

POLITECHNIKA CZĘSTOCHOWSKA  
WYDZIAŁ INŻYNIERII MECHANICZNEJ  
I INFORMATYKI

mgr Marcin Ziółkowski

**Modele i algorytmy wyznaczania  
objętości pamięci systemów  
informatycznych w warunkach  
niepewności stochastycznej**

rozprawa doktorska napisana pod kierunkiem prof. dr hab. Olega Tikhonenko

Częstochowa 2011



---

## Spis treści

<b>1</b>	<b>Wstęp</b> .....	5
1.1	Tło prowadzonych badań .....	5
1.2	Sformułowanie problemu badawczego, teza i cel pracy .....	7
<b>2</b>	<b>Elementy klasycznej teorii obsługi masowej</b> .....	11
2.1	Elementy teorii strumienia wejściowego. Strumień najprostszy	11
2.2	Modelowanie czasu obsługi zgłoszeń .....	15
2.3	Analiza ogólna klasycznych systemów obsługi. Tryb stacjonarny .....	16
2.3.1	Klasyfikacja systemów obsługi. Symbolika Kendalla	16
2.3.2	Tryb stacjonarny działania klasycznego systemu obsługi .....	18
<b>3</b>	<b>Przegląd wybranych klasycznych systemów obsługi</b> .....	21
3.1	Łańcuchy Markowa z czasem ciągłym .....	21
3.2	Metoda równań różniczkowych. Systemy markowskie .....	25
3.2.1	Metoda równań różniczkowych .....	25
3.2.2	Równania równowagi .....	26
3.2.3	Procesy narodzin i śmierci .....	28
3.2.4	Markowskie systemy obsługi $M/M/n/m$ .....	30
3.3	Jednoliniowy system obsługi $M/G/1/\infty$ .....	33
3.3.1	Procesy semimarkowskie .....	33
3.3.2	Analiza jednoliniowego systemu obsługi $M/G/1/\infty$	34
<b>4</b>	<b>Systemy informatyczne o stochastycznym charakterze działania</b> .....	41
4.1	Klasyfikacja stochastycznych systemów informatycznych .....	42

4.1.1	Modele obsługi zgłoszeń niejednorodnych . . . . .	42
4.1.2	Klasyfikacja modeli systemów informatycznych o stochastycznym charakterze działania . . . . .	44
4.2	Analiza systemu obsługi $M/M/n/(m, V)$ . . . . .	47
4.3	Zagadnienie wyznaczania charakterystyk utrat informacji . . . . .	50
4.4	Analiza systemu obsługi $M/G/1/\infty$ . . . . .	53
<b>5</b>	<b>Model pamięci operacyjnej systemu informatycznego . . . . .</b>	<b>55</b>
5.1	Wstęp . . . . .	55
5.2	Opis i analiza modelu . . . . .	56
5.3	Wyznaczenie charakterystyk objętości sumarycznej . . . . .	64
5.4	Realizacje techniczne analizowanego modelu . . . . .	69
5.5	Wyznaczenie charakterystyk utrat informacji w szczególnych przypadkach . . . . .	70
5.5.1	Schemat 1. Rozmiar procesów i czas obsługi procesów niezależne. . . . .	70
5.5.2	Schemat 1. Czas obsługi proporcjonalny do rozmiaru procesu. . . . .	71
5.5.3	Schemat 2. Rozmiar procesów i czas obsługi procesów niezależne. . . . .	73
5.5.4	Schemat 2. Czas obsługi proporcjonalny do rozmiaru procesu. . . . .	74
<b>6</b>	<b>Układy systemów połączonych poprzez wspólną pamięć (model serwera) . . . . .</b>	<b>79</b>
6.1	Wstęp . . . . .	79
6.2	Analiza modelu . . . . .	80
6.3	Wyznaczenie prawdopodobieństw odmowy . . . . .	83
6.4	Uogólnienie badanego modelu oraz pewne przypadki szczególne . . . . .	83
6.5	Realizacje techniczne omawianego modelu . . . . .	91
<b>7</b>	<b>Dodatki . . . . .</b>	<b>95</b>
7.1	Omówienie działania programu symulacyjnego dla modelu RAM - schemat 1 . . . . .	95
7.2	Omówienie działania programu symulacyjnego dla modelu RAM - schemat 2 . . . . .	98
7.3	Omówienie działania programu dla modelu serwera - nieograniczona kolejka . . . . .	99
7.4	Omówienie działania programu dla modelu serwera - ograniczona kolejka . . . . .	101

Spis treści	3
<b>8 Podsumowanie</b> .....	<b>103</b>
<b>Literatura</b> .....	<b>107</b>



## Wstęp

### 1.1 Tło prowadzonych badań

Zagadnienie wyznaczania objętości pamięci pojawiło się w latach siedemdziesiątych ubiegłego stulecia w związku z rozwojem sieci telekomunikacyjnych. Problem odpowiedniego doboru rozmiaru pamięci buforowej węzłów telekomunikacyjnych był początkowo rozwiązywany przez inżynierów telekomunikacji w oparciu o wyniki klasycznej teorii obsługi masowej. Wykorzystywano wówczas wyniki dotyczące zbadanych modeli i adaptowano je na potrzeby rozwijającej się teorii. Pierwsze próby oszacowań objętości sumarycznej w realnych systemach telekomunikacyjnych w oparciu o proste modele teorii obsługi masowej można znaleźć w pracach [27] oraz [28]. W badanych modelach (np. model z ciągłymi zgłoszeniami) zakładano jednak, że charakter zależności czasu obsługi i objętości zgłoszenia nie ma wpływu na charakterystyki utrat informacji. Wykorzystywano również założenie, że w przypadku małych prawdopodobieństw utrat informacji mogą być one wyznaczane w oparciu o charakterystyki dystrybucyjności objętości sumarycznej w analogicznych systemach lecz z nieograniczoną objętością sumaryczną. Prowadziło to do błędnego wyznaczania rozmiaru pamięci buforowej węzłów telekomunikacyjnych. Taki dobór nie gwarantował, że ustalone z góry prawdopodobieństwa utrat informacji nie będą przekroczone.

Są dwa ważne powody, dla których takie podejście nie było poprawne. Po pierwsze okazało się, że charakter zależności czasu obsługi i objętości zgłoszenia ma znaczący wpływ na charakterystyki utrat informacji. Po drugie nieuzasadnione okazało się zastępowanie w obliczeniach charakterystyk realnego systemu telekomunikacyjnego (najczęściej o ograniczonej objętości sumarycznej) odpowiednimi charakterystykami systemu o podobnym kształcie, lecz z nieograniczoną objętością sumaryczną.

Powstała zatem potrzeba utworzenia nowego kierunku badawczego opierającego się na klasycznej teorii obsługi masowej uwzględniającego fakt, że w ogólnym przypadku czas obsługi i objętość zgłoszenia mogą być zależne

oraz zakładającego ewentualne ograniczenie objętości sumarycznej. W latach osiemdziesiątych powstały uogólnienia klasycznych modeli teorii obsługi masowej uwzględniające wspomniane fakty. Na podstawie tych modeli powstały metody wyznaczania charakterystyk utrat informacji. Doprowadziło to do otrzymania algorytmów wyznaczania objętości pamięci systemów telekomunikacyjnych, a w późniejszym okresie także informatycznych.

W pracy [23] zbadany został system Erlanga  $M/G/n/0$ , w którym każde zgłoszenie charakteryzuje się dodatkowo losową objętością, przy czym czas obsługi i objętość zgłoszenia są niezależne. Dla takiego systemu zostały wyznaczone charakterystyki objętości sumarycznej. Jako początek nowopowstałego kierunku badawczego można uznać pozycję [16]. W publikacji tej po raz pierwszy zostały wprowadzone nowe metody obliczeniowe będące uogólnieniem klasycznych metod obliczeniowych znanych z teorii obsługi masowej. Metody te po raz pierwszy uwzględniały zależność czasu obsługi i objętości zgłoszenia. W pracy tej uzyskano charakterystyki objętości sumarycznej dla systemu  $M/G/1/\infty$  w trybie stacjonarnym oraz zbadano system obsługi  $M/M/n/(m, V)$ . Oba modele są używane najczęściej w sytuacjach praktycznych, ponieważ mają liczne odpowiedniki techniczne.

W pracach [29], [31], [32], oraz [33] analizowane były systemy  $M/G/1/\infty$  oraz systemy z podziałem procesora. Ten ostatni model jest również jednym z najważniejszych modeli używanych do wyznaczania charakterystyk objętości sumarycznej w realnych systemach informatycznych. W pracach [23], [42] oraz [43] były badane uogólnienia klasycznego systemu Erlanga  $M/G/n/0$ . Prace [25] oraz [26] zawierają analizy systemów obsługi zgłoszeń niejednorodnych  $M/G/1/\infty$  oraz pewnego uogólnienia tego modelu ze strumieniem wejściowym typu  $MAP$ , a praca [40] - analizę systemu obsługi zgłoszeń niejednorodnych z rekurencyjnym strumieniem wejściowym i wykładniczym czasem obsługi.

W pracy [31] zbadane zostało uogólnienie systemu  $M/G/1/\infty$ , w którym zgłoszenia przybywają grupowo. W pracach [32], [33] została przeprowadzona analiza systemu jednoliniowego w warunkach niestacjonarnych, a także zbadane zostały systemy  $M/G/n/0$  oraz  $M/G/\infty$ . W pracy [34] analizowane były bardziej skomplikowane systemy z priorytetem względnym, a w pracy [39] - systemy z priorytetem bezwzględnym.

Praca [35] może być uważana za jedyną do tej pory monografię dotyczącą tego zagadnienia. W pracy [37] uzyskane zostało uogólnienie klasycznych wzorów Little'a. Z uogólnienia tego wynika w szczególności, że charakter zależności czasu obsługi i objętości zgłoszenia ma wpływ na charakterystyki objętości sumarycznej już na poziomie jej pierwszego momentu. W pracach [21] oraz [22] zostały uzyskane charakterystyki objętości sumarycznej dla systemu  $M/G/1/(\infty, V)$ , przy czym założono, że zgłoszenia są obsługiwane zgodnie z dyscypliną LIFO. Praca [44] jest jedynym w języku polskim podręcznikiem akademickim zawierającym systematyczny przegląd metod i modeli teorii obsługi masowej wzbogaconym dodatkowo w swym ostatnim rozdziale o opis teorii systemów obsługi zgłoszeń niejednorodnych.



W związku z istotnym postępem technicznym w dziedzinie informatyki w ostatnich latach związanym przede wszystkim z rozwojem sieci komputerowych oraz zmieniającą się architekturą sprzętu komputerowego wzrosła rola omawianych metod i modeli. Okazało się, że uogólnienia modeli teorii obsługi masowej świetnie nadają się do badania systemów informatycznych o stochastycznym charakterze działania. Znane z teorii telekomunikacji pojęcie pamięci buforowej może być bowiem rozumiane w przypadku systemów informatycznych na przykład jako pamięć operacyjna lub pamięć dyskowa w zależności od charakteru badanego problemu. Kierunek badawczy zajmujący się podobnymi problemami zyskał więc na znaczeniu i pozwala na rozwiązywanie praktycznych problemów związanych z odpowiednim doбором rozmiaru pamięci buforowej systemów informatycznych. Stał się oddzielnym kierunkiem, który mocno rozwija się w ostatnich latach (wzrastająca liczba publikacji), a jego rola w przyszłości zapewne wzrośnie.

## 1.2 Sformułowanie problemu badawczego, teza i cel pracy

Jako system informatyczny o stochastycznym charakterze działania rozumiemy tu pojmowany w sensie klasycznym system obsługi składający się z jednego lub wielu urządzeń, do których w losowych chwilach czasu przybywają zgłoszenia, których obsługa odbywa się w przedziale czasu o losowej długości. Dodatkowo zakładamy, że zgłoszenia te posiadają pewien losowy rozmiar (objętość), który w ogólnym przypadku jest nieujemną zmienną losową oznaczającą rozmiar pamięci potrzebnej do zapisania informacji o zgłoszeniu.

Założenie to prowadzi do uogólnienia rozpatrywanych w teorii obsługi masowej modeli i pozwala na analizowanie modeli, w których zgłoszenia mogą być rozumiane jako pliki zapisywane na dysku komputera, procesy zajmujące pamięć operacyjną komputera czy żądania skierowane do serwera obsługującego wiele komputerów. Z uwagi na takie potraktowanie zagadnienia powstało wiele nierozwiązanych problemów. Jako przykłady takich problemów można podać między innymi:

- budowę matematycznego modelu pamięci RAM
- budowę modelu serwera obsługującego żądania wielu użytkowników

W pracy tej problemy te zostały rozwiązane. Sformułowanie problemu pozwala na określenie tezy i celu pracy.

**Teza pracy:** Wykorzystanie metod teorii obsługi masowej zgłoszeń o losowej objętości pozwala na zbudowanie modeli realnych systemów informatycznych. Wyniki uzyskane na podstawie analizy tych modeli są przeznaczone do opracowania algorytmów wyznaczania charakterystyk utrat informacji i w konsekwencji do uzyskania metod wyznaczania objętości pamięci realnych systemów informatycznych.

**Cel pracy:** Opracowanie modeli i algorytmów wyznaczania objętości pamięci wybranych realnych systemów informatycznych o stochastycznym charakterze działania

**Cele szczegółowe:**

- Zbudowanie i analiza modelu pamięci operacyjnej komputera
- Opracowanie algorytmu wyznaczania objętości pamięci na podstawie zbudowanego modelu pamięci operacyjnej
- Zbudowanie i analiza modelu serwera
- Opracowanie algorytmu wyznaczania objętości pamięci na podstawie zbudowanego modelu serwera

Poniższa praca jest zorganizowana w następujący sposób.

Rozdział 2 stanowi przegląd podstawowych wiadomości z klasycznej teorii obsługi masowej wykorzystywanych podczas analizy systemów informatycznych o stochastycznym charakterze działania.

Rozdział 3 to przegląd najważniejszych klasycznych modeli teorii obsługi masowej, których uogólnienia stanowią bazę przy analizie modeli obsługi zgłoszeń niejednorodnych. W rozdziale tym przeprowadzone są m.in. szczegółowe analizy dla modeli systemów  $M/M/n/m$  oraz  $M/G/1/\infty$ .

W rozdziale 4 zaprezentowano podstawy teorii modeli obsługi zgłoszeń niejednorodnych uwzględniającej zależność czasu obsługi i objętości zgłoszenia oraz ewentualne ograniczenie objętości sumarycznej. Przedstawione są tu główne wyniki związane z systemami  $M/M/n/(m, V)$ ,  $M/G/1/\infty$  oraz wnioski związane z wykorzystaniem analizy modeli obsługi zgłoszeń o losowej objętości, w których czas obsługi jest zależny od objętości zgłoszenia a objętość sumaryczna jest nieograniczona do szacowania charakterystyk utrat informacji w analogicznych systemach, lecz z ograniczoną objętością sumaryczną.

Rozdział 5 oraz 6 stanowią główną oraz samodzielną część realizacji celu pracy.

W rozdziale 5 przeprowadzona jest ogólna analiza modelu pamięci operacyjnej RAM opartego na uogólnieniu systemu  $M/G/1/\infty$  z nieograniczoną objętością sumaryczną. W dalszym ciągu zaprezentowany jest przegląd pewnych, ważnych z punktu widzenia praktycznego, przypadków, a także przeprowadzona jest analiza porównawcza uzyskanych wyników analitycznych opartych na oszacowaniach charakterystyk utrat informacji z wynikami otrzymanymi w trakcie prowadzenia symulacji komputerowych opartych na odpowiednio stworzonych programach.

W rozdziale 6 natomiast przeprowadzona jest analiza ogólna modelu serwera obsługującego żądania wielu użytkowników opartego na uogólnieniu systemu  $M/M/n/(m, V)$ . W dalszym ciągu przedstawione są numeryczne rezultaty związane z obliczaniem charakterystyk utrat informacji w pewnych

szczególnych przypadkach wraz z analizą porównawczą otrzymanych dokładnych wyników z wynikami otrzymanymi podczas symulacji z wykorzystaniem odpowiednich programów komputerowych.

Część wyników zawartych w rozdziałach 5 oraz 6 została opublikowana w pracach [45], [46], [47] oraz [48].

W rozdziale 7 przedstawiono natomiast omówienia zastosowanych programów komputerowych, które były wykorzystane do symulacji analizowanych modeli oraz uzyskiwania charakterystyk utrat informacji w celu przeprowadzenia analizy porównawczej z wynikami uzyskanymi w postaci analitycznej.

Autor pracy pragnie w sposób szczególny podziękować wielu osobom, których cenne wsparcie pomogło w pracy nad tą rozprawą.

Największe podziękowania należą się Panu prof. dr hab. Olegowi Tikhonenko za cenne uwagi, wsparcie merytoryczne oraz częste i owocne dyskusje podczas prowadzonych seminariów naukowych.

Autor dziękuje też władzom Instytutu Matematyki i Informatyki Akademii im. Jana Długosza oraz Wydziału Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej.

Wreszcie podziękowania należą się też wszystkim osobom wspierającym duchowo trud pracy nad tą rozprawą m.in doktorowi Mirosławowi Kurkowskiemu oraz najbliższej rodzinie i przyjaciołom.

Częstochowa, czerwiec 2011



---

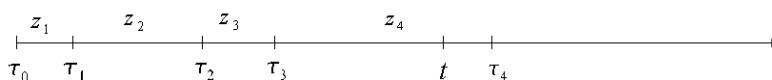
## Elementy klasycznej teorii obsługi masowej

W rozdziale tym przedstawione są podstawowe wiadomości z klasycznej teorii obsługi masowej. Informacje te są niezbędne dla lepszego zrozumienia mechanizmu działania stochastycznego systemu informatycznego oraz poznania pewnych analitycznych metod badania systemów informatycznych. Omówione zostały przede wszystkim najważniejsze wiadomości z tzw. teorii strumienia wejściowego oraz wprowadzone zostały podstawowe pojęcia związane z analizą ogólną modeli obsługi masowej. Większość omawianych pojęć klasycznej teorii obsługi masowej przedstawionych jest szerzej między innymi w pracach [6], [7], [8], [9], [10], [11], [18], [19], [30] oraz [44]. Wykorzystywane tu pojęcia klasycznego rachunku prawdopodobieństwa można natomiast znaleźć w pracach [2], [5] oraz [15].

### 2.1 Elementy teorii strumienia wejściowego. Strumień najprostszy

Założmy, że w pewnych losowych chwilach czasu  $\tau_1, \tau_2, \tau_3, \dots$  do systemu obsługi przybywają zgłoszenia. Wprowadźmy następujące oznaczenie:

$$z_k = \tau_k - \tau_{k-1}, \quad k = 1, 2, \dots$$



Rysunek 2.1. Wejściowy strumień zgłoszeń

Wielkość  $z_k$  jest oczywiście odstępem losowym czasu między sąsiednimi chwilami przybycia zgłoszeń. Załóżmy, że  $\tau_0 \equiv 0$ . Niech  $\nu(t)$  będzie liczbą zgłoszeń przybyłych do systemu w przedziale czasowym  $[0, t)$  (rys. 2.1).

Wprowadzone zostaną teraz dwie równoważne definicje strumienia wejściowego zgłoszeń.

**Definicja 2.1.** Strumieniem zgłoszeń nazywamy proces stochastyczny  $\nu(t)$ , który przyjmuje wartości naturalne, takie, że  $\nu(0) = 0$  oraz dla dowolnych  $t_1, t_2$  takich, że  $t_1 < t_2$  zachodzi nierówność  $\nu(t_1) \leq \nu(t_2)$ .

**Definicja 2.2.** Strumieniem zgłoszeń nazywamy ciąg zmiennych losowych nieujemnych  $z_k$ ;  $k = 1, 2, \dots$

W wielu realnych zagadnieniach mamy do czynienia z tzw. strumieniami rekurencyjnymi.

**Definicja 2.3.** Strumień zgłoszeń nazywamy rekurencyjnym, jeżeli dla dowolnej liczby naturalnej  $n \geq 1$  zmienne losowe  $z_1, z_2, \dots, z_n$  są niezależne i mają taki sam rozkład.

Zatem strumień zgłoszeń jest rekurencyjny, jeśli wszystkie odstępy między sąsiednimi chwilami przybycia zgłoszeń mają jednakowy rozkład określony przez pewną dystrybuantę nieujemnej zmiennej losowej  $A(t)$  tj.

$$P\{z_1 < t\} = P\{z_2 < t\} = \dots = A(t).$$

Można udowodnić, że dla strumienia rekurencyjnego zachodzi pewien związek między przekształceniem Laplace'a-Stieltjesa (PLS) dystrybuanty  $A(t)$  odstępów czasu między sąsiednimi chwilami przybycia zgłoszeń a funkcją tworzącą (FT) procesu  $\nu(t)$ .

Wprowadźmy następujące oznaczenia:

$$P_k(t) = P\{\nu(t) = k\}, \quad P(z, t) = Ez^{\nu(t)},$$

$$\alpha(q) = \int_0^\infty e^{-qt} dA(t), \quad \pi(z, q) = \int_0^\infty e^{-qt} P(z, t) dt.$$

$P_k(t)$  jest prawdopodobieństwem tego, że do systemu w przedziale czasowym  $[0, t)$  przybędzie  $k$  zgłoszeń,  $P(z, t)$  jest FT procesu  $\nu(t)$ ,  $\pi(z, q)$  – przekształceniem Laplace'a funkcji  $P(z, t)$ , natomiast  $\alpha(q)$  jest PLS dystrybuanty  $A(t)$ .

