowanym (MMPP) i nieskorelowanym (Poisson) strumieniem pakietów (Przykład 3.8).	97
3.26. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla różnych wartości obciążenia systemu (Przykład 3.12).	104

3.27. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla strumienia Poissona i różnych wartości obciążenia systemu (Przykład 3.12).	106
4.1. Rozkład długości kolejki w stanie ustalonym - Przykład 4.3.	120
4.2. Rozkład długości kolejki w różnych chwilach czasu w Przykładzie 4.3.	121
4.3. Prawdopodobieństwo przepełnienia bufora w funkcji czasu w Przykładzie 4.3 dla różnych początkowych długości kolejki.	122
4.4. Prawdopodobieństwo przepełnienia bufora w funkcji czasu dla skorelowanego (BMAP) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.3).	123
4.5. Prawdopodobieństwo przepełnienia bufora w stanie ustalonym w funkcji rozmiaru bufora dla skorelowanego (BMAP) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.3).	124
4.6. Rozkład długości kolejki w stanie ustalonym w Przykładzie 4.4 dla bufora 100 KB.	125
4.7. Rozkład długości kolejki w stanie ustalonym w Przykładzie 4.4 dla bufora 200 KB.	126
4.8. Prawdopodobieństwo przepełnienia bufora w stanie ustalonym w funkcji rozmiaru bufora dla obciążenia ρ od 75% do 99% (Przykład 4.4).	127
4.9. Średnie opóźnienie w funkcji czasu w Przykładzie 4.5 dla różnych wartości $X(0)$	128
4.10. Średnie opóźnienie w funkcji czasu w Przykładzie 4.5 dla różnych wartości $J(0)$	130
4.11. Średnie opóźnienie w stanie ustalonym w funkcji rozmiaru bufora dla różnych wartości obciążenia ρ (Przykład 4.6).	129
4.12. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla skorelowanego (BMAP) i nieskorelowanego (M^X) modelu ruchu (Przykład 4.7).	134
4.13. Średni czas do przepełnienia bufora w funkcji rozmiaru bufora dla różnych wartości obciążenia (Przykład 4.8).	135
5.1. Algorytm EM dla BMAP.	152

Tabele

2.1. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym (Przykład 2.1).	40
2.2. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym (Przykład 2.2).	42
2.3. Prawdopodobieństwo utraty l kolejnych pakietów w okresie przepełnienia bufora w stanie ustalonym (Przykład 2.3).	44
3.1. Podstawowe parametry zapisanego ruchu IP i dopasowanego MMPP w Przykładzie 3.2.	65
3.2. Prawdopodobieństwa utraty l kolejnych pakietów w k -tym okresie przepełnienia bufora w Przykładzie 3.10.	100
3.3. Prawdopodobieństwa utraty l kolejnych pakietów w okresie przepełnienia bufora (w stanie ustalonym) dla skorelowanego (MMPP) i nieskorelowanego (Poisson) modelu ruchu (Przykład 3.11).	100
4.1. Podstawowe parametry zapisanego ruchu IP i dopasowanego BMAP w Przykładzie 4.2.	116

INFORMATION FOR AUTHORS

The journal *STUDIA INFORMATICA* publishes both fundamental and applied Memoirs and Notes in the field of informatics. The Editors' aim is to provide an active forum for disseminating the original results of theoretical research and applications practice of informatics understood as a discipline focused on the investigations of laws that rule processes of coding, storing, processing, and transferring of information or data.

Papers are welcome from fields of informatics inclusive of, but not restricted to *Computer Science, Engineering, and Life and Physical Sciences*.

All manuscripts submitted for publication will be subject to critical review. Acceptability will be judged according to the paper's contribution to the art and science of informatics.

In the first instance, all text should be submitted as hardcopy, conventionally mailed, and for accepted paper accompanying with the electronically readable manuscript to:

Dr. Marcin SKOWRONEK
Institute of Informatics
Silesian University of Technology
ul. Akademicka 16
44-100 Gliwice, Poland
Tel.: +48 32 237-12-15
Fax: +48 32 237-27-33
e-mail: marcins@polsl.pl

MANUSCRIPT REQUIREMENTS

All manuscripts should be written in Polish or in English. Manuscript should be typed on one side paper only, and submitted in duplicate. The name and affiliation of each author should be followed by the title of the paper (as brief as possible). An abstract of not more than 50 words is required. The text should be logically divided under numbered headings and subheadings (up to four levels). Each table must have a title and should be cited in the text. Each figure should have a caption and have to be cited in the text. References should be cited with a number in square brackets that corresponds to a proper number in the reference list. The accuracy of the references is the author's responsibility. Abbreviations should be used sparingly and given in full at first mention (e.g. "Central Processing Unit (CPU)"). In case when the manuscript is provided in Polish (English) language, the summary and additional abstract (up to 300 words with reference to the equations, tables and figures) in English (Polish) should be added.

After the paper has been reviewed and accepted for publication, the author has to submit to the Editor a hardcopy and electronic version of the manuscript.

It is strongly recommended to submit the manuscript in a form downloadable from web site <http://zti.iinf.polsl.gliwice.pl/makiety/>.

To subscribe: *STUDIA INFORMATICA* (PL ISSN 0208-7286) is published by Silesian University of Technology Press (Wydawnictwo Politechniki Śląskiej) ul. Akademicka 5, 44-100 Gliwice, Poland, Tel./Fax +48 32 237-13-81. 2007 annual subscription rate: US\$40. Single number price approx. US\$8-15 according to the issue volume.