**Lemat 2.1.** Przy wprowadzonych oznaczeniach zachodzi następujący związek:

$$\pi(z, q) = \frac{1 - \alpha(q)}{q[1 - z\alpha(q)]}. \quad (2.1)$$

Ze związku tego wynika, że jeśli w strumieniu rekurencyjnym mamy daną dystrybuantę odstępu czasu między sąsiednimi chwilami przybycia zgłoszeń, to możemy w oparciu o aparat przekształceń Laplace'a i funkcji tworzących wyznaczyć prawdopodobieństwa  $P_k(t)$ . Istotnie, korzystając z metody residuów mając dane przekształcenie Laplace'a funkcji tworzącej  $P(z, t)$  niekiedy jesteśmy w stanie wyznaczyć tę funkcję, następnie korzystając z własności funkcji tworzącej lub rozwijając funkcję  $P(z, t)$  w szereg potęgowy możemy wyznaczyć funkcję  $P_k(t)$ .

Szczególnym przypadkiem strumienia rekurencyjnego jest strumień najprostszy. Strumienie najprostsze odgrywają bardzo ważną rolę z uwagi na ich szerokie występowanie praktyczne w modelach obsługi masowej, a także użyteczność obliczeniową.

**Definicja 2.4.** Strumieniem najprostszym z parametrem  $a$  nazywamy strumień rekurencyjny, w którym odstępy między sąsiednimi chwilami przybycia zgłoszeń mają rozkład wykładniczy z parametrem  $a$  tzn.

$$A(t) = P\{z_k < t\} = 1 - e^{-at}, \quad t > 0, \quad a > 0, \quad k = 1, 2, \dots$$

W dalszym ciągu przedstawione zostaną podstawowe własności strumieni najprostszych.

Wyznamy na początku liczbę losową zgłoszeń strumienia najprostszego z parametrem  $a$  przybyłych do systemu w przedziale czasowym  $[0, t)$  tj. określimy funkcje  $P_k(t)$ .

Można wykazać, że PLS  $\alpha(q)$  dystrybuanty  $A(t)$  w tym przypadku ma postać:

$$\alpha(q) = \frac{a}{a + q}.$$

Korzystając ze wzoru (2.1) otrzymujemy:

$$\pi(z, q) = \frac{1}{q + a - az}$$

Stąd korzystając z metody residuów otrzymujemy następującą postać FT procesu  $\nu(t)$ :

$$P(z, t) = e^{-(a-az)t} = e^{-at(1-z)}.$$

Zauważmy, że jest to funkcja tworząca rozkładu Poissona z parametrem  $at$ . Stąd:

$$P_k(t) = \frac{(at)^k}{k!} e^{-at}.$$

Oznacza to, że dla ustalonego  $t$  zmienna losowa  $\nu(t)$  ma rozkład Poissona z parametrem  $at$ .

Strumień najprostszy ma trzy charakterystyczne własności. Są to własności **stacjonarności**, **braku następstw** i **pojedynczości**.

Niech  $\nu(a, b) = \nu(b) - \nu(a)$  oznacza liczbę zgłoszeń przybyłych do systemu w przedziale czasu  $[a, b)$ ,  $\pi_2(T, T+t) = \sum_{k=2}^{\infty} P_k(T, T+t)$  - niech oznacza prawdopodobieństwo pojawienia się co najmniej dwóch zgłoszeń w przedziale  $[T, T+t)$ .

**Stacjonarność** oznacza, że dla każdej liczby naturalnej  $n \geq 1$ , każdego ciągu liczb rzeczywistych  $t_1, \dots, t_n$ , takich, że  $0 < t_1 < \dots < t_n$  i każdego  $t_0 > 0$  rozkład wektora losowego  $(\nu(t_0 + t_1) - \nu(t_0), \dots, \nu(t_0 + t_n) - \nu(t_0))$  nie zależy od  $t_0$ .

W szczególności rozkład liczby zgłoszeń strumienia najprostszego przybyłych do systemu w przedziale czasowym o długości  $t$  nie zależy od położenia tego przedziału na osi czasu.

**Brak następstw** oznacza, że dla dowolnego ciągu przedziałów rozłącznych  $[a_1, b_1), [a_2, b_2), \dots$  zmienne losowe  $\nu(a_1, b_1), \nu(a_2, b_2), \dots$  są niezależne. Oznacza to, że prawdopodobieństwo przybycia  $k$  zgłoszeń w dowolnym przedziale czasu  $[T, T+t)$  nie zależy od tego, co działo się przed chwilą  $T$ .

**Pojedynczość** oznacza, że praktycznie niemożliwe jest pojawienie się w tej samej chwili czasu dwóch lub więcej zgłoszeń tzn. dla dowolnego  $t \geq 0$  i  $\Delta t \rightarrow 0$  mamy:

$$\pi_2(t, t + \Delta t) = o(\Delta t).$$

Obliczmy dla strumienia najprostszego wartość oczekiwaną liczby zgłoszeń, które przybyły do systemu w przedziale  $[0, t)$  (a wobec stacjonarności także w dowolnym przedziale czasowym o długości  $t$ ). Mamy:

$$E\nu(t) = \sum_{k=0}^{\infty} k P_k(t) = \sum_{k=0}^{\infty} k \frac{(at)^k}{k!} e^{-at} = e^{-at} \sum_{k=1}^{\infty} \frac{(at)^k}{(k-1)!} =$$



$$= ate^{-at} \sum_{k=1}^{\infty} \frac{(at)^{k-1}}{(k-1)!} = ate^{-at} e^{at} = at.$$

Stąd wynika, że parametr  $a$  ma sens intensywności tzn. średniej liczby zgłoszeń przybyłych w ciągu jednostki czasu. Natomiast średni odstęp między sąsiednimi chwilami przybycia zgłoszeń wynosi  $\frac{1}{a}$ .

## 2.2 Modelowanie czasu obsługi zgłoszeń

Założmy, że pewne urządzenie obsługuje przybywające do niego zgłoszenia. W ogólnym przypadku możemy założyć, że czas obsługi jest nieujemną zmienną losową. Ponumerujmy zgłoszenia zgodnie z kolejnością ich przybycia do urządzenia obsługi liczbami  $1, 2, \dots$  i oznaczmy przez  $\xi_k$ ,  $k \geq 1$ , czas obsługi zgłoszenia o numerze  $k$ . Założmy dodatkowo, że czas obsługi nie zależy od charakterystyk strumienia wejściowego. Obsługa jest w pełni określona, jeśli dla dowolnej liczby naturalnej  $n \geq 1$  określony jest rozkład wektora losowego  $(\xi_1, \dots, \xi_n)$ .

**Definicja 2.5.** Obsługę nazywamy rekurencyjną, jeżeli dla dowolnej liczby naturalnej  $n \geq 1$  zmienne losowe  $\xi_1, \dots, \xi_n$  są niezależne i spełniają ten sam rozkład.

Jest oczywiste, że dla określenia obsługi rekurencyjnej wystarczy podać dystrybuantę czasu obsługi  $B(t) = P\{\xi_k < t\} = P\{\xi < t\}$ ,  $k \geq 1$ .

**Definicja 2.6.** Obsługę rekurencyjną, dla której  $B(t) = 1 - e^{-\mu t}$  nazywa się obsługą wykładniczą z parametrem  $\mu$ .

Wykładniczy rozkład czasu obsługi jest dość często spotykanym rozkładem w teorii obsługi masowej. Oprócz rozkładu wykładniczego jako często występujący rozkład czasu obsługi można wymienić rozkład degeneratywny (czas obsługi stały, taki sam dla wszystkich zgłoszeń).

Jeżeli chodzi o obsługę deterministyczną (czas obsługi stały równy  $t_0$ ), to dystrybuanta czasu obsługi w takim przypadku ma oczywiście postać:

$$B(t) = \begin{cases} 0, & \text{dla } t \leq t_0, \\ 1, & \text{dla } t > t_0. \end{cases}$$

W przypadku obsługi wykładniczej z parametrem  $\mu$  wielkość  $\mu\Delta t + o(\Delta t)$  określa prawdopodobieństwo zakończenia obsługi w przedziale czasu  $[t, t + \Delta t)$  pod warunkiem, że obsługa trwała w chwili  $t$ . Natomiast

wielkość  $1 - \mu\Delta t + o(\Delta t)$  jest prawdopodobieństwem tego, że obsługa nie zakończy się w przedziale czasowym  $[t, t + \Delta t)$  pod tym samym warunkiem (są to wnioski wynikające z własności rozkładu wykładniczego).

## 2.3 Analiza ogólna klasycznych systemów obsługi. Tryb stacjonarny

### 2.3.1 Klasyfikacja systemów obsługi. Symbolika Kendalla

Klasyfikację systemów obsługi prowadzimy zwykle według następujących cech:

#### 1. Typ wejściowego strumienia zgłoszeń

Najczęściej w analizowanych modelach strumień wejściowy jest rekurencyjny. Stąd jego konkretny typ określa się przez podanie dystrybuanty  $A(t)$  odstępów czasu między sąsiednimi chwilami przybycia zgłoszeń.

#### 2. Rozkład łączny czasu obsługi zgłoszeń

W przeważającej części analizowanych modeli obsługa zgłoszeń jest rekurencyjna. Wówczas dla jej określenia wystarczy podać dystrybuantę  $B(t)$  czasu obsługi.

#### 3. Liczba urządzeń obsługi (serwerów)

Najczęściej analizowane są systemy mające jednakowe urządzenia obsługi, które działają jednocześnie. Oznacza to, że w dowolnej chwili czasu  $t$  możliwe jest jednoczesne obsługiwanie nie więcej niż  $n$  zgłoszeń, gdzie  $n$  jest liczbą serwerów.

#### 4. Dyscyplina obsługi

W teorii obsługi masowej analizowane są najczęściej następujące dyscypliny:

- **FIFO** (first in – first out) – zgłoszenia są obsługiwane w kolejności ich przybycia do systemu;
- **LIFO** (last in – first out) – zgłoszenia są obsługiwane w kolejności odwrotnej do porządku ich przybycia tzn. jako pierwsze obsługiwane

jest zgłoszenie, które przybyło do systemu jako ostatnie;

- **PS** (processor sharing) – zgłoszenia znajdujące się w systemie są obsługiwane jednocześnie, natomiast prędkość obsługi każdego z  $k$  zgłoszeń znajdujących się w systemie jest  $k$  razy mniejsza od prędkości obsługi, gdyby w systemie znajdowało się tylko jedno zgłoszenie.

Dla krótkiego oznaczenia dostatecznie prostych systemów obsługi używa się symboliki Kendalla:  $A/B/n/m$ .

Litera  $A$  oznacza typ strumienia wejściowego. Na przykład jeżeli  $A = M$  (Markov), strumień wejściowy jest najprostszy;  $A = D$  – strumień wejściowy deterministyczny;  $A = G$  – strumień wejściowy jest rekurencyjny o dowolnej dystrybucji  $A(t)$  odstępów czasu między sąsiednimi chwilami przybycia zgłoszeń.

Litera  $B$  oznacza typ rozkładu czasu obsługi (uważanej za rekurencyjną) przez urządzenia (uważane za identyczne). Analogicznie  $B = M$  oznacza rozkład wykładniczy czasu obsługi;  $B = D$  – rozkład degeneratywny (czyli czas obsługi zgłoszeń jest stały);  $B = G$  – oznacza obsługę rekurencyjną o dowolnej dystrybucji  $B(t)$  czasu obsługi jednakowej dla wszystkich urządzeń obsługi.

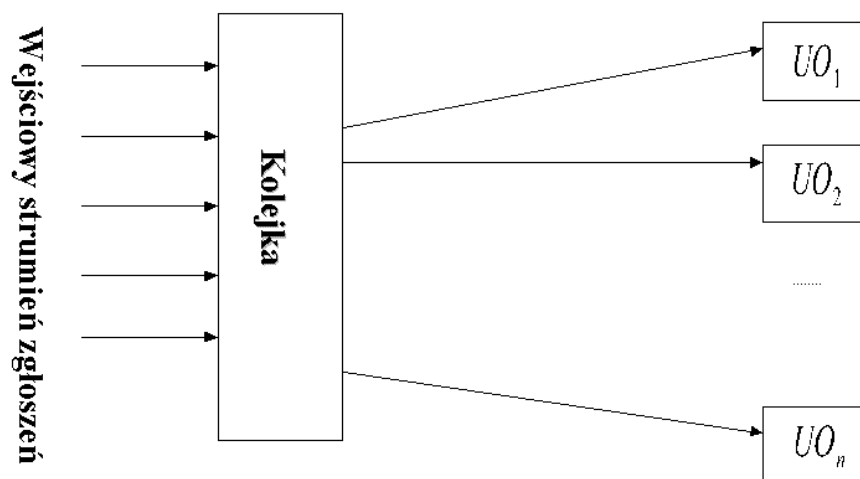
Litera  $n$  oznacza liczbę identycznych urządzeń obsługi (serwerów):  $1 \leq n \leq \infty$ .

Litera  $m$  oznacza liczbę miejsc oczekiwania w kolejce zgłoszeń:  $0 \leq m \leq \infty$ . Przyjmujemy, że kolejka jest wspólna dla wszystkich urządzeń.

W przypadku, gdy  $m = 0$ , zgłoszenia przybywające do systemu w chwilach, gdy w systemie wszystkie urządzenia obsługi są zajęte, ulegają zniszczeniu (utraceniu). Systemy takie nazywamy **systemami z utratą zgłoszeń**. Jeżeli  $m = \infty$  i wszystkie zgłoszenia oczekują „cierpliwie” na początek swojej obsługi, to odpowiadające systemy nazywamy **systemami z oczekiwaniem**, jeżeli dodatkowo wszystkie oczekujące zgłoszenia zostaną całkowicie obsłużone, to odpowiadające systemy nazywamy **systemami bez utrat zgłoszeń**. Jeżeli  $0 < m < \infty$ , to system nazywamy **systemem z ograniczoną liczbą miejsc oczekiwania**.

W przypadku, gdy  $n = 1$ , system nazywamy **jednoliniowym**, natomiast, gdy  $n > 1$ , to system nazywamy **wieloliniowym** ( $n$ -liniowym).

Na rysunku 2.2 przedstawiono ogólny schemat działania klasycznego systemu obsługi.



Rysunek 2.2. Ogólny schemat działania klasycznego systemu obsługi

### 2.3.2 Tryb stacjonarny działania klasycznego systemu obsługi

Rozpatrzmy dostatecznie ogólny system obsługi typu  $G/G/n/m$ , gdzie  $1 \leq n \leq \infty$ ,  $0 \leq m \leq \infty$ . Rozpatrzmy strumień wejściowy posiadający własność stacjonarności (niekoniecznie najprostsz) z intensywnością  $\lambda$ . Intensywność strumienia stacjonarnego wyraża wartość średnią liczby zgłoszeń przybyłych do systemu w ciągu jednostki czasu, wielkość ta istnieje dla dowolnego strumienia stacjonarnego i w ogólnym przypadku wyraża się wzorem:

$$\lambda = \frac{1}{\int_0^{\infty} t dA(t)},$$

$A(t)$  jest dystrybuantą odstępu czasu między sąsiednimi chwilami przybywania zgłoszeń do systemu.

Założmy dodatkowo, że intensywność jest skończona oraz założmy istnienie pierwszego momentu (wartości oczekiwanej)  $\beta_1$  czasu obsługi. Oznaczmy przez  $\eta(t)$  liczbę zgłoszeń znajdujących się w systemie w chwili  $t$ . Proces losowy  $\eta(t)$  jest podstawowym procesem badanym w teorii obsługi masowej. Jednym z najważniejszych zadań teorii obsługi masowej jest wyznaczenie

charakterystyk tego procesu dla konkretnych modeli.

Powstaje naturalne pytanie, jak zachowuje się ten proces przy  $t \rightarrow \infty$ . Jest oczywiste, że w analizowanych warunkach jego przebieg zależy od wielkości  $n, m, \lambda, \beta_1$ .

Jeżeli  $n = \infty$ , to należy się spodziewać, że proces  $\eta(t)$  przy dowolnych warunkach początkowych wraz ze wzrostem czasu będzie dążył do pewnego procesu stacjonarnego  $\eta$ , gdzie  $\eta$  jest zmienną losową, której rozkład nie zależy od  $t$ , tzn.  $\eta(t) \Rightarrow \eta$  przy  $t \rightarrow \infty$  w sensie zbieżności według rozkładu. W przypadku, gdy  $1 \leq n < \infty$ ,  $0 \leq m < \infty$ , część przybywających zgłoszeń będzie utracona. Przy ustalonych  $n$  i  $m$  część ta będzie tym większa, im większe będą wartości  $\lambda$  i  $\beta_1$ . W tym przypadku proces  $\eta(t)$  również wraz ze wzrostem czasu będzie dążył do pewnego procesu stacjonarnego. Najciekawszy jest przypadek, gdy  $1 \leq n < \infty$ ,  $m = \infty$ . W tym przypadku warunki stacjonarne działania systemu obsługi będą istniały w pewnych ściśle określonych warunkach.

Jest oczywiste, że w przypadku tym, z uwagi na fakt, że kolejka zgłoszeń oczekujących na obsługę jest nieograniczona, liczba zgłoszeń obecnych w systemie może wraz ze wzrostem czasu dążyć do nieskończoności. Można udowodnić, że warunki stacjonarne dla takiego systemu istnieją, gdy  $\frac{\lambda\beta_1}{n} < 1$ . Inaczej mówiąc, warunki stacjonarne dla tego systemu będą spełnione, gdy średnia liczba zgłoszeń przybywających do systemu w ciągu jednostki czasu nie przekracza „maksymalnej prędkości obsługi zgłoszeń” tzn. średniej liczby zgłoszeń obsłużonych przez  $n$  niezależnych urządzeń obsługi w ciągu jednostki czasu. Wielkość  $\rho = \frac{\lambda\beta_1}{n}$  w takim przypadku nazywa się **ładowaniem systemowym**.



## Przegląd wybranych klasycznych systemów obsługi

W rozdziale tym zaprezentowany został przegląd metod analizy klasycznych systemów obsługi oraz ich wykorzystanie do badania wybranych modeli. Uwaga jest skupiona na tych klasycznych modelach teorii obsługi masowej, których uogólnienia występują przy analizie systemów informatycznych o stochastycznym charakterze działania. Pokazane jest zatem na początku zastosowanie aparatu łańcuchów Markowa z czasem ciągłym do analizy markowskich systemów  $M/M/n/m$  - czyli metoda równań różniczkowych, a następnie - wykorzystanie metody zmiennej dodatkowej do analizy jednoliniowego systemu  $M/G/1/\infty$  - metoda ta jest wnioskiem z teorii procesów semimarkowskich. Jak widać zatem, pojęciem podstawowym i najważniejszym jest pojęcie łańcucha Markowa, które, jak się okazuje, jest kluczem otwierającym możliwości analizy klasycznych systemów obsługi. Dlatego na początku jest szeroko omówione właśnie to pojęcie a następnie zostało wyjaśnione, dlaczego ten właśnie aparat jest wygodnym narzędziem badania klasycznych systemów obsługi. Przedstawiona tu analiza klasycznych systemów obsługi jest oparta o pozycje [7], [18], [19], [30], [35] oraz [44]. Wiadomości dotyczące teorii łańcuchów Markowa oraz procesów semimarkowskich zaczerpnięto z prac [3], [4], [8] oraz [14], a podstawowe wiadomości dotyczące analizy zespolonej (m.in. metoda residuów) zostały zaczerpnięte z pracy [13].

### 3.1 Łańcuchy Markowa z czasem ciągłym

W punkcie tym wprowadzone zostało pojęcie łańcucha Markowa z czasem ciągłym oraz omówione są jego podstawowe własności. Jak można zauważyć, już z samej definicji wynika, że jest to bardzo wygodne narzędzie analizy klasycznych systemów obsługi ze względu na jego własność „zapominania o przeszłości”. Dzięki temu można wykorzystać aparat łańcuchów Markowa z czasem ciągłym do wypisywania równań różniczkowych dla stanów odpowiednich procesów badanych przy analizie klasycznych systemów obsługi.

Przejdźmy teraz do określenia tego pojęcia.

**Definicja 3.1.** Niech  $T$  będzie zbiorem liczb rzeczywistych nieujemnych oraz niech  $X$  będzie zbiorem skończonym lub przeliczalnym. Proces stochastyczny  $\xi(t)$  przyjmujący wartości ze zbioru  $X$  nazywamy łańcuchem Markowa z czasem ciągłym, jeżeli dla dowolnej liczby naturalnej  $n = 1, 2, \dots$ , dowolnych  $t_1, t_2, \dots, t_n \in T$ , takich, że  $t_1 < t_2 < \dots < t_n$  i dowolnych  $x_0, x_1, \dots, x_n \in X$  zachodzi równość:

$$\begin{aligned} P\{\xi(t_n) = x_n | \xi(t_0) = x_0, \xi(t_1) = x_1, \dots, \xi(t_{n-1}) = x_{n-1}\} = \\ = P\{\xi(t_n) = x_n | \xi(t_{n-1}) = x_{n-1}\}. \end{aligned}$$

Elementy zbioru  $X$  nazywać będziemy stanami procesu  $\xi(t)$ . Równość zawarta w definicji 3.1 oznacza, że przy ustalonym stanie  $x_{n-1}$  procesu  $\xi(t)$  w chwili czasu  $t_{n-1}$  jego dalsze zachowanie nie zależy od tego, w jaki sposób proces trafił do stanu  $x_{n-1}$  (przyszłość przy ustalonej terażniejszości nie zależy od przeszłości). I tutaj właśnie widać zalety stosowania tego aparatu przy analizie szerokiej klasy systemów obsługi. Mianowicie, analizując odpowiednie procesy dla konkretnych modeli nie musimy zastanawiać się nad przebiegiem całego procesu, tylko nad jego przebiegiem od pewnej chwili czasu, to co działo się przed tą chwilą nie wpływa na dalszy przebieg procesu. To ułatwia analizę i prowadzi do dużo prostszych obliczeń. Badanie większości klasycznych systemów obsługi polega na zbudowaniu odpowiedniego łańcucha Markowa z czasem ciągłym (lub jego wielowymiarowego uogólnienia - procesu markowowskiego) i jego dalszym analizowaniu.

W dalszym ciągu będziemy zakładać, że  $X = \{i\}$ ,  $i = 0, 1, \dots$  (na przykład wszystkie stany procesu  $\xi(t)$  są ponumerowane). Typowym przykładem takiego procesu jest używany w teorii obsługi masowej proces  $\eta(t)$ , gdzie  $\eta(t)$  jest liczbą zgłoszeń znajdujących się w systemie w chwili  $t$ .

Wprowadzimy teraz kilka określeń związanych z łańcuchami Markowa z czasem ciągłym.

**Definicja 3.2.** Łańcuch Markowa z czasem ciągłym nazywa się jednorodnym, gdy prawdopodobieństwo warunkowe  $P\{\xi(t + t_0) = j | \xi(t_0) = i\}$  dla ustalonych  $i, j \in X$  oraz dowolnych  $t, t_0 \geq 0$  nie zależy od  $t_0$ .

Nadal analizować będziemy tylko jednorodne łańcuchy Markowa z czasem ciągłym i będziemy je nazywać krótko - łańcuchami Markowa.



Wprowadźmy następujące oznaczenia:

$$P_{ij}(t) = P\{\xi(t) = j | \xi(0) = i\},$$

$$P_j(t) = P\{\xi(t) = j\}.$$

Wielkości te są odpowiednio nazywane prawdopodobieństwem przejścia ze stanu  $i$  do stanu  $j$  w ciągu czasu  $t$  oraz prawdopodobieństwem obecności procesu w stanie  $j$  w chwili czasu  $t$ .

**Definicja 3.3.** Zespół prawdopodobieństw  $\{P_j(0), j \in X\}$  ( $\sum_{j \in X} P_j(0) = 1$ ) nazywamy rozkładem początkowym łańcucha Markowa.

**Definicja 3.4.** Łańcuch Markowa nazywa się ergodycznym, jeżeli dla wszystkich jego stanów  $j \in X$  istnieją granice  $\pi_j = \lim_{t \rightarrow \infty} P_j(t)$  ( $\sum_{j \in X} \pi_j = 1$ ).

Zespół liczb  $\{\pi_j, j \in X\}$  nazywamy rozkładem ergodycznym łańcucha Markowa, natomiast granice  $\pi_j$  nazywamy prawdopodobieństwami końcowymi łańcucha Markowa. Jeżeli dla każdego stanu  $j \in X$  mamy  $\pi_j > 0$ , to łańcuch Markowa nazywamy ostro ergodycznym.

**Definicja 3.5.** Rozkład prawdopodobieństwa  $\{p_j, j \in X\}$  ( $p_j \geq 0, \sum_{j \in X} p_j = 1$ ) nazywa się rozkładem stacjonarnym łańcucha Markowa, jeżeli dla dowolnego  $j \in X$  i dowolnego  $t > 0$  zachodzi równość:

$$p_j = \sum_{i \in X} p_i P_{ij}(t).$$

Jeżeli w dowolnej chwili czasu łańcuch Markowa posiada rozkład stacjonarny, to prawdopodobieństwa obecności procesu w każdym ze stanów ze zbioru  $X$  nie zależą od czasu; taki łańcuch nazywamy **stacjonarnym**. Stąd wynika w szczególności, że rozkład początkowy  $\{P_j(0)\}$  stacjonarnego łańcucha Markowa jest rozkładem stacjonarnym (tj.  $P_j(0) = p_j$  dla dowolnego  $j \in X$ ), ponieważ spełnia on równania z definicji 3.5. Rozkład ergodyczny (o ile istnieje) jest oczywiście stacjonarny, więc jego prawdopodobieństwa końcowe są wyznaczone przez te same równania. Oznacza to, że jeżeli łańcuch Markowa  $\xi(t)$  o rozkładzie początkowym  $\{P_j(0), j \in X\}$  i prawdopodobieństwach przejścia  $P_{ij}(t)$ ,  $i \in X$  jest ergodyczny z prawdopodobieństwami końcowymi  $\pi_j$ ,  $j \in X$ , to istnieje łańcuch Markowa stacjonarny  $\xi_1(t)$  o takich samych prawdopodobieństwach przejścia i rozkładzie początkowym  $P_j^1(0) = \pi_j$ . Wynika stąd, że w pewnym sensie zachowanie ergodycznego łańcucha Markowa o prawdopodobieństwach przejścia  $P_{ij}(t)$  wraz z ciągiem czasu (przy  $t \rightarrow \infty$ ) zbliża się do zachowania łańcucha stacjonarnego  $\xi_1(t)$ .

Mówi się wtedy, że istnieje jedyny rozkład stacjonarny łańcucha  $\xi(t)$ , który odpowiada jego rozkładowi ergodycznemu.

**Definicja 3.6.** Łańcuch Markowa nazywamy nieredukowalnym, jeżeli dla dowolnych jego stanów  $i, j \in X$  istnieje taka liczba skończona  $t_{ij}$  ( $0 < t_{ij} < \infty$ ), że  $P_{ij}(t_{ij}) > 0$ .

Nieredukowalność łańcucha Markowa oznacza praktyczną możliwość przejścia w ciągu skończonego czasu z dowolnego stanu  $i \in X$  do dowolnego stanu  $j \in X$ .

**Definicja 3.7.** Łańcuch Markowa nazywa się ciągłym stochastycznie, gdy dla dowolnych jego stanów  $i, j \in X$  mamy  $\lim_{t \rightarrow 0} P_{ij}(t) = P_{ij}(0) = \delta_{i,j}$ , gdzie  $\delta_{i,j}$  jest symbolem Kroneckera zdefiniowanym następująco:

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

Dla ciągłego stochastycznie łańcucha Markowa można dowieść istnienia granic:

$$\lambda_i = \lim_{\Delta t \rightarrow 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t}; \quad \lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t}, \quad i \neq j,$$

gdzie  $0 \leq \lambda_i \leq \infty$ ,  $0 \leq \lambda_{ij} < \infty$ . Liczba  $\lambda_i$  nazywa się **intensywnością wyjścia ze stanu  $i$** , a  $\lambda_{ij}$  – **intensywnością przejścia ze stanu  $i$  do stanu  $j$** . Jeżeli  $\lambda_i = \infty$ , to stan  $i$  nazywa się momentalnym, w przeciwnym przypadku stan nazywa się niemomentalnym (regulacyjnym). Stan regulacyjny, dla którego  $\lambda_i = 0$  nazywa się pochłaniającym. Można dowieść, że zawsze jest spełniona nierówność  $\lambda_i \geq \sum_{j \neq i} \lambda_{ij}$ . Jeżeli dla regulacyjnego stanu  $i$  zachodzi równość  $\lambda_i = \sum_{j \neq i} \lambda_{ij}$ , to taki stan nazywa się zachowawczym. Jeżeli wszystkie stany łańcucha są zachowawcze, to łańcuch nazywamy zachowawczym.

**Definicja 3.8.** Stan  $j$  ciągłego stochastycznie łańcucha Markowa nazywa się osiągalnym ze stanu  $i$ , jeżeli albo  $i = j$ , albo  $\lambda_{ij} > 0$ , albo istnieje taki ciąg stanów  $i_1, i_2, \dots, i_n \in X$ , że  $\lambda_{ii_1} \lambda_{i_1 i_2} \dots \lambda_{i_n j} > 0$ .

W dalszym ciągu przedstawione są twierdzenia ergodyczne używane w teorii obsługi masowej. Twierdzenia ergodyczne podają warunki w jakich istnieje rozkład stacjonarny łańcucha Markowa odpowiadający jego rozkładowi ergodycznemu. Twierdzenia te są istotne z uwagi na fakt badania systemów obsługi w trybie stacjonarnym. Często w praktyce szukamy właśnie rozkładów stacjonarnych (granicznych) dla analizowanych procesów losowych. Dlatego też warto mieć wiedzę, kiedy takie rozkłady w ogóle istnieją.

Podstawowe twierdzenia dla łańcuchów Markowa wykorzystywane w teorii obsługi masowej są następujące.

**Twierdzenie 3.1.** Jeżeli zbiór stanów  $X$  ciągłego stochastycznie łańcucha Markowa z czasem ciągłym jest skończony i istnieje stan osiągalny z dowolnego jego stanu, to taki łańcuch jest ergodyczny i posiada tylko jeden rozkład stacjonarny odpowiadający rozkładowi ergodycznemu.

**Twierdzenie 3.2.** Zachowawczy łańcuch Markowa z przeliczalnym zbiorem stanów jest ergodyczny, jeżeli jest nieredukowalny i układ równań:

$$x_j = \sum_{i \in X} x_i \frac{\lambda_{ij}}{\lambda_i}, \quad j \in X, \quad \lambda_i > 0$$

posiada rozwiązania nietrywialne  $\{x_i, i \in X\}$  takie, że  $\sum_{i \in X} |x_i| < \infty$ . W tym przypadku istnieje jedyny rozkład stacjonarny łańcucha, który odpowiada jego rozkładowi ergodycznemu, i z dokładnością do czynnika normalizacyjnego zgadza się on z rozwiązaniem  $\{x_i, i \in X\}$ .

## 3.2 Metoda równań różniczkowych. Systemy markowskie

W tym punkcie wprowadzona została metoda równań różniczkowych (metoda  $\Delta t$ ) badania łańcuchów Markowa. W dalszym ciągu za pomocą metody  $\Delta t$  przeprowadzona jest natomiast analiza markowskich systemów obsługi typu  $M/M/n/m$ . Zaprezentowana jest też, wynikająca z metody  $\Delta t$ , metoda przekrojów badania markowskich systemów obsługi w trybie stacjonarnym.

### 3.2.1 Metoda równań różniczkowych

Przedstawimy teraz tzw. proste równania Kołmogorowa dla prawdopodobieństw obecności łańcucha Markowa w stanie  $j$  w chwili czasu  $t$ . Metoda analizy łańcuchów Markowa w oparciu o te równania nazywa się metodą równań różniczkowych albo inaczej metodą  $\Delta t$ . Metodę tą wykorzystuje się do analizy klasycznych markowskich systemów obsługi  $M/M/n/m$ , gdyż jak zobaczymy później, proces  $\eta(t)$  dla takich systemów jest łańcuchem Markowa ze skończonym lub przeliczalnym zbiorem stanów.

Można wykazać, że dla ciągłego stochastycznie łańcucha Markowa ze skończonym zbiorem stanów  $X$  zachodzą następujące równania (nazywane równaniami prostymi Kołmogorowa):

$$P'_j(t) = -\lambda_j P_j(t) + \sum_{k \neq j} \lambda_{kj} P_k(t), \quad j \in X \quad (3.1)$$

z warunkiem początkowym  $P_j(0) = p_j^{(0)}$ , gdzie  $\{p_j^{(0)}, j \in X\}$  jest rozkładem początkowym badanego łańcucha.

Aby wprowadzone równania były prawdziwe także w przypadku przeliczalnego zbioru stanów  $X$ , wystarczy, aby intensywności  $\lambda_i, i \in X$  były wielkościami ograniczonymi tzn. istnieje taka liczba  $C > 0$ , że  $\lambda_i \leq C$  dla każdego  $i \in X$ .

### 3.2.2 Równania równowagi

Jeżeli dla danego łańcucha Markowa spełnione są warunki stacjonarne (inaczej mówiąc łańcuch Markowa jest ergodyczny), to z równań (3.1) poprzez odpowiednie przejścia graniczne przy  $t \rightarrow \infty$  uzyskujemy algebraiczne równania, za pomocą których możemy wyznaczyć prawdopodobieństwa końcowe  $p_j = \lim_{t \rightarrow \infty} P_j(t)$  analizowanego łańcucha Markowa. Otrzymane w taki sposób równania nazywamy **równaniami równowagi**.

Ponieważ prawdopodobieństwa końcowe nie zależą od czasu i są wielkościami stałymi, to przy  $t \rightarrow \infty$  pochodna  $P'_j(t)$  dąży do zera. Wówczas z równań (3.1) otrzymujemy następujące równania:

$$-\lambda_j p_j + \sum_{k \neq j} \lambda_{kj} p_k = 0, \quad j \in X,$$

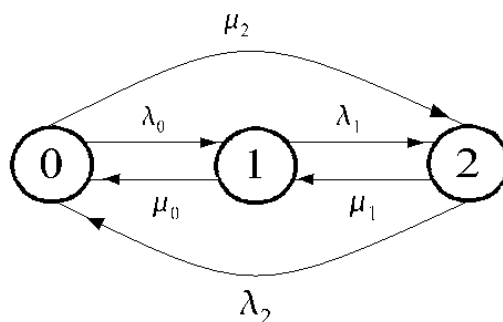
czyli

$$\lambda_j p_j = \sum_{k \neq j} \lambda_{kj} p_k, \quad j \in X. \quad (3.2)$$

Równania równowagi (3.2) mają następujący sens probabilistyczny. Nazwijmy iloczyn  $\lambda_j p_j$  strumieniem prawdopodobieństwa ze stanu  $j$ . Iloczyn  $\lambda_{kj} p_k$  nazwijmy strumieniem prawdopodobieństwa ze stanu  $k$  do stanu  $j$ . Wówczas ze wzoru (3.2) wynika, że w warunkach stacjonarnych strumień prawdopodobieństwa ze stanu dowolnego jest zrównoważony sumą strumieni prawdopodobieństw ze wszystkich innych stanów do tego stanu.

Wprowadzone równania Kołmogorowa dla prawdopodobieństw bezwarunkowych oraz równania równowagi w dość łatwy sposób można wypisywać korzystając z interpretacji łańcuchów Markowa w postaci grafów stochastycznych. Rozpatrzmy następujący przykład.

**Przykład 3.1.** Dany jest łańcuch Markowa o trzech stanach, którego graf jest przedstawiony poniżej.



**Rysunek 3.1.** Graf stochastyczny przykładowego trzystanowego łańcucha Markowa

W analizowanym łańcuchu Markowa intensywność wyjścia z dowolnego stanu jest sumą intensywności przejścia z tego stanu do innych stanów (tak jest w przypadku łańcuchów Markowa ze skończonym zbiorem stanów - są one bowiem zachowawcze). Zatem otrzymujemy następujące równania na funkcje  $P_j(t)$ :

$$P_0'(t) = -(\lambda_0 + \mu_2)P_0(t) + \mu_0P_1(t) + \lambda_2P_2(t),$$

$$P_1'(t) = -(\lambda_1 + \mu_0)P_1(t) + \lambda_0P_0(t) + \mu_1P_2(t),$$

$$P_2'(t) = -(\lambda_2 + \mu_1)P_2(t) + \mu_2P_0(t) + \lambda_1P_1(t).$$

Patrząc na postać tych równań zauważamy, że wypisuje się je automatycznie korzystając z grafu stochastycznego. Dokonując odpowiednich przejść granicznych (lub korzystając z interpretacji równań równowagi) otrzymujemy natomiast następujące równania:

$$(\lambda_0 + \mu_2)p_0 = \mu_0p_1 + \lambda_2p_2,$$

$$(\lambda_1 + \mu_0)p_1 = \lambda_0p_0 + \mu_1p_2,$$

$$(\lambda_2 + \mu_1)p_2 = \mu_2p_0 + \lambda_1p_1.$$

Oczywiście zarówno równania dla prawdopodobieństw bezwarunkowych  $P_j(t)$ , jak i równania równowagi mogą być rozwiązane w oparciu o pewne algebraiczne metody rozwiązywania odpowiednio układów równań różniczkowych (np. w oparciu o przekształcenia Laplace'a) oraz układów równań algebraicznych, przy czym oczywiście dokładamy warunek normalizacyjny –

odpowiednio  $\sum_{j \in X} P_j(t) = 1$  lub  $\sum_{j \in X} p_j = 1$ .

Przejdziemy teraz do omówienia szczególnej klasy łańcuchów Markowa, a mianowicie procesów narodzin i śmierci.

### 3.2.3 Procesy narodzin i śmierci

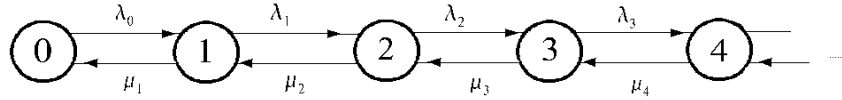
**Definicja 3.9.** Procesem narodzin i śmierci nazywa się łańcuch Markowa  $\xi(t)$  z przeliczalnym zbiorem stanów  $X = \{0, 1, \dots\}$ , dla którego ze stanu  $n$  ( $n \geq 1$ ) możliwe jest przejście bezpośrednie tylko do stanów  $n - 1$  i  $n + 1$ , a ze stanu 0 tylko do stanu 1 (dla takich procesów  $\lambda_{ij} = 0$ , jeżeli  $|i - j| \geq 2$ ).

Z definicji 3.9 wynika, że w procesach narodzin i śmierci możliwa jest tylko zmiana numeru stanu „o jeden” (z wyjątkiem stanu 0, gdzie mamy tylko możliwe przejście  $0 \rightarrow 1$ ). Stąd też wynika ich nazwa, przejście  $n \rightarrow n + 1$  można interpretować jako narodziny, a przejście  $n \rightarrow n - 1$  jako śmierć osobnika w populacji złożonej z  $n$  osobników.

Procesy narodzin i śmierci odgrywają ważną rolę w teorii obsługi masowej. Na przykład w modelach systemów typu  $M/M/n/m$  o przeliczalnej liczbie stanów proces  $\eta(t)$  jest procesem narodzin i śmierci. Graf stochastyczny dla procesów narodzin i śmierci jest oczywiście grafem liniowym.

Założmy, że analizowany proces narodzin i śmierci jest ciągły stochastycznie, a wszystkie jego stany są zachowawcze. Przyjmijmy dla prostoty następujące oznaczenia:  $\lambda_{nn+1} = \lambda_n$ ,  $\lambda_{nn-1} = \mu_n$ .

Poglądowo proces ten można przedstawić w postaci następującego grafu liniowego (rys. 3.2).



**Rysunek 3.2.** Graf stochastyczny procesu narodzin i śmierci

Można dowieść, że w przypadku analizowanego procesu narodzin i śmierci równania Kołmogorowa (3.1) dla prawdopodobieństw bezwarunkowych  $P_j(t)$  są spełnione w przypadku, gdy równość  $P_{kj}(\Delta t) = o(\Delta t)$  jest spełniona dla wszystkich  $k, j$  takich, że  $|k - j| \geq 2$ . Przyjmują one wówczas postać:

$$P'_j(t) = \lambda_{j-1}P_{j-1}(t) - (\lambda_j + \mu_j)P_j(t) + \mu_{j+1}P_{j+1}(t),$$

gdzie  $i, j = 0, 1, \dots$ ;  $\mu_0 \equiv 0$ ,  $\lambda_{-1} \equiv 0$ .

W przypadku, gdy wielkości  $\lambda_k$  wraz ze wzrostem  $k$  szybko wzrastają, w ciągu skończonego czasu proces może z dodatnim prawdopodobieństwem wyjść z przestrzeni stanów  $X$  tzn. osobników w populacji może być nieskończenie wiele, co oznacza, że warunek normalizacyjny  $\sum_{k=0}^{\infty} P_k(t) = 1$  nie będzie spełniony. Aby warunek ten był spełniony wystarcza, aby szereg  $\sum_{k=1}^{\infty} \prod_{i=1}^k \frac{\mu_i}{\lambda_i}$  był rozbieżny. Jeżeli dodatkowo spełniony jest warunek  $\sum_{k=1}^{\infty} \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i} < \infty$ , to proces  $\xi(t)$  jest ergodyczny i istnieją jego prawdopodobieństwa końcowe, które możemy wyznaczyć np. na podstawie równań równowagi. Oba omawiane warunki są spełnione, gdy istnieją stan  $N \in X$  i liczba  $\gamma$  ( $0 < \gamma < 1$ ) takie, że dla wszystkich stanów  $i \geq N$  spełniona jest nierówność:

$$\frac{\lambda_i}{\mu_{i+1}} \leq \gamma. \quad (3.3)$$

Układ równań równowagi przyjmuje w tym przypadku postać:

$$\begin{cases} 0 = -\lambda_0 p_0 + \mu_1 p_1; \\ 0 = \lambda_{j-1} p_{j-1} - (\lambda_j + \mu_j) p_j + \mu_{j+1} p_{j+1}, \quad j = 1, 2, \dots \end{cases} \quad (3.4)$$

Jeżeli oznaczymy:  $z_k = -\lambda_k p_k + \mu_{k+1} p_{k+1}$ ;  $k = 0, 1, \dots$ , to wówczas równania (3.4) przyjmują postać:

$$z_0 = 0; \quad z_j - z_{j-1} = 0; \quad j = 1, 2, \dots$$

Stąd wynika, że  $z_k = 0$  dla wszystkich  $k = 0, 1, \dots$ , a więc

$$p_k = \frac{\lambda_{k-1}}{\mu_k} p_{k-1} = \frac{\lambda_{k-1}}{\mu_k} \cdot \frac{\lambda_{k-2}}{\mu_{k-1}} \cdot \dots \cdot \frac{\lambda_0}{\mu_1} p_0,$$

czyli

$$p_k = \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i} p_0. \quad (3.5)$$

Z warunku normalizacyjnego otrzymujemy natomiast

$$p_0 = \left[ 1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i} \right]^{-1}.$$

**Uwaga.** W przypadku procesów narodzin i śmierci wypisywanie równań równowagi upraszcza się do następującej metody zwanej **metodą przekrojów**. Między dowolnymi dwoma stanami rysujemy przerywaną linię. W procesach narodzin i śmierci iloczyn prawdopodobieństwa i intensywności z lewej strony tej linii (przekroju) jest równy analogicznemu iloczynowi z prawej strony przekroju. Układ równań wynikający z tej metody ma następującą postać:

$$\lambda_{n-1} p_{n-1} = \mu_n p_n; \quad n = 1, 2, \dots \quad (3.6)$$

i jest równoważny układowi (3.4).

W dalszym ciągu pokażemy, jak stosuje się poznane metody do analizy systemów markowowskich.

### 3.2.4 Markowowskie systemy obsługi $M/M/n/m$

Niech  $\eta(t)$  oznacza liczbę zgłoszeń znajdujących się w systemie obsługi w chwili  $t$ . Jeżeli  $\eta(t)$  jest łańcuchem Markowa, to odpowiadający system nazywa się markowowskim systemem obsługi. Można wykazać, że markowowskie są wszystkie systemy obsługi typu  $M/M/n/m$ , gdzie  $1 \leq n \leq \infty$ ,  $0 \leq m \leq \infty$ .

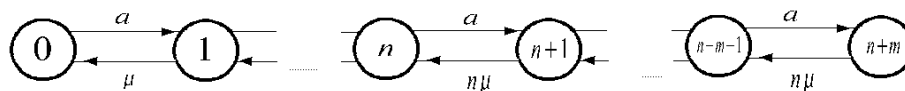
W kolejnych rozważaniach przyjmujemy następujące oznaczenia:  $a$  jest parametrem (intensywnością) wejściowego strumienia zgłoszeń;  $\mu$  jest parametrem rozkładu czasu obsługi;  $\rho = \frac{a}{n\mu}$  (jest to ładowanie systemowe w przypadku, gdy  $n < \infty$ ,  $m = \infty$ ). Dla systemów o nieskończonej liczbie urządzeń ( $n = \infty$ ) będziemy używać oznaczenia  $y = \frac{a}{\mu}$ .

Dalej skupimy się na wyprowadzeniu równań dla prawdopodobieństw bezwarunkowych i stacjonarnych dla kilku szczególnych przypadków systemów  $M/M/n/m$ .



1. **SYSTEM OBSŁUGI**  $M/M/n/m$ ,  $1 \leq n < \infty$ ,  $0 \leq m < \infty$ .

Dla systemu  $M/M/n/m$ ,  $1 \leq n < \infty$ ,  $0 \leq m < \infty$  zbiór stanów jest skończony. Jeżeli  $\rho = \frac{a}{n\mu} < \infty$ , to na mocy twierdzenia 3.1 spełnione są warunki istnienia trybu stacjonarnego dla tego systemu. Proces  $\eta(t)$  nie jest w tym przypadku procesem narodzin i śmierci. Zbudujemy proces  $\eta_1(t)$  równoważny procesowi  $\eta(t)$ . Będziemy uważać, że proces  $\eta_1(t)$  ma nieskończoną liczbę stanów, intensywności przejścia do stanu  $n+m$  włącznie pokrywają się z odpowiednimi intensywnościami procesu  $\eta(t)$ , a począwszy od stanu  $n+m+1$  intensywności przejścia są zerowe. Proces  $\eta_1(t)$  jest już procesem narodzin i śmierci, więc możemy w oparciu o metodę przekrojów wypisać równania równowagi i znaleźć prawdopodobieństwa stacjonarne  $p_k$ . Aby to zrobić narysujemy graf przejść dla tego systemu (rys. 3.3).



**Rysunek 3.3.** Graf stochastyczny dla systemu obsługi  $M/M/n/m$ ,  $1 \leq n < \infty$ ,  $0 \leq m < \infty$

Równania Kołmogorowa dla prawdopodobieństw bezwarunkowych  $P_j(t)$  wyglądają następująco:

$$\begin{aligned} P'_0(t) &= -aP_0(t) + \mu P_1(t); \\ P'_k(t) &= aP_{k-1}(t) - (a + k\mu)P_k(t) + (k+1)\mu P_{k+1}(t), \quad 0 < k < n; \\ P'_k(t) &= aP_{k-1}(t) - (a + n\mu)P_k(t) + n\mu P_{k+1}(t), \quad n \leq k < n+m; \\ P'_{n+m}(t) &= aP_{n+m-1}(t) - n\mu P_{n+m}(t). \end{aligned} \quad (3.7)$$

Korzystając natomiast z metody przekrojów otrzymujemy następujące równania równowagi:

$$\begin{cases} ap_{k-1} = k\mu p_k, & \text{gdy } 1 \leq k \leq n; \\ ap_{k-1} = n\mu p_k, & \text{gdy } n \leq k \leq n+m. \end{cases}$$

Rozwiązując powyższy układ otrzymujemy wyrażenia na obliczanie prawdopodobieństw stacjonarnych obecności  $k$  zgłoszeń w systemie:

$$p_k = \begin{cases} \frac{(n\rho)^k}{k!} p_0, & \text{gdy } 0 \leq k \leq n; \\ \frac{n^n \rho^k}{n!} p_0, & \text{gdy } n < k \leq n+m, \end{cases} \quad (3.8)$$

gdzie na mocy warunku normalizacyjnego mamy:

$$p_0 = \left[ \sum_{k=0}^n \frac{(n\rho)^k}{k!} + \frac{n^n \rho^{n+1} (1 - \rho^m)}{n!(1 - \rho)} \right]^{-1}.$$

Zauważmy, że prawdopodobieństwo  $p_{n+m}$  odgrywa rolę prawdopodobieństwa utraty zgłoszenia.

**Uwaga.** Dokonując przejścia granicznego przy  $n \rightarrow \infty$  (w takim przypadku liczba miejsc oczekiwania nie wpływa na charakterystyki procesu  $\eta(t)$ ) otrzymujemy stacjonarny rozkład liczby zgłoszeń w systemie  $M/M/\infty$ :

$$p_k = \frac{y^k}{k!} e^{-y}, \quad k = 0, 1, \dots$$

## 2. SYSTEM OBSŁUGI $M/M/n/\infty$ , $1 \leq n < \infty$ .

Dla tego systemu graf będzie bardzo podobny do grafu z poprzedniego systemu z tym, że w danym przypadku zbiór stanów jest nieskończony. Proces  $\eta(t)$  jest więc procesem narodzin i śmierci. Warunki stacjonarne są spełnione, gdy  $\rho < 1$  (wynika to ze wzoru (3.3)). Równania Kołmogorowa dla prawdopodobieństw bezwarunkowych są bardzo podobne do równań (3.7). Równania równowagi mają postać:

$$\begin{cases} ap_{k-1} = k\mu p_k, & \text{gdy } 1 \leq k \leq n; \\ ap_{k-1} = n\mu p_k, & \text{gdy } k \geq n+1. \end{cases}$$

Rozwiązanie tego układu ma postać podobną do wzoru (3.8):

$$p_k = \begin{cases} \frac{(n\rho)^k}{k!} p_0, & \text{gdy } 0 \leq k \leq n; \\ \frac{n^n \rho^k}{n!} p_0, & \text{gdy } k \geq n+1, \end{cases} \quad (3.9)$$

gdzie na mocy warunku normalizacyjnego oraz wzoru na sumę szeregu geometrycznego zbieżnego mamy:

$$p_0 = \left[ \sum_{k=0}^n \frac{(n\rho)^k}{k!} + \frac{n^n \rho^{n+1}}{n!(1 - \rho)} \right]^{-1}.$$

W przypadku  $n = 1$  (system  $M/M/1/\infty$ ) wzór (3.9) upraszcza się do następującego wzoru:

$$p_k = (1 - \rho)\rho^k, \quad k = 0, 1, \dots \quad (3.10)$$

Ze wzoru (3.10) wynika, że dla systemu  $M/M/1/\infty$  stacjonarny rozkład liczby zgłoszeń jest rozkładem geometrycznym z parametrem  $\rho$ .

### 3.3 Jednoliniowy system obsługi $M/G/1/\infty$

W punkcie tym przedstawiona jest analiza klasycznego systemu  $M/G/1/\infty$ , który jest często wykorzystywany w praktyce do modelowania realnych systemów informatycznych. Aby uzyskać efektywne wyniki dotyczące wyznaczenia charakterystyk tego systemu należy najpierw w pewnym sensie uogólnić wprowadzony w punkcie 3.1 aparat łańcuchów Markowa. Wprowadzone zostanie zatem pojęcie procesu semimarkowskiego.

Jest jasne, że w ogólnym przypadku dla systemu  $M/G/1/\infty$  proces  $\eta(t)$  nie jest łańcuchem Markowa, natomiast występuje tu pewna markowska składowa, mianowicie czas między sąsiednimi chwilami przybycia zgłoszeń spełnia rozkład wykładniczy, podobnie jest w przypadku odstępu czasu od dowolnej chwili zakończenia obsługi zgłoszenia do chwili przybycia kolejnego zgłoszenia, co wynika z własności braku pamięci rozkładu wykładniczego. Z tego powodu można uogólnić metodę równań różniczkowych wprowadzając pewien dwuwymiarowy proces markowski, za pomocą analizy którego można uzyskać formuły pozwalające wyznaczać charakterystyki stacjonarne liczby zgłoszeń dla tego systemu. Takie postępowanie nosi nazwę metody zmiennej dodatkowej, a jej uogólnienia będą nam służyć przy analizie kolejnych systemów.

#### 3.3.1 Procesy semimarkowskie

**Definicja 3.10.** Niech  $X = \{0, \dots, N\}$  lub  $X = \{0, 1, \dots\}$ ,  $T = [0, \infty)$ . Ciąg wektorów losowych  $\{(\xi_n, T_n)\}$ ,  $n = 0, 1, \dots$ , gdzie  $\xi_n$  przyjmują wartości ze zbioru  $X$ , a  $T_n$  – ze zbioru  $T$ , nazywamy ciągiem semimarkowskim, gdy zachodzi równość:

$$\begin{aligned} P\{\xi_n = j, T_n < t | \xi_0 = i_0, T_0 < t_0; \xi_1 = i_1, T_1 < t_1; \dots; \xi_{n-1} = i, T_{n-1} < t_{n-1}\} = \\ = P\{\xi_n = j, T_n < t | \xi_{n-1} = i\} = Q_{ij}(t) \end{aligned} \quad (3.11)$$

dla dowolnych  $n$  naturalnych, dowolnych  $t_0, \dots, t_{n-1} \in T$  i dowolnych  $i, j, i_0, \dots, i_{n-2} \in X$ .

Funkcja  $Q_{ij}(t)$  nazywa się funkcją przejściową ciągu  $\{(\xi_n, T_n)\}$ .

Zachodzi następujące twierdzenie.

**Twierdzenie 3.3.** Ciąg losowy  $\{(\xi_n, T_n), n = 0, 1, \dots\}$  spełnia własność (3.11) wtedy i tylko wtedy, gdy są spełnione poniższe równości:

$$\begin{aligned} P\{\xi_n = j | T_n < t_n; \xi_{n-1} = i, T_{n-1} < t_{n-1}; \dots; \xi_0 = i_0, T_0 < t_0\} = \\ = P\{\xi_n = j | T_n < t_n, \xi_{n-1} = i\}, \end{aligned} \quad (3.12)$$

$$\begin{aligned} P\{T_n < t_n | \xi_{n-1} = i, T_{n-1} < t_{n-1}; \xi_{n-2} = i_{n-2}, T_{n-2} < t_{n-2}; \dots; \xi_0 = i_0, T_0 < t_0\} = \\ = P\{T_n < t_n | \xi_{n-1} = i\}. \end{aligned} \quad (3.13)$$

Na podstawie ciągu semimarkowskiego  $\{(\xi_n, T_n)\}$  możemy zbudować proces stochastyczny  $\{\xi(t), t \in T\}$  w sposób następujący.

Założmy, że  $P\{T_0 = 0\} = 1$ . Będziemy zakładać, że  $\xi(t) = \xi_0$ , gdy  $0 \leq t < T_1$ ;  $\xi(t) = \xi_1$ , gdy  $T_1 \leq t < T_1 + T_2$ ;  $\xi(t) = \xi_2$ , gdy  $T_1 + T_2 \leq t < T_1 + T_2 + T_3$  itd. Będziemy zakładać również, że w dowolnym skończonym przedziale czasu  $[0, t]$  z prawdopodobieństwem jeden występuje skończona liczba skoków procesu  $\xi(t)$ . Otrzymany w ten sposób proces  $\{\xi(t), t \geq 0\}$  nazywa się **procesem semimarkowskim** zbudowanym według ciągu semimarkowskiego  $\{(\xi_n, T_n)\}$ .

Omówimy teraz na czym polega metoda zmiennej dodatkowej analizy procesu semimarkowskiego  $\xi(t)$ . Wprowadźmy następujące oznaczenia:

$$\nu(t) = \min \left( n \mid \sum_{i=1}^n T_i \geq t \right),$$

$$\xi_*(t) = T_1 + T_2 + \dots + T_{\nu(t)} - t,$$

gdzie  $\nu(t)$  oznacza liczbę zmiany stanów (skoków) procesu semimarkowskiego  $\xi(t)$  do pierwszej po chwili czasu  $t$  zmiany stanu. Proces  $\xi_*(t)$  nazywa się **cyklem**. Chwile czasu  $0, T_1, T_1 + T_2, \dots$  nazwiemy **0-momentami**. Wówczas cykl  $\xi_*(t)$  jest długością przedziału czasowego od chwili  $t$  do następnego po tej chwili 0-momentu.

Wprowadźmy teraz dwuwymiarowy proces markowski  $(\xi(t), \xi_*(t))$ . Proces ten nazywamy **procesem uliniowanym**. Analiza systemu obsługi za pomocą takiego procesu nosi nazwę **metody zmiennej dodatkowej**.

**Uwaga.** Proces markowski  $(\xi(t), \xi_*(t))$ , gdzie  $\xi_*(t)$  jest odstępem czasu od ostatniego przed chwilą  $t$  0-momentu do chwili  $t$  także będziemy nazywać uliniowanym.

### 3.3.2 Analiza jednoliniowego systemu obsługi $M/G/1/\infty$

Przejdziemy teraz do wyznaczenia charakterystyk stacjonarnych liczby zgłoszeń w systemie  $M/G/1/\infty$  na podstawie analizy pewnego procesu uliniowanego.

Wprowadzimy najpierw bardzo użyteczne pojęcie **intensywności obsługi rekurencyjnej**.

Niech  $P(t_0, t_0 + \Delta t)$  oznacza prawdopodobieństwo zakończenia obsługi w przedziale  $[t_0, t_0 + \Delta t)$ , jeśli wiadomo, że obsługa trwała w chwili czasu  $t_0$  i jej długość przed tą chwilą była równa  $t_0$  ( $\xi \geq t_0$ ).

**Definicja 3.11.** Intensywnością obsługi w punkcie  $t_0$  nazywamy liczbę:

$$\mu(t_0) = \lim_{\Delta t \rightarrow 0} \frac{P(t_0, t_0 + \Delta t)}{\Delta t}.$$

Zauważmy, że na mocy definicji 3.11 i wzoru na prawdopodobieństwo warunkowe zachodzi następujący związek:

$$\begin{aligned} \mu(t_0) &= \lim_{\Delta t \rightarrow 0} \frac{P(t_0, t_0 + \Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P\{t_0 \leq \xi < t_0 + \Delta t\}}{\Delta t P\{\xi \geq t_0\}} = \\ &= \lim_{\Delta t \rightarrow 0} \frac{B(t_0 + \Delta t) - B(t_0)}{\Delta t(1 - B(t_0))} = \frac{b(t_0)}{1 - B(t_0)}, \end{aligned}$$

gdzie  $b(t)$  oznacza uogólnioną gęstość zmiennej losowej o dystrybucancie  $B(t)$  (dla zmiennych losowych absolutnie ciągłych jest to gęstość w sensie zwykłym, natomiast dla zmiennych innego typu gęstość uogólnioną wprowadzamy w oparciu o dystrybucję Diraca [51]).

Z definicji 3.11 wynika, że intensywność obsługi jest określona dla dowolnego nieujemnego  $t$  oraz zachodzi równość:

$$P(t, t + \Delta t) = \mu(t)\Delta t + o(\Delta t).$$

Stosując metodę zmiennej dodatkowej, znajdziemy rozkład stacjonarny  $\{p_k, k = 0, 1, \dots\}$  liczby zgłoszeń w systemie  $M/G/1/\infty$  w dowolnej chwili czasu. W tym celu będziemy analizować markowowski proces uliniowany  $(\eta(t), \xi^*(t))$ , gdzie  $\xi^*(t)$  jest długością przedziału czasowego od początku obsługi zgłoszenia obsługiwanego w chwili  $t$  do chwili  $t$  (można łatwo wykazać, że zastąpienie analizy procesu uliniowanego  $(\xi(t), \xi^*(t))$  procesem  $(\eta(t), \xi^*(t))$  nie łamie własności markowowości ze względu na własność braku pamięci rozkładu wykładniczego).

Wprowadźmy następujące oznaczenia:

$$\begin{aligned} P_n(t) &= P\{\eta(t) = n\}, n = 0, 1, \dots; \\ P_n(x, t)dx &= P\{\eta(t) = n, \xi^*(t) \in [x, x + dx)\}, n = 1, 2, \dots \end{aligned}$$

Korzystając z teorii procesów semimarkowowskich można łatwo dowieść, że w przypadku  $\rho = a\beta_1 < 1$  przy  $t \rightarrow \infty$  istnieją granice  $P_n(x, t) \rightarrow p_n(x)$ ,  $P_n(t) \rightarrow p_n$ . FT liczby zgłoszeń znajdujących się w systemie w trybie

stacjonarnym określimy jako  $P(z) = \sum_{n=0}^{\infty} p_n z^n$ . Naszym zadaniem będzie wyznaczenie FT  $P(z)$ .

Jest oczywiste, że przy  $n \geq 1$  zachodzi równość:

$$p_n = \int_0^{\infty} p_n(x) dx. \quad (3.14)$$

Wprowadźmy jeszcze jedną FT:

$$p(z, x) = \sum_{n=1}^{\infty} p_n(x) z^n. \quad (3.15)$$

Wówczas ze wzoru (3.14) otrzymujemy równość:

$$P(z) = p_0 + \int_0^{\infty} p(z, x) dx. \quad (3.16)$$

Znajdziemy funkcję  $p(z, x)$ . Wprowadźmy oznaczenie

$$\mu(x) = b(x)[1 - B(x)]^{-1},$$

gdzie  $b(x)$  jest uogólnioną gęstością czasu obsługi. Funkcja  $\mu(x)$  jest intensywnością obsługi wprowadzoną w def. 3.11. Wyznamy prawdopodobieństwo  $P_n(x + \Delta t, t + \Delta t) dx$ . Warunkiem koniecznym i wystarczającym na to, aby w systemie w chwili  $t + \Delta t$  było  $n$  zgłoszeń oraz w tej chwili była spełniona relacja  $\xi^*(t + \Delta t) \in [x + \Delta t, x + \Delta t + dx)$  jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie było  $n$  zgłoszeń,  $\xi^*(t) \in [x, x + dx)$ , w ciągu czasu  $\Delta t$  nie była zakończona obsługa zgłoszenia obsługiwanego w chwili  $t$  oraz nie przybywały do systemu inne zgłoszenia. Prawdopodobieństwo tego zdarzenia wynosi  $P_n(x, t) dx [1 - (a + \mu(x)) \Delta t] + o(\Delta t) dx$ ;

2) albo w przypadku  $n \geq 2$  w systemie w chwili  $t$  było  $n - 1$  zgłoszeń,  $\xi^*(t) \in [x, x + dx)$  i w czasie  $\Delta t$  po chwili  $t$  do systemu przybyło zgłoszenie. Prawdopodobieństwo tego zdarzenia wynosi

$$(1 - \delta_{n,1}) a P_{n-1}(x, t) dx \Delta t + o(\Delta t) dx,$$

gdzie  $\delta_{n,1}$  jest standardowo określonym symbolem Kroneckera. Ponieważ prawdopodobieństwa innych zdarzeń, które doprowadzają system do badanego stanu są równe  $o(\Delta t) dx$ , ostatecznie otrzymujemy równanie:

$$\begin{aligned} P_n(x + \Delta t, t + \Delta t) &= P_n(x, t) [1 - (a + \mu(x)) \Delta t] + \\ &+ (1 - \delta_{n,1}) a P_{n-1}(x, t) \Delta t + o(\Delta t). \end{aligned} \quad (3.17)$$

Znajdziemy teraz prawdopodobieństwo  $P_0(t + \Delta t)$ . Aby w chwili  $t + \Delta t$  w systemie nie było ani jednego zgłoszenia, konieczne i wystarczające jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie nie było zgłoszeń i w ciągu czasu  $\Delta t$  zgłoszenia do systemu nie przybywały. Prawdopodobieństwo tego zdarzenia wynosi  $P_0(t)(1 - a\Delta t) + o(\Delta t)$ ;

2) albo w chwili  $t$  w systemie znajdowało się jedno zgłoszenie, którego obsługa została zakończona w czasie  $\Delta t$ . Prawdopodobieństwo tego zdarzenia jest równe  $\int_0^t P_1(x, t)\mu(x)dx\Delta t + o(\Delta t)$ .

Prawdopodobieństwa pozostałych zdarzeń doprowadzających system do tego stanu są równe  $o(\Delta t)$ . Stąd otrzymujemy równanie:

$$P_0(t + \Delta t) = P_0(t)(1 - a\Delta t) + \int_0^t P_1(x, t)\mu(x)dx\Delta t + o(\Delta t). \quad (3.18)$$

Prawdopodobieństwo tego, że w chwili  $t + \Delta t$  w systemie znajdowało się  $n$  zgłoszeń i przy tym  $\xi^*(t + \Delta t) < \Delta t$ , jest równe  $\int_0^{\Delta t} P_n(u, t + \Delta t)du$ . Z drugiej strony system może wpaść do tego stanu, jeżeli w chwili  $t$  było w nim  $n + 1$  zgłoszeń i w czasie  $\Delta t$  jedno zgłoszenie zostało obsłużone. Prawdopodobieństwo takiego zdarzenia jest równe  $\int_0^t P_{n+1}(x, t)\mu(x)dx\Delta t + o(\Delta t)$ . Należy jednak pamiętać, że w przypadku  $n = 1$  system może wpaść do oznaczonego stanu, jeśli w chwili  $t$  nie było w nim żadnego zgłoszenia i w czasie  $\Delta t$  przybyło zgłoszenie. Prawdopodobieństwo tego zdarzenia wynosi  $P_0(t)a\Delta t + o(\Delta t)$ . Ostatecznie otrzymujemy równanie:

$$\int_0^{\Delta t} P_n(u, t + \Delta t)du = \Delta t \int_0^t P_{n+1}(x, t)\mu(x)dx + \delta_{n,1}P_0(t)a\Delta t + o(\Delta t). \quad (3.19)$$

Z równań (3.17)–(3.19) przy  $\Delta t \rightarrow 0$  wynika następujący układ równań:

$$\frac{\partial P_n(x, t)}{\partial t} + \frac{\partial P_n(x, t)}{\partial x} = -(a + \mu(x))P_n(x, t) + (1 - \delta_{n,1})aP_{n-1}(x, t), \quad n = 1, 2, \dots; \quad (3.20)$$

$$\frac{\partial P_0(t)}{\partial t} = -aP_0(t) + \int_0^t P_1(x, t)\mu(x)dx; \quad (3.21)$$

$$P_n(0^+, t) = \int_0^t P_{n+1}(x, t)\mu(x)dx + \delta_{n,1}aP_0(t), n = 1, 2, \dots \quad (3.22)$$

Ponieważ przy  $\rho < 1$  istnieją granice  $P_0(t) \rightarrow p_0$  oraz  $P_n(x, t) \rightarrow p_n(x)$  przy  $t \rightarrow \infty$ ,  $n = 1, 2, \dots$ , to z równań (3.20)–(3.22) wynika następujący układ dla prawdopodobieństw stacjonarnych:

$$\frac{\partial p_n(x)}{\partial x} = -(a + \mu(x))p_n(x) + (1 - \delta_{n,1})ap_{n-1}(x), n = 1, 2, \dots; \quad (3.23)$$

$$0 = -ap_0 + \int_0^\infty p_1(x)\mu(x)dx; \quad (3.24)$$

$$p_n(0^+) = \int_0^\infty p_{n+1}(x)\mu(x)dx + \delta_{n,1}ap_0, n = 1, 2, \dots \quad (3.25)$$

Równanie z numerem  $n$  we wzorze (3.23) mnożymy przez  $z^n$ , a następnie sumujemy wszystkie otrzymane równania względem  $n$ . W rezultacie otrzymujemy następujące równanie na FT  $p(z, x)$  określoną wzorem (3.15):

$$\frac{\partial p(z, x)}{\partial x} = -(a - az + \mu(x))p(z, x). \quad (3.26)$$

W sposób analogiczny przechodząc do FT we wzorze (3.25) otrzymujemy:

$$p(z, 0) = z^{-1} \int_0^\infty p(z, x)\mu(x)dx - \int_0^\infty p_1(x)\mu(x)dx + azp_0. \quad (3.27)$$

Ze wzoru (3.24) wynika, że  $\int_0^\infty p_1(x)\mu(x)dx = ap_0$ , zatem wzór (3.27) można przedstawić w postaci:

$$p(z, 0) = z^{-1} \int_0^\infty p(z, x)\mu(x)dx - a(1 - z)p_0. \quad (3.28)$$

Biorąc pod uwagę postać funkcji  $\mu(x)$  (def. 3.11) można rozwiązanie równania (3.26) uzyskać w postaci:

$$p(z, x) = [1 - B(x)]e^{-(a-az)x}p(z, 0). \quad (3.29)$$

Podstawienie tak określonej funkcji  $p(z, x)$  do wzoru (3.28) prowadzi do wyrażenia:

$$p(z, 0) = \frac{p(z, 0)}{z} \int_0^\infty [1 - B(x)]e^{-(a-az)x}\mu(x)dx - a(1 - z)p_0.$$

Ponieważ  $\mu(x) = \frac{b(x)}{1-B(x)}$ , zatem  $p(z, 0) = \frac{(a-az)p_0z}{\beta(a-az)-z}$ . Wówczas na mocy wzoru (3.27) mamy:



$$p(z, x) = \frac{(a - az)p_0 z}{\beta(a - az) - z} [1 - B(x)] e^{-(a-az)x}. \quad (3.30)$$

Ze wzorów (3.16) i (3.30) otrzymujemy

$$P(z) = p_0 + \frac{p_0 z (1 - \beta(a - az))}{\beta(a - az) - z},$$

lub w innej postaci

$$P(z) = \frac{p_0 (1 - z) \beta(a - az)}{\beta(a - az) - z}.$$

Ponieważ  $p_0 = 1 - \rho$ , co wynika z warunku normalizacyjnego  $P(1) = 1$ , to ostatecznie otrzymujemy:

$$P(z) = \frac{(1 - \rho)(1 - z) \beta(a - az)}{\beta(a - az) - z}. \quad (3.31)$$

Wzór (3.31) jest nazywany wzorem Pollaczka–Chinczyna. Ze wzoru tego możemy na podstawie własności FT wyznaczyć wartości pierwszych dwóch momentów liczby zgłoszeń znajdujących się w systemie w warunkach stacjonarnych:

$$\begin{aligned} E\eta &= P'(1) = \rho + \frac{a^2 \beta_2}{2(1 - \rho)}; \\ E\eta^2 &= P''(1) + P'(1) = \rho + \frac{a^2 \beta_2}{2(1 - \rho)} + \\ &+ a^2 \beta_2 + \frac{a^3 \beta_3}{3(1 - \rho)} + \frac{a^4 \beta_2^2}{2(1 - \rho)^2} + \frac{a^2 \beta_2 \rho}{1 - \rho}, \end{aligned}$$

gdzie  $\beta_2, \beta_3$  są odpowiednio drugim i trzecim momentem czasu obsługi.



## Systemy informatyczne o stochastycznym charakterze działania

W rozdziale tym została wprowadzona nowa klasa modeli systemów informatycznych, które są w pewnym sensie uogólnieniem klasycznych modeli systemów obsługi prezentowanych w poprzednim rozdziale. Jako system informatyczny o stochastycznym charakterze działania rozumiemy tu system obsługi procesów przybywających do systemu komputerowego składający się z jednego lub większej ilości komputerów obsługujących te procesy.

Uogólnienie polega na przypisaniu przybywającym do systemu procesom pewnego losowego rozmiaru, który odpowiada objętości pamięci (operacyjnej lub dyskowej) potrzebnej do zapisania informacji o procesie. Wprowadzenie tej dodatkowej zmiennej losowej charakteryzującej przybywające (uruchamiane) procesy pozwala na badanie bardzo ciekawych modeli o dużej wartości praktycznej. Istotnie, zgłoszenie może być tutaj rozumiane jako pewien proces potrzebujący określonych zasobów pamięci komputera. Badanie takich modeli służy do prawidłowego wyboru rozmiaru pamięci na etapie projektowania systemu informatycznego w tym sensie, że pewne charakterystyki utrat zgłoszeń spowodowane ograniczeniem rozmiaru pamięci mają znajdować się w określonym z góry zakresie.

Ze względu na sposób opisywania matematycznego stochastycznych systemów informatycznych dla ich oznaczenia będziemy nadal korzystać z oznaczeń, czasem zmodyfikowanych, odpowiednich systemów obsługi.

Na początku rozdziału przedstawiona jest klasyfikacja stochastycznych systemów informatycznych. Jest jasne, że modele takich systemów należy dzielić zarówno ze względu na ograniczony lub nieograniczony rozmiar pamięci, jak i ze względu na charakter zależności czasu obsługi i rozmiaru procesu.

Następnie przedstawiona jest analiza pewnych wybranych systemów informatycznych, w szczególności systemu  $M/M/n/(m, V)$ , w którym czas obsługi i rozmiar procesu są niezależne oraz systemu  $M/G/1/\infty$ . Dla pierwszego z systemów wyprowadzone są charakterystyki ilości procesów obecnych w systemie oraz prawdopodobieństwa utrat. Jeżeli chodzi o kolejny system, to ponieważ

jest to system bez utrat, zostały dla niego wyprowadzone charakterystyki objętości sumarycznej. Jak później zostało pokazane, te na pierwszy rzut oka niepraktyczne obliczenia mogą być wykorzystane do szacowania prawdopodobieństw utrat informacji w systemach analogicznych lecz z ograniczonym rozmiarem pamięci. Podstawy teorii obsługi zgłoszeń niejednorodnych, które stanowią bazę dla rozważań prowadzonych w tym rozdziale można znaleźć przede wszystkim w monografiach [35] oraz [44], a także w artykułach [16], [17], [36],[38],[49] oraz [50]. Pewne aspekty praktycznych zastosowań znajdują się m.in w pracach [20] oraz [24].

## 4.1 Klasyfikacja stochastycznych systemów informatycznych

### 4.1.1 Modele obsługi zgłoszeń niejednorodnych

Próby badania systemów obsługi, w których przybywające zgłoszenia charakteryzują się pewnym losowym rozmiarem, były prowadzone już w oparciu o klasyczne modele teorii obsługi masowej. W literaturze z sieci telekomunikacyjnych można znaleźć analizę takich modeli ([27], [28]). Niestety klasyczna teoria obsługi nie wystarcza do uzyskania dobrych wyników dotyczących charakterystyk takich systemów. Okazuje się bowiem, że dla badania systemów obsługi zgłoszeń niejednorodnych (tj. zgłoszeń, które oprócz czasu przybycia i czasu obsługi różnią się również pewnym losowym rozmiarem) należy wprowadzić uogólnienie w sensie matematycznym poznanych metod i modeli, a także należy brać pod uwagę charakter zależności czasu obsługi i objętości zgłoszenia oraz ewentualne ograniczenie objętości sumarycznej. Zaniedbanie tych zależności i ograniczeń prowadzi bowiem do błędnych wyników i w praktyce do złego doboru objętości pamięci systemu, takiego, który nie gwarantuje, że z góry przyjęte prawdopodobieństwa utrat zgłoszeń nie będą przekroczone.

W realnych systemach informatycznych objętość sumaryczna zwykle jest ograniczona wielkością stałą  $V > 0$ , którą będziemy nazywać rozmiarem pamięci. Dokładniej, proces losowy  $\sigma(t)$  (objętość wszystkich procesów zajmujących pamięć w chwili czasu  $t$ ) zwany objętością sumaryczną jest ograniczony.

Jednym z głównych problemów opracowywania systemów informatycznych jest problem wyznaczenia rozmiaru pamięci  $V$  w taki sposób, aby można zapewnić, by z góry przyjęte charakterystyki utrat informacji nie były przekroczone. Zauważmy również, że mechanizm utrat zgłoszeń w systemach klasycznych oraz w systemach z ograniczoną objętością sumaryczną nie jest taki sam. Istotnie, w systemach klasycznych zgłoszenie zostaje utracone, gdy nie ma wolnych urządzeń i wolnych miejsc oczekiwania w kolejce, w systemach z

ograniczoną objętością sumaryczną proces zostaje utracony także wtedy, gdy suma jego rozmiaru i objętości sumarycznej wszystkich procesów zajmujących pamięć w chwili jego przybycia przekroczy wartość rozmiaru pamięci  $V$ .

W ramach klasycznej teorii obsługi do wyznaczenia rozmiaru pamięci systemu  $V$  wykorzystywano założenie, że w przypadku małego prawdopodobieństwa utraty  $p_u$ , wielkość  $V$  można zastąpić przez pierwiastek równania  $1 - D(x) = p_u$ , gdzie  $D(x)$  jest dystrybuantą objętości sumarycznej systemu obsługi o takich samych parametrach jak analizowany system, lecz z nieograniczonym rozmiarem pamięci ( $V = \infty$ ). Okazuje się, że takie uproszczenia nie gwarantują bezpiecznego doboru rozmiaru pamięci.

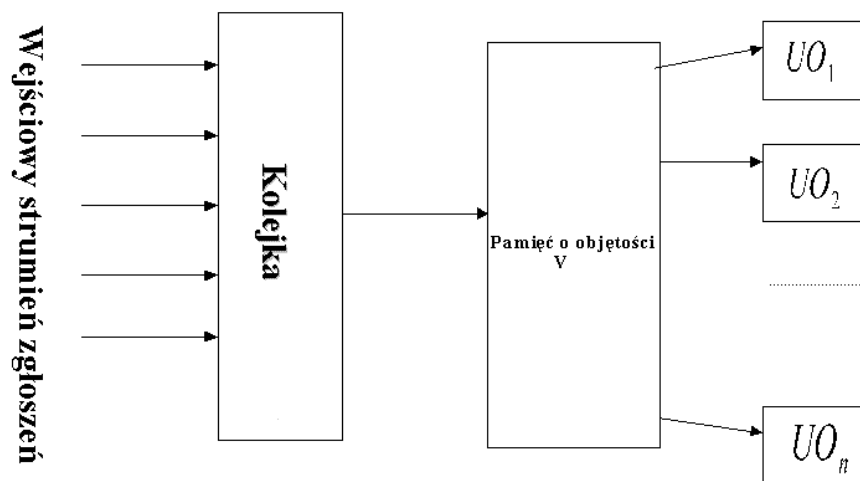
W dalszym ciągu wprowadzona jest klasyfikacja modeli systemów informatycznych oraz przedstawione są wybrane modele należące do wybranych klas.

W celu porównania rozpatrzmy możliwie ogólny system klasyczny tj. system  $G/G/n/m$  (rys. 2.2). Jest jasne, że w ramach klasycznej teorii obsługi masowej nie możemy wyznaczyć ani charakterystyk objętości sumarycznej, ani charakterystyk utrat informacji wynikających z jej ograniczenia.

Założmy dodatkowo, że każdy proces niezależnie od innych oraz czasu swego przybycia charakteryzuje się pewnym losowym rozmiarem  $\zeta$ , będącym dodatnio określoną zmienną losową. Zakładamy, że czas obsługi  $\xi$  i rozmiar procesu  $\zeta$  są w ogólnym przypadku zależne. Niech dana będzie dystrybuanta wektora losowego  $(\zeta, \xi)$ :

$$F(x, t) = P\{\zeta < x, \xi < t\}.$$

W tym przypadku oczywiście dystrybuanta rozmiaru procesu ma postać  $L(x) = P\{\zeta < x\} = F(x, \infty)$ , natomiast dystrybuanta czasu obsługi ma postać  $B(t) = P\{\xi < t\} = F(\infty, t)$ . Oznaczmy przez  $\sigma(t)$  całkowitą objętość sumaryczną procesów zajmujących pamięć w chwili czasu  $t$ . Jeżeli dla danego systemu istnieje tryb stacjonarny, to  $\sigma(t) \Rightarrow \sigma$  przy  $t \rightarrow \infty$  w sensie zbieżności według rozkładu. Zmienna losowa  $\sigma$  jest nazywana stacjonarną objętością sumaryczną. Przez  $V$  będziemy oznaczać rozmiar pamięci systemu ( $0 < V \leq \infty$ ). Jest jasne, że  $0 \leq \sigma(t) \leq V$ . System, który różni się od systemu przedstawionego na rysunku 2.2 tym, że procesy w tym systemie są charakteryzowane pewnym losowym rozmiarem oraz objętość sumaryczna jest ograniczona przez wielkość  $V$ , będziemy oznaczać przez  $G/G/n/(m, V)$ . Struktura takiego systemu jest przedstawiona na rys. 4.1.



Rysunek 4.1. Ogólny schemat działania systemu obsługi zgłoszeń niejednorodnych z ograniczoną objętością sumaryczną

#### 4.1.2 Klasyfikacja modeli systemów informatycznych o stochastycznym charakterze działania

Klasyfikację takich modeli prowadzi się zwykle ze względu na charakter zależności zmiennych losowych  $\xi$  i  $\zeta$  oraz ze względu na wartość wielkości  $V$ . Istotnie, można tu wydzielić cztery klasy takich modeli.

1. Modele, dla których  $V = \infty$ , a zmienne losowe  $\xi$  i  $\zeta$  są niezależne ( $F(x, t) = L(x)B(t)$ ).
2. Modele, dla których  $V < \infty$ , a zmienne losowe  $\xi$  i  $\zeta$  są niezależne ( $F(x, t) = L(x)B(t)$ ).
3. Modele, dla których  $V = \infty$ , a zmienne losowe  $\xi$  i  $\zeta$  są zależne ( $F(x, t) \neq L(x)B(t)$ ).
4. Modele, dla których  $V < \infty$ , a zmienne losowe  $\xi$  i  $\zeta$  są zależne ( $F(x, t) \neq L(x)B(t)$ ).

Wewnątrz każdej z klas możemy analizować modele różnego typu.

Jeżeli chodzi o modele z pierwszej klasy, to ich badanie możliwe jest za pomocą metod klasycznej teorii obsługi, ponieważ z uwagi na niezależność zmiennych  $\xi$  oraz  $\zeta$  rozkład objętości sumarycznej jest wyznaczony przez rozkład liczby procesów oraz postać dystrybuanty  $L(x)$ . Jeżeli chodzi o modele z drugiej i czwartej klasy, to ich podstawowymi charakterystykami są rozkład liczby procesów w określonej chwili czasu oraz w warunkach stacjonarnych oraz prawdopodobieństwa utrat. Modele z drugiej klasy dobrze odwzorowują realne systemy, w których można zrezygnować z zależności czasu obsługi i rozmiaru procesu. W modelach z trzeciej klasy badamy zwykle charakterystyki stacjonarne objętości sumarycznej, natomiast okazuje się, że można modele te wykorzystać do szacowania charakterystyk utrat informacji w analogicznych systemach lecz z ograniczoną objętością sumaryczną. Modele z czwartej klasy są najbardziej skomplikowane, ich analiza matematyczna jest możliwa tylko w niektórych szczególnych przypadkach. Zauważmy jeszcze, że w modelach z trzeciej klasy nie ma w zasadzie utrat procesów, dlatego będziemy je oznaczać tak samo jak ich klasyczne odpowiedniki oraz korzystać z wyników klasycznych dotyczących na przykład rozkładu liczby zgłoszeń.

Teraz wprowadzone zostaną pewne pojęcia i oznaczenia potrzebne do analizy wybranych modeli.

Niech  $D(x, t)$  oznacza dystrybuantę objętości sumarycznej w chwili czasu  $t$  tj.  $D(x, t) = P\{\sigma(t) < x\}$ . Wówczas w przypadku istnienia trybu stacjonarnego mamy  $\sigma(t) \Rightarrow \sigma$  przy  $t \rightarrow \infty$ , gdzie  $\sigma$  jest stacjonarną objętością sumaryczną. Dla dystrybuanty tej zmiennej losowej wprowadzamy oznaczenie  $D(x)$ . Jest jasne, że w warunkach stacjonarnych  $D(x, t) \rightarrow D(x)$  przy  $t \rightarrow \infty$ . Wprowadzamy również oznaczenie  $\delta(s)$  dla PLS dystrybuanty  $D(x)$ . Najczęściej znalezienie postaci jawnej dystrybuanty  $D(x)$  jest niemożliwe, dlatego musimy ograniczyć się do wyznaczenia funkcji  $\delta(s)$ , co pozwala na obliczenie jej podstawowych charakterystyk np. momentów.

Oznaczmy przez

$$\alpha(s, q) = Ee^{-s\zeta - q\xi} = \int_0^\infty \int_0^\infty e^{-sx - qt} dF(x, t)$$

podwójne PLS dystrybuanty łącznej  $F(x, t)$ . Wówczas PLS  $\varphi(s)$ ,  $\beta(q)$  dystrybuant  $L(x)$  oraz  $B(t)$  odpowiednio możemy wyznaczyć korzystając ze wzorów  $\varphi(s) = \alpha(s, 0)$ ,  $\beta(q) = \alpha(0, q)$ . Moment mieszany rzędu  $i + j$  wektora losowego  $(\zeta, \xi)$  wyznaczamy wówczas ze wzoru:

$$\alpha_{ij} = E(\zeta^i \xi^j) = (-1)^{i+j} \frac{\partial^{i+j}}{\partial s^i \partial q^j} \alpha(s, q) \Big|_{s=0, q=0}.$$

Wtedy momenty zmiennych losowych  $\xi$  i  $\zeta$  rzędu  $i$  oznaczamy odpowiednio przez  $\beta_i$  i  $\varphi_i$ . Jest oczywiste, że dla dowolnych  $i \geq 1$  mamy  $\beta_i = \alpha_{0i}$  i  $\varphi_i = \alpha_{i0}$ .

Zauważmy teraz, że objętość sumaryczna  $\sigma(t)$  może być dla dowolnego systemu przedstawiona w postaci sumy:

$$\sigma(t) = \sigma_1(t) + \sigma_2(t),$$

gdzie  $\sigma_1(t)$  jest objętością sumaryczną procesów oczekujących w chwili czasu  $t$ , a  $\sigma_2(t)$  jest objętością sumaryczną procesów obsługiwanych w tej chwili czasu, z takiego podziału procesów będziemy korzystać dość często.

Niech  $\chi(t)$  oznacza objętość procesu zajmującego urządzenie w chwili  $t$ , a  $\xi^*(t)$  będzie odcinkiem czasu od początku obsługi procesu obsługiwanego w chwili czasu  $t$  do chwili  $t$ ,  $F(x, t)$  niech standardowo oznacza dystrybuantę łączną zmiennych losowych  $\zeta$  i  $\xi$  (rozmiar procesu i czas obsługi), natomiast  $B(t) = F(\infty, t)$  – dystrybuantę czasu obsługi. Przez  $E_y(x) = P\{\chi(t) < x | \xi^*(t) = y\}$  oznaczmy dystrybuantę warunkową rozmiaru procesu zajmującego urządzenie w chwili  $t$  pod warunkiem, że od początku jego obsługi do tej chwili upłynęło  $y$  jednostek czasu. Zachodzi następujący związek używany często przy analizie wielu modeli systemów informatycznych.

**Lemat 4.1.** Spełniona jest następująca równość:

$$dE_y(x) = [1 - B(y)]^{-1} \int_{u=y}^{\infty} dF(x, u).$$

Zauważmy, że funkcja  $E_y(x)$  może być również przedstawiona w postaci:

$$E_y(x) = \int_0^x dE_y(u) = P\{\zeta < x | \xi \geq y\} = \frac{P\{\zeta < x, \xi \geq y\}}{P\{\xi \geq y\}} = \frac{L(x) - F(x, y)}{1 - B(y)}.$$

Z lematu 4.1 wynika, że PLS dystrybuanty  $E_y(x)$  ma postać:

$$e_y(s) = \int_0^{\infty} e^{-sx} dE_y(x) = [1 - B(y)]^{-1} \int_{x=0}^{\infty} e^{-sx} \int_{u=y}^{\infty} dF(x, u).$$

Przejdziemy teraz do analizy wybranych modeli systemów informatycznych.



## 4.2 Analiza systemu obsługi $M/M/n/(m, V)$

W punkcie tym przeanalizowany jest przykładowy model systemu informatycznego należącego do drugiej klasy, tzn. model, w którym zmienne losowe oznaczające czas obsługi  $\xi$  i rozmiar procesu  $\zeta$  są niezależne, natomiast rozmiar pamięci  $V$  jest ograniczony. Modelem, który jest analizowany jest model systemu  $M/M/n/(m, V)$ . Dla systemu tego wyznaczony jest rozkład stacjonarny liczby procesów obecnych w systemie oraz stacjonarne prawdopodobieństwo utraty procesu. Przy analizie modelu używane są standardowe oznaczenia wprowadzone dla markowowskich systemów obsługi.

Dla danego systemu tryb stacjonarny istnieje, gdy tylko spełniona jest nierówność  $\rho = a/(n\mu) < \infty$  ( $a$  oznacza parametr najprostszego wejściowego strumienia procesów,  $\mu$  – parametr rozkładu wykładniczego czasu obsługi). Przez  $L(x)$  oznaczmy dystrybuantę rozmiaru procesu. Omówmy teraz krótko działanie tego systemu.

Jeżeli w chwili  $\tau$  przybycia do systemu procesu o rozmiarze  $x$  są zajęte wszystkie urządzenia obsługi (serwery) i wszystkie miejsca oczekiwania lub objętość sumaryczna  $\sigma(\tau^-)$  wszystkich procesów obecnych w systemie bezpośrednio przed chwilą przybycia procesu jest taka, że  $\sigma(\tau^-) + x > V$ , to przybywający proces zostanie utracony, co nie wpływa na dalszą pracę systemu. W przeciwnym wypadku proces zostanie do systemu przyjęty i będzie obsługiwany jednym z wolnych serwerów tak jak w klasycznym systemie  $M/M/n/m$ , przy czym wówczas  $\eta(\tau) = \eta(\tau^-) + 1$  oraz  $\sigma(\tau) = \sigma(\tau^-) + x$ . Natomiast w chwili  $t$  zakończenia obsługi procesu mamy  $\eta(t) = \eta(t^-) - 1$  oraz  $\sigma(t) = \sigma(t^-) - x$ . Założymy dodatkowo, że procesy są obsługiwane zgodnie z dyscypliną FIFO (to założenie nie wpływa na wyniki, lecz ułatwia wprowadzenie odpowiedniego procesu markowowskiego i wypisanie równań dla tego procesu). Przy przyjętych założeniach i oznaczeniach system może być charakteryzowany przez markowowski proces stochastyczny  $(\eta(t), \sigma_1(t), \dots, \sigma_{\eta(t)}(t))$ , gdzie  $\sigma_1(t), \dots, \sigma_{\eta(t)}(t)$  są rozmiarami procesów znajdujących się w systemie w chwili  $t$  ponumerowanych w kolejności przybycia. Jest jasne, że zachodzi równość  $\sigma(t) = \sum_{i=1}^{\eta(t)} \sigma_i(t)$ .

Wprowadzimy teraz funkcje, które charakteryzują wskazany wcześniej proces losowy:

$$P_k(t) = P\{\eta(t) = k\}, \quad k = \overline{0, n+m};$$

$$F_k(y, t) = P\{\eta(t) = k, \sigma(t) < y\}, \quad k = \overline{1, n+m}.$$

Jest jasne, że dla  $k = \overline{1, n+m}$  jest spełniona równość  $P_k(t) = F_k(V, t)$ .

Wypiszemy teraz równania dla wprowadzonych funkcji. Równania te wynikają z analizy zmiany stanów wprowadzonego procesu. Jak można zauważyć, dzięki pewnym własnościom rozkładu wykładniczego równania te niewiele różnią się od prostych równań Kołmogorowa dla zwykłego systemu  $M/M/n/m$  (wzór (3.7)). Rzeczywiście, w równaniach różnicowych poprzedzających równania różniczkowe składniki odwzorowujące moment zakończenia obsługi nie zmieniają się w stosunku do analogicznych równań systemu klasycznego, natomiast zmianie ulegną składniki odwzorowujące moment przybycia procesu, gdyż musimy uwzględnić ograniczenie objętości sumarycznej. Ostatecznie więc składniki typu  $aP_0(t)$  należy zastąpić przez składniki  $aP_0(t)L(V)$ , a składniki typu  $aP_k(t)$  należy zastąpić składnikami  $a \int_0^V F_k(V-y, t)dL(y)$ . Po tych zmianach otrzymujemy następujące równania:

$$P'_0(t) = -aP_0(t)L(V) + \mu P_1(t); \quad (4.1)$$

$$P'_1(t) = aP_0(t)L(V) - a \int_0^V F_1(V-y, t)dL(y) - \mu P_1(t) + 2\mu P_2(t); \quad (4.2)$$

$$P'_k(t) = a \int_0^V F_{k-1}(V-y, t)dL(y) - a \int_0^V F_k(V-y, t)dL(y) - k\mu P_k(t) + (k+1)\mu P_{k+1}(t), \quad k = \overline{2, n-1}; \quad (4.3)$$

$$P'_k(t) = a \int_0^V F_{k-1}(V-y, t)dL(y) - a \int_0^V F_k(V-y, t)dL(y) - n\mu P_k(t) + n\mu P_{k+1}(t), \quad k = \overline{n, n+m-1}; \quad (4.4)$$

$$P'_{n+m}(t) = a \int_0^V F_{n+m-1}(V-y, t)dL(y) - n\mu P_{n+m}(t). \quad (4.5)$$

Wskutek istnienia trybu stacjonarnego istnieją niezależne od warunków początkowych granice:

$$f_k(y) = \lim_{t \rightarrow \infty} F_k(y, t), \quad k = \overline{1, n+m};$$

$$p_k = \lim_{t \rightarrow \infty} P_k(t), \quad k = \overline{0, n+m}.$$

Oczywiście również dla  $k = \overline{1, n+m}$  mamy:

$$p_k = f_k(V). \quad (4.6)$$

Dokonując zatem przejść granicznych przy  $t \rightarrow \infty$  w równaniach (4.1)–(4.5) otrzymujemy następujące równania stacjonarne:

$$0 = -ap_0L(V) + \mu p_1; \quad (4.7)$$

$$0 = ap_0L(V) - a \int_0^V f_1(V-y)dL(y) - \mu p_1 + 2\mu p_2; \quad (4.8)$$

$$0 = a \int_0^V f_{k-1}(V-y)dL(y) - a \int_0^V f_k(V-y)dL(y) - k\mu p_k + (k+1)\mu p_{k+1}, \quad k = \overline{2, n-1}; \quad (4.9)$$

$$0 = a \int_0^V f_{k-1}(V-y)dL(y) - a \int_0^V f_k(V-y)dL(y) - n\mu p_k + n\mu p_{k+1}, \quad k = \overline{n, n+m-1}; \quad (4.10)$$

$$0 = a \int_0^V f_{n+m-1}(V-y)dL(y) - n\mu p_{n+m}. \quad (4.11)$$

Wprowadźmy oznaczenie  $L_k(y) = L_*^{(k)}(y)$  tj. dystrybuanta  $L_k(y)$  jest spłotem według Stieltjesa rzędu  $k$  dystrybuanty  $L(y)$ :

$$L_0(y) \equiv 1, \quad L_k(y) = \int_0^y L_{k-1}(y-x)dL(x), \quad k = 1, 2, \dots$$

Za pomocą podstawienia można sprawdzić, że rozwiązanie układu (4.7)–(4.11) ma postać:

$$f_k(y) = \begin{cases} \frac{(n\rho)^k}{k!} p_0 L_k(y), & \text{gdy } k = \overline{1, n}; \\ \frac{n^n \rho^k}{n!} p_0 L_k(y), & \text{gdy } k = \overline{n+1, n+m}. \end{cases}$$

Wówczas na mocy wzoru (4.6) możemy uzyskać formuły określające prawdopodobieństwa stacjonarne  $p_k$ :

$$p_k = \begin{cases} \frac{(n\rho)^k}{k!} p_0 L_k(V), & \text{gdy } k = \overline{1, n}; \\ \frac{n^n \rho^k}{n!} p_0 L_k(V), & \text{gdy } k = \overline{n+1, n+m}. \end{cases} \quad (4.12)$$

Natomiast na mocy warunku normalizacyjnego mamy:

$$p_0 = \left[ \sum_{k=0}^n \frac{(n\rho)^k}{k!} L_k(V) + \frac{n^n}{n!} \sum_{k=n+1}^{n+m} \rho^k L_k(V) \right]^{-1}. \quad (4.13)$$

Prawdopodobieństwo utraty procesu dla tego systemu nie jest oczywiście równe  $p_{n+m}$ , gdyż procesy mogą być utracone również wówczas, gdy są w systemie wolne urządzenia lub wolne miejsca oczekiwania. Prawdopodobieństwo utraty  $p_u$  wyznaczmy z tzw. warunku równowagi, na mocy którego średnia liczba procesów przybywających w trybie stacjonarnym w ciągu jednostki czasu, które nie są utracone, jest równa średniej liczbie procesów obsłużonych w ciągu jednostki czasu. Stąd otrzymujemy następujące równanie:

$$a(1 - p_u) = \mu \sum_{k=1}^{n-1} k p_k + n\mu \left( 1 - \sum_{k=0}^{n-1} p_k \right).$$

Rozwiązanie tego równania prowadzi do wzoru:

$$p_u = 1 - (n\rho)^{-1} \sum_{k=1}^{n-1} kp_k - \rho^{-1} \left(1 - \sum_{k=0}^{n-1} p_k\right), \quad (4.14)$$

gdzie  $p_k$  są określone przez wzory (4.12).

### 4.3 Zagadnienie wyznaczania charakterystyk utrat informacji

W punkcie tym zastanowimy się nad wartością praktyczną modeli należących do trzeciej klasy, czyli klasy modeli, w których czas obsługi i rozmiar procesu są zależne, a rozmiar pamięci jest nieograniczony. Okazuje się, że charakterystyki objętości sumarycznej w takich modelach można wykorzystać do przybliżonego wyznaczania charakterystyk utrat informacji w analogicznych systemach, lecz z ograniczoną objętością sumaryczną.

Rozpatrzmy system obsługi dowolnego typu z najprostszym strumieniem wejściowym bez utrat, który oznaczymy jako  $SO_\infty$  oraz system , który różni się od poprzedniego tym, że objętość sumaryczna jest w nim ograniczona wielkością stałą  $V > 0$ . Wprowadzony system oznaczajmy  $SO_V$ . W dalszym ciągu charakterystyki pierwszego z systemów oznaczajmy wskaźnikiem dolnym  $\ll \infty \gg$ , natomiast charakterystyki drugiego z systemów - wskaźnikiem dolnym  $\ll V \gg$ .

Przejdźmy teraz do omówienia charakterystyk utrat dla systemów z ograniczoną objętością sumaryczną. Taką charakterystyką podstawową jest jest prawdopodobieństwo utraty procesu  $p_u$ . Jeśli  $D_V(x)$  oznacza dystrybuantę objętości sumarycznej w systemie  $SO_V$ , a  $L(x)$  - dystrybuantę rozmiaru procesu, to stacjonarne prawdopodobieństwo utraty procesu wyznaczamy ze wzoru:

$$p_u = 1 - \int_0^V D_V(V-x)dL(x).$$

Charakterystyka ta ma jednak taką wadę, że charakteryzuje ona część względną utraconych procesów, lecz nie część utraconej objętości informacji. Jest jasne, że przybywający do systemu proces ma tym większą szansę zostać utraconym, im większy ma on rozmiar. Dlatego wśród procesów utraconych większą część stanowią procesy o większym rozmiarze. Wprowadzimy teraz kolejną charakterystykę utrat nazywaną prawdopodobieństwem utraty jednostki informacji.

Rozważmy procesy, które są utracone w chwili przybycia do systemu. Niech  $\gamma$  oznacza rozmiar takiego procesu, a  $F_\gamma(x)$  – dystrybuantę zmiennej losowej  $\gamma$ . Jest jasne, że wartość oczekiwana  $\kappa$  zmiennej losowej  $\gamma$  wyraża się wzorem:

$$\kappa = E\gamma = \int_0^\infty x dF_\gamma(x),$$

gdzie  $dF_\gamma(x) = P\{\gamma \in [x, x + dx)\}$  jest prawdopodobieństwem tego, że rozmiar utraconego procesu jest równy  $x$  (dokładniej znajduje się w przedziale  $[x, x + dx)$ ). Z uwagi na niezależność rozmiaru przybywającego procesu od objętości sumarycznej innych procesów znajdujących się w systemie w momencie przybycia procesu otrzymujemy równość:

$$P\{\gamma \in [x, x + dx)\} = P\{\zeta \in [x, x + dx), \sigma_V > V - x\} = [1 - D_V(V - x)]dL(x),$$

skąd na mocy równości  $D_V(V - x) = 0$  dla  $x > V$  otrzymujemy następujące wyrażenie pozwalające wyznaczyć wielkość  $\kappa$ :

$$\begin{aligned} \kappa &= E\gamma = \int_0^\infty x[1 - D_V(V - x)]dL(x) = \\ &= \varphi_1 - \int_0^\infty xD_V(V - x)dL(x) = \varphi_1 - \int_0^V xD_V(V - x)dL(x). \end{aligned}$$

Jest jasne, że  $\kappa < \varphi_1$ .

Określmy teraz prawdopodobieństwo utraty jednostki informacji, jest to wielkość  $Q$  wyrażona poprzez formułę:

$$Q = \frac{\kappa}{\varphi_1} = 1 - \frac{1}{\varphi_1} \int_0^V xD_V(V - x)dL(x).$$

Wielkość  $Q$  charakteryzuje część względną utraty informacji w stosunku do jednostki rozmiaru procesu. Dokładniej  $Q$  wyraża część względną informacji utraconej z powodu ograniczenia objętości sumarycznej. Jest ona bardziej obiektywną charakterystyką niż prawdopodobieństwo utraty  $p_u$ , jednak jej obliczanie w większości przypadków jest o wiele trudniejsze.

Założmy teraz, że dla systemu  $SO_\infty$  (więc również dla  $SO_V$ ) istnieje tryb stacjonarny. Niech  $\sigma_\infty$  oraz  $\sigma_V$  będą odpowiednio stacjonarnymi objętościami sumarycznymi dla systemów  $SO_\infty$  i  $SO_V$ , a  $D_\infty(x)$  oraz  $D_V(x)$  – odpowiadającymi im dystrybuantami. Zachodzi następujące twierdzenie.

**Twierdzenie 4.1.** Dla dowolnej liczby nieujemnej  $x$  spełniona jest nierówność:

$$D_\infty(x) \leq D_V(x).$$

Twierdzenie 4.1 ma niebagatelne znaczenie. Wynika z niego bowiem, że możemy wykorzystać wyniki uzyskane dla systemów obsługi z nieograniczoną

objętością sumaryczną do szacowania charakterystyk utraty informacji dla analogicznych systemów z ograniczonym rozmiarem pamięci. Istotnie, w sposób oczywisty z nierówności zawartej w tezie twierdzenia wynikają następujące oszacowania:

$$p_u = 1 - \int_0^V D_V(V-x)dL(x) \leq 1 - \int_0^V D_\infty(V-x)dL(x) = p_u^*. \quad (4.15)$$

$$Q = 1 - \frac{1}{\varphi_1} \int_0^V xD_V(V-x)dL(x) \leq 1 - \frac{1}{\varphi_1} \int_0^V xD_\infty(V-x)dL(x) = Q^*. \quad (4.16)$$

Wielkości  $p_u^*$  oraz  $Q^*$  są zatem granicami górnymi dla wielkości  $p_u$  i  $Q$ , a ponieważ w realnych systemach dopuszczalne prawdopodobieństwa utrat są bardzo małe, więc wielkości te niewiele różnią się od  $p_u$  oraz  $Q$  i mogą występować w roli ich estymatorów. W każdym razie objętość pamięci wybrana według wzorów z prawych części nierówności (4.15), (4.16) gwarantuje, że odpowiednie prawdziwe prawdopodobieństwa utrat nie będą większe niż  $p_u^*$  oraz  $Q^*$ .

Uzyskujemy więc następujące wnioski dotyczące badań realnych systemów informatycznych.

1. Dla systemów informatycznych, w których dokładna analiza w przypadku ograniczonej objętości sumarycznej jest bardzo skomplikowana, ważne jest badanie analogicznych systemów o nieograniczonym rozmiarze pamięci i zależnym od rozmiaru procesu czasem obsługi. Dla takich systemów w miarę możliwości powinno się wyznaczać wielkości  $p_u^*$  i  $Q^*$ .

Niestety obliczenia te mogą być utrudnione z uwagi na fakt, że rzadko udaje się wyznaczyć w sposób jawny dystrybucję  $D_\infty(x)$ . Ponieważ  $L(0) = 0$ , to w nierówności (4.15) możemy aproksymować splot  $\Phi(x) = \int_0^x D_\infty(x-u)dL(u)$  dystrybucją  $\Phi^*(x) = \frac{\gamma(q, cx)}{\Gamma(q)}$ , gdzie  $\gamma(q, cx) = \int_0^{cx} t^{q-1} e^{-t} dt$  jest funkcją gamma niezupełną, natomiast  $\Gamma(q) = \gamma(q, \infty)$  jest funkcją gamma. Jest jasne, że pierwsze dwa momenty dystrybucji  $\Phi^*(x)$  wynoszą  $f_1^* = q/c$ ,  $f_2^* = q(q+1)/c^2$ . Natomiast z uwagi na fakt, że  $\Phi(x)$  jest dystrybucją sumy dwóch niezależnych zmiennych losowych  $\sigma_\infty$  oraz  $\zeta$ , jej momenty wyrażają się formułami  $f_1 = \delta_1 + \varphi_1$ ,  $f_2 = \delta_2 + \varphi_2 + 2\delta_1\varphi_1$ , gdzie  $\varphi_1$ ,  $\varphi_2$ ,  $\delta_1$ ,  $\delta_2$  są odpowiednio pierwszymi dwoma momentami objętości sumarycznej oraz rozmiaru procesu. Wówczas parametry  $q$  i  $c$  dystrybucji  $\Phi^*(x)$ , które obliczamy w oparciu o równość momentów  $f_1^*$  i  $f_1$  oraz  $f_2^*$  i  $f_2$ , okazują się równe

$$q = \frac{f_1^2}{f_2 - f_1^2}, \quad c = \frac{f_1}{f_2 - f_1^2}.$$

W nierówności (4.16) dokonujemy najczęściej aproksymacji funkcji  $D_\infty(x)$  funkcją  $D_\infty^*(x) = p_0 + (1-p_0)\frac{\gamma(p, gx)}{\Gamma(p)}$  ([12]). Parametry  $p$  i  $g$  tutaj również wyznaczamy z założenia równości pierwszych dwóch momentów dystrybuant  $D_\infty(x)$  oraz  $D_\infty^*(x)$ . Wówczas

$$p = \frac{\delta_1^2}{(1-p_0)\delta_2 - \delta_1^2}, \quad g = \frac{(1-p_0)\delta_1}{(1-p_0)\delta_2 - \delta_1^2}.$$

Jest jasne, że zastąpienie dystrybuant  $D_\infty(x)$  oraz  $\Phi(x)$  przez wskazane funkcje aproksymacyjne może doprowadzić do tego, że nierówności (4.15)–(4.16) nie będą już spełnione, lecz odpowiedni błąd nie przekracza błędu aproksymacji.

2. Przy badaniu modeli systemów bez utrat informacji z zależnym od rozmiaru procesu czasem obsługi ważne ze względów praktycznych jest wyznaczenie przynajmniej pierwszych dwóch momentów objętości sumarycznej.

**Uwaga.** Liczba procesów znajdujących się w systemie w warunkach stacjonarnych jest przypadkiem szczególnym objętości sumarycznej dla systemów z nieograniczonym rozmiarem pamięci. Rzeczywiście, mając wzory na charakterystyki objętości sumarycznej możemy zawsze wyznaczyć charakterystyki liczby procesów. Wystarczy w tym celu przyjąć, że czas obsługi w analizowanym systemie ma rozkład o dystrybuancie  $B(t) = F(\infty, t)$ , gdzie  $F(x, t)$  jest dystrybuantą wektora losowego  $(\zeta, \xi)$ , a rozmiar każdego procesu  $\zeta \equiv 1$ .

#### 4.4 Analiza systemu obsługi $M/G/1/\infty$

W punkcie tym przedstawione są wyniki dla systemu obsługi  $M/G/1/\infty$  z nieograniczonym rozmiarem pamięci oraz zależnym od rozmiaru procesu czasem obsługi. Dla systemu tego wyznaczone są w szczególności charakterystyki objętości sumarycznej.

Wprowadźmy następujące oznaczenia:  $a$  niech będzie parametrem najprostszego strumienia wejściowego procesów,  $\zeta$  – rozmiarem procesu,  $\xi$  – czasem obsługi,  $F(x, t) = P\{\zeta < x, \xi < t\}$  – dystrybuantą wektora losowego  $(\zeta, \xi)$ ,  $\alpha(s, q)$  – PLS dystrybuanty  $F(x, t)$ ,  $\alpha_{ij}$  – momentem mieszanym rzędu  $i + j$  dystrybuanty  $F(x, t)$ ,  $L(x)$  oraz  $B(t)$  niech będą odpowiednio dystrybuantami zmiennych losowych  $\zeta$  oraz  $\xi$ , a  $\varphi(s)$  oraz  $\beta(q)$  – odpowiadającymi im PLS. Przez  $\varphi_i$ ,  $\beta_i$  oznaczmy natomiast momenty rzędu  $i$  odpowiednio zmiennych losowych  $\zeta$  oraz  $\xi$ . Załóżmy, że dla danego systemu istnieje tryb

stacjonarny tzn. jest spełniony warunek  $\rho = a\beta_1 < 1$ . Zakładamy również, że kolejność obsługi procesów nie zależy od ich rozmiaru. Oznaczmy przez  $\sigma$  stacjonarną objętość sumaryczną w tym systemie i załóżmy, że jest ona nieograniczona ( $V = \infty$ ). Niech  $D(x)$  będzie dystrybuantą zmiennej losowej  $\sigma$ , a  $\delta(s)$  – odpowiadającym jej PLS. Zachodzi następujące twierdzenie.

**Twierdzenie 4.2.** PLS  $\delta(s)$  stacjonarnej objętości sumarycznej  $\sigma$  dla systemu  $M/G/1/\infty$  wyraża poniższa formuła:

$$\delta(s) = (1 - \rho) \left[ 1 + \frac{\varphi(s) - \alpha(s, a - a\varphi(s))}{\beta(a - a\varphi(s)) - \varphi(s)} \right]. \quad (4.17)$$

**Wniosek.** Pierwsze dwa momenty zmiennej losowej  $\sigma$  obliczamy ze wzorów:

$$\delta_1 = -\delta'(0) = a\alpha_{11} + \frac{a^2\beta_2\varphi_1}{2(1-\rho)}, \quad (4.18)$$

$$\begin{aligned} \delta_2 = \delta''(0) = & a(\alpha_{21} + a\varphi_1\alpha_{12}) + \\ & + \frac{a^3\beta_2\varphi_1\alpha_{11}}{1-\rho} + \frac{a^2\beta_2\varphi_2}{2(1-\rho)} + \frac{a^3\beta_3\varphi_1^2}{3(1-\rho)} + \frac{a^4\beta_2^2\varphi_1^2}{2(1-\rho)^2}. \end{aligned} \quad (4.19)$$

Ze wzorów tych można korzystać dla oszacowania charakterystyk utrat informacji w systemie  $M/G/1/(\infty, V)$  z ograniczoną objętością sumaryczną.



## Model pamięci operacyjnej systemu informatycznego

### 5.1 Wstęp

Celem tego rozdziału jest przedstawienie pewnej modyfikacji systemu  $M/G/1/\infty$  z nieograniczoną objętością sumaryczną i zależnym od rozmiaru procesu czasem obsługi analizowanego m.in w pracach [16] oraz [44]. Wspomniana modyfikacja polega na tym, że procesy są podzielone na dwie klasy, które będziemy nazywali procesami wewnętrznymi oraz procesami zewnętrznymi. Wprowadzenie i analiza takiego systemu ma sens praktyczny z uwagi na jego realny charakter. Istotnie, często w praktyce mamy do czynienia z sytuacją, kiedy taki podział wydaje się potrzebny i uzasadniony. Na przykład, programy komputerowe można podzielić na programy zewnętrzne przybywające do systemu i programy wewnętrzne (np. usługi systemu operacyjnego czy inne aplikacje), które są niezbędne do uruchomienia programów zewnętrznych. Często bowiem do uruchomienia pewnej aplikacji zewnętrznej jest potrzebna inna aplikacja o wyższym priorytecie systemowym. Takich przykładów można znaleźć wiele zwłaszcza w rozwiązaniach internetowych. Na przykład, przy uruchamianiu różnych aplikacji pobierane są potrzebne dodatki do bieżących aplikacji czy też programy, bez których udziału uruchomienie aplikacji staje się niemożliwe. W takich sytuacjach jasne jest również, że program zewnętrzny potrzebuje dodatkowej pamięci niezbędnej do obsługi związanego z nim programu wewnętrznego. Zakładamy dodatkowo, że programy wewnętrzne mają większy priorytet obsługi tzn. w czasie obsługi programu wewnętrznego obsługa programu zewnętrznego jest zatrzymywana i kontynuowana dopiero po zakończeniu obsługi programu wewnętrznego. Takie założenie ma również swoje praktyczne uzasadnienie. Analizowany model może być na przykład traktowany jako model pamięci operacyjnej komputera. W dalszej kolejności zastosowano otrzymane wyniki związane z charakterystykami objętości sumarycznej do oszacowania charakterystyk utrat informacji dla systemów o analogicznym kształcie, lecz z ograniczoną objętością sumaryczną. Początki analizy wspomnianego modelu znalazły się

w pracach [46] oraz [47], a w bieżącym rozdziale wyniki uzyskane w tych publikacjach zostały znacznie rozszerzone i uzupełnione między innymi o analizę przypadków szczególnych oraz obliczenia prawdopodobieństw utrat w oparciu o odpowiednie aproksymacje z wykorzystaniem funkcji gamma. Porównano również otrzymane estymatory charakterystyk utrat informacji z wynikami otrzymanymi podczas symulacji komputerowej. Przejdziemy teraz do analizy dwóch schematów związanych z omawianym modelem.

## 5.2 Opis i analiza modelu

Zajmiemy się teraz analizą systemu  $M/G/1/\infty$ , w którym procesy (aplikacje) zostały podzielone na zewnętrzne i wewnętrzne. Jeżeli chodzi o procesy zewnętrzne, to zakładamy, że tworzą one strumień najprostszy z parametrem  $a$ . Dodatkowo zakładamy, że każdy proces zewnętrzny niezależnie od innych charakteryzuje się pewnym losowym rozmiarem  $\zeta$  oraz czasem obsługi  $\xi$ , który w ogólnym przypadku zależy od rozmiaru. Niech  $F(x, t) = P\{\zeta < x, \xi < t\}$  będzie dystrybuantą łączną zmiennych losowych  $\zeta$  i  $\xi$ , natomiast  $L(x) = F(x, \infty)$  oraz  $B(t) = F(\infty, t)$  niech oznaczają dystrybuanty zmiennych  $\zeta$  i  $\xi$  odpowiednio.

Procesy wewnętrzne przybywają do systemu tylko w chwilach, gdy system jest zajęty obsługą procesów zewnętrznych. Zakładamy tu, że jeśli w chwili czasu  $T$  rozpoczęła się obsługa zewnętrznego procesu i nie zakończy się ona w przedziale  $[T, T + t)$ , to z prawdopodobieństwem  $E(t) = 1 - e^{-ct}$ ,  $c > 0$ , do systemu może przybyć proces wewnętrzny. Oznaczmy przez  $\gamma$  i  $\kappa$  rozmiar i czas obsługi procesu wewnętrznego odpowiednio. Niech  $\Phi(x, t) = P\{\gamma < x, \kappa < t\}$  będzie dystrybuantą łączną zmiennych losowych  $\gamma$  i  $\kappa$ . Załóżmy również, że przybywający proces wewnętrzny przerywa obsługę procesu zewnętrznego. Obsługa procesu zewnętrznego jest kontynuowana natychmiast po zakończeniu obsługi procesu wewnętrznego. Zakładamy, że wtedy znów z tym samym prawdopodobieństwem do systemu może przybyć proces wewnętrzny, który przerywa proces obsługi itd. Niech  $R(x) = \Phi(x, \infty)$  oraz  $H(t) = \Phi(\infty, t)$  będą dystrybuantami zmiennych  $\gamma$  i  $\kappa$  odpowiednio.

Oznaczmy przez  $\sigma(t)$  objętość sumaryczną wszystkich procesów (zarówno zewnętrznych jak i wewnętrznych) zajmujących pamięć operacyjną w chwili czasu  $t$  i załóżmy, że objętość ta jest nieograniczona. Działanie systemu można opisać następująco: jeśli  $T$  oznacza chwilę czasu, w której do systemu przybywa proces zewnętrzny lub wewnętrzny, to  $\sigma(T) = \sigma(T^-) + x$ , gdzie  $x$  jest rozmiarem przybywającego procesu. Oznaczmy przez  $\eta(t)$  liczbę procesów zewnętrznych zajmujących pamięć operacyjną w chwili  $t$ . W chwili  $t$  przybycia zewnętrznego procesu mamy oczywiście  $\eta(t) = \eta(t^-) + 1$ . Natomiast jeśli  $t$

jest chwilą zakończenia obsługi procesu zewnętrznego, wtedy  $\eta(t) = \eta(t^-) - 1$ .

Jeżeli chodzi o zmiany objętości sumarycznej w chwilach zakończenia obsługi procesów wewnętrznych, to można wydzielić dwa schematy, które są dalej analizowane.

- **SCHEMAT 1.** Jeśli  $T$  oznacza chwilę zakończenia obsługi procesu wewnętrznego, wówczas  $\sigma(T) = \sigma(T^-) - x$ , gdzie  $x$  jest rozmiarem procesu wewnętrznego. Jeśli natomiast  $T$  jest chwilą zakończenia obsługi procesu zewnętrznego, to  $\sigma(T) = \sigma(T^-) - y$ , gdzie  $y$  oznacza rozmiar procesu zewnętrznego.
- **SCHEMAT 2.** Jeśli  $T$  jest chwilą zakończenia obsługi procesu wewnętrznego, to  $\sigma(T) = \sigma(T^-)$ . Jeśli natomiast  $T$  jest chwilą zakończenia obsługi procesu zewnętrznego, to wówczas  $\sigma(T) = \sigma(T^-) - y$ , gdzie  $y$  jest objętością sumaryczną procesu zewnętrznego oraz wszystkich procesów wewnętrznych przybyłych do systemu podczas obsługi procesu zewnętrznego.

Zatem w schemacie pierwszym wszystkie procesy po zakończeniu swojej obsługi zwalniają pamięć operacyjną, natomiast w schemacie drugim do momentu zakończenia obsługi procesu zewnętrznego wszystkie procesy wewnętrzne, które przybyły do systemu w trakcie obsługi procesu zewnętrznego, zachowują się w pamięci i zwalniają pamięć operacyjną dopiero razem z procesem zewnętrznym.

Wprowadźmy następujące oznaczenia.

Niech

$$\alpha(s, q) = \int_0^\infty \int_0^\infty e^{-sx-qt} dF(x, t) \text{ oraz } g(s, q) = \int_0^\infty \int_0^\infty e^{-sx-qt} d\Phi(x, t)$$

będą podwójnymi przekształceniami Laplace'a–Stieltjesa (PLS) dystrybuant  $F(x, t)$  i  $\Phi(x, t)$  odpowiednio. Oznaczmy przez

$$\alpha_{ij} = (-1)^{i+j} \frac{\partial^{i+j} \alpha(s, q)}{\partial s^i \partial q^j} \Big|_{s=0, q=0}, \quad g_{ij} = (-1)^{i+j} \frac{\partial^{i+j} g(s, q)}{\partial s^i \partial q^j} \Big|_{s=0, q=0}$$

momenty mieszane rzędu  $i + j$  dystrybuant  $F(x, t)$  oraz  $\Phi(x, t)$  odpowiednio,  $i, j = 1, 2, \dots$ . Niech  $\varphi(s) = \alpha(s, 0)$ ,  $\beta(q) = \alpha(0, q)$ ,  $r(s) = g(s, 0)$ ,  $h(q) = g(0, q)$  będą zwykłymi PLS zmiennych losowych  $\zeta$ ,  $\xi$ ,  $\gamma$  oraz  $\kappa$  odpowiednio, natomiast  $\varphi_i$ ,  $\beta_i$ ,  $r_i$ ,  $h_i$  – momentami  $i$ -tego rzędu tych zmiennych.

W dalszym ciągu oznaczmy przez  $D(x, t) = P\{\sigma(t) < x\}$  dystrybuantę objętości sumarycznej wszystkich procesów zajmujących pamięć operacyjną w chwili czasu  $t$ .

Można łatwo wykazać, że warunki stacjonarne dla takiego systemu istnieją, jeżeli jest spełniony warunek  $\rho = a\beta_1(1 + ch_1) < 1$ . W takich warunkach istnieje granica  $D(x) = \lim_{t \rightarrow \infty} D(x, t) = P\{\sigma < x\}$ , gdzie  $\sigma$  jest stacjonarną objętością sumaryczną.

Wprowadźmy następujące oznaczenia. Niech  $\nu(t)$  będzie funkcją wskaźnikową trybu obsługi tzn.  $\nu(t) = 0$ , jeśli w chwili  $t$  trwa obsługa zewnętrznego procesu oraz  $\nu(t) = 1$ , jeśli w chwili  $t$  trwa obsługa procesu wewnętrznego (oczywiście funkcja ta jest nieokreślona, gdy w systemie w chwili czasu  $t$  nie ma procesów). Oznaczmy przez  $\xi_{(0)}^*(t)$  długość odcinka czasowego, w ciągu którego trwała obsługa zewnętrznego procesu zajmującego pamięć w chwili  $t$ . Funkcja ta jest nieokreślona, gdy  $\eta(t) = 0$ . Przez  $\xi_{(1)}^*(t)$  oznaczmy długość odcinka czasowego, w ciągu którego trwała obsługa wewnętrznego procesu zajmującego urządzenie w chwili  $t$ , przy warunku  $\nu(t) = 1$  (jest jasne, że funkcja  $\xi_{(1)}^*(t)$  jest także nieokreślona, gdy  $\nu(t) = 0$ ).

Analizowany system jest charakteryzowany poprzez poniższy proces markowski:

$$\left( \eta(t), \nu(t), \xi_{(0)}^*(t), \xi_{(1)}^*(t) \right) \quad (5.1)$$

Proces ten jest określony poprzez poniższe funkcje:

$$P_0(t) = P\{\eta(t) = 0\}; \quad (5.2)$$

$$P_k(0, x, t)dx = P\{\eta(t) = k, \nu(t) = 0, \xi_{(0)}^*(t) \in [x; x + dx]\}, k = 1, 2, \dots; \quad (5.3)$$

$$P_k(1, x, y, t)dx dy = P\{\eta(t) = k, \nu(t) = 1, \xi_{(0)}^*(t) \in [x; x + dx), \xi_{(1)}^*(t) \in [y; y + dy)\}, k = 1, 2, \dots \quad (5.4)$$

Oznaczmy przez  $b_{(0)}(t)$  oraz  $b_{(1)}(t)$  uogólnione gęstości zmiennych losowych  $\xi$  i  $\kappa$  odpowiednio ([51]) a przez  $\mu_{(0)}(t) = \frac{b_{(0)}(t)}{1 - B(t)}$  i  $\mu_{(1)}(t) = \frac{b_{(1)}(t)}{1 - H(t)}$  odpowiednio intensywności obsługi procesów zewnętrznych i wewnętrznych. W przypadku istnienia trybu stacjonarnego istnieją następujące granice:

$$p_0 = P\{\eta = 0\} = \lim_{t \rightarrow \infty} P_0(t); \quad (5.5)$$

$$\begin{aligned} p_k(0, x)dx &= P\{\eta = k, \nu = 0, \xi_{(0)}^* \in [x; x + dx)\} = \\ &= \lim_{t \rightarrow \infty} P_k(0, x, t)dx, k = 1, 2, \dots; \end{aligned} \quad (5.6)$$

$$\begin{aligned} p_k(1, x, y)dx dy &= P\{\eta = k, \nu = 1, \xi_{(0)}^* \in [x; x + dx), \xi_{(1)}^* \in [y; y + dy)\} = \\ &= \lim_{t \rightarrow \infty} P_k(1, x, y, t)dx dy, k = 1, 2, \dots, \end{aligned} \quad (5.7)$$

gdzie  $\eta$ ,  $\xi_{(0)}^*$ ,  $\xi_{(1)}^*$  są stacjonarnymi odpowiednikami funkcji  $\eta(t)$ ,  $\xi_{(0)}^*(t)$  i  $\xi_{(1)}^*(t)$ .

Niech dalej  $\delta_{i,k}$  będzie standardowo określonym symbolem Kroneckera:

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

Wyprowadzimy teraz równania różnicowe dla funkcji (5.2)–(5.4).

Najpierw znajdziemy  $P_0(t + \Delta t)$ . Aby w chwili  $t + \Delta t$  w systemie nie było procesów zewnętrznych konieczne i dostateczne jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie nie było procesów i w ciągu czasu  $\Delta t$  procesy nie przybywały do systemu; prawdopodobieństwo tego zdarzenia wynosi  $P_0(t)(1 - a\Delta t) + o(\Delta t)$ ;

2) albo w chwili  $t$  w systemie był jeden proces zewnętrzny i został obsłużony w czasie  $\Delta t$ , przy czym tryb obsługi nie uległ zmianie w ciągu czasu  $\Delta t$  (tzn. były obsługiwane wyłącznie procesy zewnętrzne); prawdopodobieństwo tego zdarzenia wynosi  $\int_0^t P_1(0, x, t)\mu_{(0)}(x)\Delta t dx + o(\Delta t)$ .

Prawdopodobieństwo pozostałych zdarzeń prowadzących do stanu  $\{\eta(t + \Delta t) = 0\}$  wynosi  $o(\Delta t)$ . Stąd otrzymujemy następujące równanie:

$$P_0(t + \Delta t) = P_0(t)(1 - a\Delta t) + \int_0^t P_1(0, x, t)\mu_{(0)}(x)\Delta t dx + o(\Delta t). \quad (5.8)$$

Znajdziemy teraz  $P_k(0, x + \Delta t, t + \Delta t)dx$ . Aby w chwili  $t + \Delta t$  w systemie było  $k$  procesów zewnętrznych, przy czym  $\nu(t + \Delta t) = 0$  oraz  $\xi_{(0)}^*(t + \Delta t) \in [x + \Delta t, x + \Delta t + dx)$  konieczne i dostateczne jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie było  $k$  procesów zewnętrznych,  $\nu(t) = 0$  oraz  $\xi_{(0)}^*(t) \in [x, x + dx)$  i w ciągu czasu  $\Delta t$  nie przybywały do systemu żadne procesy (ani wewnętrzne, ani zewnętrzne) oraz obsługa procesu zewnętrznego obsługiwanego w chwili czasu  $t$  nie została zakończona; prawdopodobieństwo tego zdarzenia wynosi  $P_k(0, x, t)dx[1 - (a + c + \mu_{(0)}(x))\Delta t] + o(\Delta t)dx$ ;

2) albo w chwili  $t$  w systemie było  $k$  procesów zewnętrznych,  $\nu(t) = 1$ ,  $\xi_{(0)}^*(t) \in [x, x + dx)$ ,  $\xi_{(1)}^*(t) \in [y, y + dy)$  i w ciągu czasu  $\Delta t$  proces wewnętrzny obsługiwany w chwili  $t$  zakończył swoją obsługę; prawdopodobieństwo tego zdarzenia wynosi  $\int_0^t P_k(1, x, y, t)\mu_{(1)}(y)dx dy \Delta t + o(\Delta t)dx$ ;

3) albo w przypadku  $k \geq 2$  w chwili czasu  $t$  w systemie było  $k - 1$  procesów zewnętrznych,  $\nu(t) = 0$ ,  $\xi_{(0)}^*(t) \in [x, x + dx)$  i w ciągu czasu  $\Delta t$  do systemu przybył proces zewnętrzny; prawdopodobieństwo tego zdarzenia wynosi  $(1 - \delta_{k,1})aP_{k-1}(0, x, t)dx \Delta t + o(\Delta t)dx$ .

Prawdopodobieństwo pozostałych zdarzeń prowadzących do stanu  $\{\eta(t + \Delta t) = k, \nu(t + \Delta t) = 0, \xi_{(0)}^*(t + \Delta t) \in [x + \Delta t, x + \Delta t + dx]\}$  wynosi  $o(\Delta t)dx$ . Z powyższych rozważań wynika poniższe równanie:

$$P_k(0, x + \Delta t, t + \Delta t) = P_k(0, x, t)[1 - (a + c + \mu_{(0)}(x))\Delta t] + \int_0^t P_k(1, x, y, t)\mu_{(1)}(y)dy\Delta t + (1 - \delta_{k,1})aP_{k-1}(0, x, t)\Delta t + o(\Delta t), k = 1, 2, \dots \quad (5.9)$$

W dalszym ciągu znajdziemy  $P\{\eta(t + \Delta t) = k, \nu(t + \Delta t) = 0, \xi_{(0)}^*(t + \Delta t) < \Delta t\}$ . Jest jasne, że aby system znalazł się w oznaczonym stanie konieczne i dostateczne jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie było  $k + 1$  procesów zewnętrznych i w czasie  $\Delta t$  jeden z procesów został obsłużony, przy czym tryb obsługi nie uległ zmianie; prawdopodobieństwo tego zdarzenia wynosi  $\int_0^t P_{k+1}(0, x, t)\mu_{(0)}(x)\Delta tdx + o(\Delta t)$ ;

2) albo w przypadku  $k = 1$  w chwili czasu  $t$  w systemie nie było procesów i w ciągu czasu  $\Delta t$  do systemu przybył proces zewnętrzny; prawdopodobieństwo tego zdarzenia wynosi  $\delta_{k,1}P_0(t)a\Delta t + o(\Delta t)$ .

Prawdopodobieństwo pozostałych zdarzeń prowadzących do omawianego stanu wynosi  $o(\Delta t)$ . Z drugiej strony prawdopodobieństwo  $P\{\eta(t + \Delta t) = k, \nu(t + \Delta t) = 0, \xi_{(0)}^*(t + \Delta t) < \Delta t\}$  jest równe  $\int_0^{\Delta t} P_k(0, x, t + \Delta t)dx$ . Stąd otrzymujemy równanie:

$$\int_0^{\Delta t} P_k(0, x, t + \Delta t)dx = \int_0^t P_{k+1}(0, x, t)\mu_{(0)}(x)dx\Delta t + \delta_{k,1}P_0(t)a\Delta t + o(\Delta t). \quad (5.10)$$

Dalej znajdziemy  $P_k(1, x + \Delta t, y + \Delta t, t + \Delta t)dxdy$ . Aby system znalazł się w oznaczonym stanie konieczne i dostateczne jest spełnienie jednego z warunków:

1) albo w chwili  $t$  w systemie znajdowało się  $k$  procesów zewnętrznych,  $\nu(t) = 1, \xi_{(0)}^*(t) \in [x, x + dx), \xi_{(1)}^*(t) \in [y, y + dy)$  oraz w czasie  $\Delta t$  nie przybył do systemu żaden proces zewnętrzny, a obsługa procesu wewnętrznego obsługiwanego w chwili czasu  $t$  nie została zakończona; prawdopodobieństwo tego zdarzenia wynosi  $P_k(1, x, y, t)dxdy[1 - (a + \mu_{(1)}(y))\Delta t] + o(\Delta t)dxdy$ ;

2) albo w przypadku  $k \geq 2$  w systemie w chwili czasu  $t$  było  $k - 1$  procesów zewnętrznych,  $\nu(t) = 1, \xi_{(0)}^*(t) \in [x, x + dx), \xi_{(1)}^*(t) \in [y, y + dy)$  oraz w czasie  $\Delta t$  do systemu przybył proces zewnętrzny; prawdopodobieństwo tego zdarzenia wynosi  $(1 - \delta_{k,1})aP_{k-1}(1, x, y, t)\Delta tdx dy$ .

Prawdopodobieństwa pozostałych zdarzeń prowadzących do omawianego stanu wynosi  $o(\Delta t)dxdy$ . Stąd otrzymujemy poniższe równanie:

$$P_k(1, x + \Delta t, y + \Delta t, t + \Delta t) = P_k(1, x, y, t)[1 - (a + \mu_{(1)}(y))\Delta t] + \\ + (1 - \delta_{k,1})aP_{k-1}(1, x, y, t)\Delta t + o(\Delta t) \quad (5.11)$$

Na koniec znajdziemy prawdopodobieństwo  $P\{\eta(t + \Delta t) = k, \nu(t + \Delta t) = 1, \xi_{(0)}^*(t + \Delta t) \in [x, x + dx), \xi_{(1)}^*(t + \Delta t) < \Delta t\}$ . Aby system znalazł się w oznaczonym stanie potrzeba i wystarcza, aby w chwili  $t$  było w systemie  $k$  procesów zewnętrznych i w ciągu czasu  $\Delta t$  do systemu przybył proces wewnętrzny. Prawdopodobieństwo tego zdarzenia wynosi  $P_k(0, x, t)dx c\Delta t + o(\Delta t)dx$ . Prawdopodobieństwo innych zdarzeń prowadzących do tego stanu wynosi  $o(\Delta t)dx$ . Z drugiej strony omawiane prawdopodobieństwo wynosi

$\int_0^{\Delta t} P_k(1, x, y, t + \Delta t)dxdy$ . Stąd wynika jeszcze jedno równanie:

$$\int_0^{\Delta t} P_k(1, x, y, t + \Delta t)dy = cP_k(0, x, t)\Delta t + o(\Delta t). \quad (5.12)$$

Dokonując przejść granicznych przy  $\Delta t \rightarrow 0$ , z wykorzystaniem reguły de l'Hospitala z równań (5.8)–(5.12) uzyskujemy następujący układ:

$$\frac{\partial P_0(t)}{\partial t} = -aP_0(t) + \int_0^t P_1(0, x, t)\mu_{(0)}(x)dx; \quad (5.13)$$

$$\frac{\partial P_k(0, x, t)}{\partial t} + \frac{\partial P_k(0, x, t)}{\partial x} = -(a + c + \mu_{(0)}(x))P_k(0, x, t) + \\ + \int_0^t P_k(1, x, y, t)\mu_{(1)}(y)dy + (1 - \delta_{k,1})aP_{k-1}(0, x, t), \quad k = 1, 2, \dots; \quad (5.14)$$

$$P_k(0, 0^+, t) = \int_0^t P_{k+1}(0, x, t)\mu_{(0)}(x)dx + \delta_{k,1}aP_0(t), \quad k = 1, 2, \dots; \quad (5.15)$$

$$\frac{\partial P_k(1, x, y, t)}{\partial t} + \frac{\partial P_k(1, x, y, t)}{\partial y} = -(a + \mu_{(1)}(y))P_k(1, x, y, t) + \\ + (1 - \delta_{k,1})aP_{k-1}(1, x, y, t), \quad k = 1, 2, \dots; \quad (5.16)$$

$$P_k(1, x, 0^+, t) = cP_k(0, x, t). \quad (5.17)$$

Dokonując odpowiednich przejść granicznych przy  $t \rightarrow \infty$  w równaniach (5.13)–(5.17) otrzymujemy następujące równania stacjonarne dla funkcji  $p_0, p_k(0, x), p_k(1, x, y)$ :

$$0 = -ap_0 + \int_0^\infty p_1(0, x)\mu_{(0)}(x)dx; \quad (5.18)$$

$$\frac{\partial p_k(0, x)}{\partial x} = -(a + c + \mu_{(0)}(x))p_k(0, x) +$$

$$+ \int_0^{\infty} p_k(1, x, y) \mu_{(1)}(y) dy + (1 - \delta_{k,1}) a p_{k-1}(0, x), \quad k = 1, 2, \dots; \quad (5.19)$$

$$p_k(0, 0^+) = \int_0^{\infty} p_{k+1}(0, x) \mu_{(0)}(x) dx + \delta_{k,1} a p_0, \quad k = 1, 2, \dots; \quad (5.20)$$

$$\begin{aligned} \frac{\partial p_k(1, x, y)}{\partial y} &= -(a + \mu_{(1)}(y)) p_k(1, x, y) + \\ &+ (1 - \delta_{k,1}) a p_{k-1}(1, x, y), \quad k = 1, 2, \dots; \end{aligned} \quad (5.21)$$

$$p_k(1, x, 0^+) = c p_k(0, x). \quad (5.22)$$

Aby rozwiązać powyższy układ wprowadźmy następujące funkcje tworzące:

$$p_{(0)}(z, x) = \sum_{k=1}^{\infty} p_k(0, x) z^k, \quad p_{(1)}(z, x, y) = \sum_{k=1}^{\infty} p_k(1, x, y) z^k.$$

Równanie z numerem  $k$  we wzorze (5.21) mnożymy przez  $z^k$  i sumujemy te równania dla wszystkich  $k$ , otrzymujemy stąd równanie:

$$\frac{\partial p_{(1)}(z, x, y)}{\partial y} = -(a - az + \mu_{(1)}(y)) p_{(1)}(z, x, y). \quad (5.23)$$

Rozwiązanie równania (5.23) (biorąc pod uwagę postać funkcji  $\mu_{(1)}(t)$ ) otrzymujemy w postaci następującej:

$$p_{(1)}(z, x, y) = [1 - H(y)] e^{-(a-az)y} p_{(1)}(z, x, 0). \quad (5.24)$$

Z równań (5.22) po pomnożeniu każdego z nich przez  $z^k$  i zsumowaniu po wszystkich  $k$  otrzymujemy natomiast równość:

$$p_{(1)}(z, x, 0) = c p_{(0)}(z, x),$$

skąd

$$p_{(1)}(z, x, y) = c [1 - H(y)] e^{-(a-az)y} p_{(0)}(z, x).$$

Teraz każde z równań (5.19) mnożymy przez  $z^k$  i sumujemy po wszystkich  $k$ . Stąd mamy równanie:

$$\frac{\partial p_{(0)}(z, x)}{\partial x} = -(a - az + c + \mu_{(0)}(x)) p_{(0)}(z, x) + \int_0^{\infty} p_{(1)}(z, x, y) \mu_{(1)}(y) dy. \quad (5.25)$$

Wykorzystując związek (5.24) zauważamy, że

$$\begin{aligned} \int_0^{\infty} p_{(1)}(z, x, y) \mu_{(1)}(y) dy &= \int_0^{\infty} c [1 - H(y)] e^{-(a-az)y} p_{(0)}(z, x) \mu_{(1)}(y) dy = \\ &= \int_0^{\infty} c p_{(0)}(z, x) b_{(1)}(y) e^{-(a-az)y} dy = c p_{(0)}(z, x) \int_0^{\infty} b_{(1)}(y) e^{-(a-az)y} dy = \end{aligned}$$



$$= ch(a - az)p_{(0)}(z, x).$$

Z powyższych obliczeń wynika, że równanie (5.25) może być przedstawione w postaci następującej:

$$\frac{\partial p_{(0)}(z, x)}{\partial x} = -(a - az + c + \mu_{(0)}(x) - ch(a - az))p_{(0)}(z, x). \quad (5.26)$$

Równanie (5.26) ma rozwiązanie

$$p_{(0)}(z, x) = [1 - B(x)]e^{-(a-az+c-ch(a-az))x}p_{(0)}(z, 0). \quad (5.27)$$

Z równania (5.20) po analogicznych przekształceniach przechodząc do FT otrzymujemy równanie:

$$p_{(0)}(z, 0) = z^{-1} \int_0^\infty p_{(0)}(z, x)\mu_{(0)}(x)dx - \int_0^\infty p_{(1)}(0, x)\mu_{(0)}(x)dx + azp_0.$$

Ponieważ z równania (5.18) wynika, że  $\int_0^\infty p_{(1)}(0, x)\mu_{(0)}(x)dx = ap_0$ , więc

$$p_{(0)}(z, 0) = z^{-1} \int_0^\infty p_{(0)}(z, x)\mu_{(0)}(x)dx - (a - az)p_0. \quad (5.28)$$

Ze wzorów (5.27)–(5.28) wynika, że

$$p_{(0)}(z, 0) = z^{-1}p_{(0)}(z, 0) \int_0^\infty b_{(1)}(x)e^{-(a-az+c-ch(a-az))x}dx - (a - az)p_0,$$

skąd łatwo otrzymujemy równość:

$$p_{(0)}(z, 0) = \frac{(a - az)p_0z}{\beta(a - az + c - ch(a - az)) - z}.$$

Wstawiając ten wynik do wzoru (5.27), otrzymujemy równość:

$$p_{(0)}(z, x) = \frac{(a - az)p_0z}{\beta(a - az + c - ch(a - az)) - z} [1 - B(x)]e^{-(a-az+c-ch(a-az))x}. \quad (5.29)$$

Natomiast z równań (5.29), (5.24) otrzymujemy:

$$\begin{aligned} p_{(1)}(z, x, y) &= \frac{c(a - az)p_0z}{\beta(a - az + c - ch(a - az)) - z} [1 - B(x)]e^{-(a-az+c-ch(a-az))x} \times \\ &\quad \times [1 - H(y)]e^{-(a-az)y}. \end{aligned} \quad (5.30)$$

### 5.3 Wyznaczenie charakterystyk objętości sumarycznej

Wyznamy teraz charakterystyki stacjonarnej objętości sumarycznej dla badanych schematów. Załóżmy, że dla badanego systemu istnieje tryb stacjonarny.

- **Schemat 1.** Dla schematu 1 na mocy wzoru na prawdopodobieństwo całkowite otrzymujemy w warunkach stacjonarnych oczywistą równość:

$$D(x) = p_0 + \sum_{k=1}^{\infty} \int_0^{\infty} P\{\sigma < x \mid \eta = k, \nu = 0, \xi_{(0)}^* = y\} p_k(0, y) dy + \\ + \sum_{k=1}^{\infty} \int_{y=0}^{\infty} \int_{u=0}^{\infty} P\{\sigma < x \mid \eta = k, \nu = 1, \xi_{(0)}^* = y, \xi_{(1)}^* = u\} p_k(1, y, u) dy du. \quad (5.31)$$

Oznaczmy przez  $A * B(x)$  spłot według Stieltjesa dystrybuant  $A(x)$  i  $B(x)$ , a przez  $A_*^{(n)}(x)$  – spłot  $n$ -tego rzędu dystrybuanty  $A(x)$ ,  $n = 0, 1, \dots$ , tzn.

$$A_*^{(0)}(x) \equiv 1, \quad A_*^{(n)}(x) = \int_0^x A_*^{(n-1)}(x-u) dA(u), \quad n = 1, 2, \dots$$

Przy przyjętych oznaczeniach są jasne poniższe równości:

$$P\{\sigma < x \mid \eta = k, \nu = 0, \xi_{(0)}^* = y\} = L_*^{(k-1)} * E_y^{(0)}(x),$$

gdzie, jak wiadomo z lematu 4.1,  $E_y^{(0)}(x) = \frac{L(x) - F(x, y)}{1 - B(y)}$ . Analogicznie otrzymujemy

$$P\{\sigma < x \mid \eta = k, \nu = 1, \xi_{(0)}^* = y, \xi_{(1)}^* = u\} = L_*^{(k-1)} * E_y^{(0)} * E_u^{(1)}(x),$$

gdzie  $E_u^{(1)}(x) = \frac{R(x) - \Phi(x, u)}{1 - H(u)}$ .

Zatem dystrybuantę objętości sumarycznej w warunkach stacjonarnych możemy przedstawić w postaci:

$$D(x) = p_0 + \sum_{k=1}^{\infty} \int_0^{\infty} L_*^{(k-1)} * E_y^{(0)}(x) p_k(0, y) dy + \\ + \sum_{k=1}^{\infty} \int_{y=0}^{\infty} \int_{u=0}^{\infty} L_*^{(k-1)} * E_y^{(0)} * E_u^{(1)}(x) p_k(1, y, u) dy du. \quad (5.32)$$

Przechodząc do PLS, otrzymujemy

$$\begin{aligned} \delta(s) &= p_0 + \sum_{k=1}^{\infty} \int_0^{\infty} (\varphi(s))^{k-1} e_y^{(0)}(s) p_k(0, y) dy + \\ &+ \sum_{k=1}^{\infty} \int_{y=0}^{\infty} \int_{u=0}^{\infty} (\varphi(s))^{k-1} e_y^{(0)}(s) e_u^{(1)}(s) p_k(1, y, u) dy du, \end{aligned} \quad (5.33)$$

gdzie

$$e_y^{(0)}(s) = [1 - B(y)]^{-1} \int_{x=0}^{\infty} e^{-sx} \int_{w=y}^{\infty} dF(x, w),$$

oraz

$$e_u^{(1)}(s) = [1 - H(u)]^{-1} \int_{x=0}^{\infty} e^{-sx} \int_{w=u}^{\infty} d\Phi(x, w).$$

We wzorze (5.33) występują dwie sumy, które można łatwo wyznaczyć wykorzystując wzory (5.29)–(5.30) oraz wykonując odpowiednie przekształcenia:

$$\begin{aligned} S_1 &= \sum_{k=1}^{\infty} \int_0^{\infty} (\varphi(s))^{k-1} e_y^{(0)}(s) p_k(0, y) dy = \\ &= (\varphi(s))^{-1} \int_0^{\infty} e_y^{(0)}(s) \sum_{k=1}^{\infty} p_k(0, y) (\varphi(s))^k dy = \\ &= (\varphi(s))^{-1} \int_0^{\infty} p_{(0)}(\varphi(s), y) e_y^{(0)}(s) dy. \end{aligned}$$

Stąd dalej po odpowiednich przekształceniach oraz wprowadzając oznaczenie  $\kappa(z) = a - az + c - ch(a - az)$  otrzymujemy:

$$S_1 = \frac{p_0(a - a\varphi(s))[\varphi(s) - \alpha(s, \kappa(\varphi(s)))]}{\kappa((\varphi(s))[\beta(\kappa(\varphi(s))) - \varphi(s)]}.$$

W analogiczny sposób otrzymujemy:

$$S_2 = \frac{p_0 c [\varphi(s) - \alpha(s, \kappa(\varphi(s)))] [r(s) - g(s, a - a\varphi(s))]}{\kappa(\varphi(s)) [\beta(\kappa(\varphi(s))) - \varphi(s)]}.$$

Z otrzymanego wzoru można uzyskać wzory na pierwsze dwa momenty objętości sumarycznej w warunkach stacjonarnych:

$$\delta_1 = -\delta'(0) = a(\alpha_{11}(1 + ch_1) + c\beta_1 g_{11}) + \frac{a^2 \varphi_1 [\beta_2(1 + ch_1)^2 + c\beta_1 h_2]}{2(1 - a\beta_1(1 + ch_1))} \quad (5.34).$$

$$\delta_2 = \delta''(0) = \frac{a^3 \beta_1^3 c g_{21} (1 + ch_1)^2}{(1 - \rho)^2} - \frac{2a^3 \alpha_{12} \beta_1 \varphi_1 (1 + ch_1)^3}{(1 - \rho)^2} + \frac{a^4 \alpha_{12} \beta_1^2 \varphi_1 (1 + ch_1)^4}{(1 - \rho)^2} -$$

$$\begin{aligned}
& - \frac{a^4 \alpha_{11} \beta_1 \beta_2 \varphi_1 (1 + ch_1)^4}{(1 - \rho)^2} + \frac{a^4 \beta_2^2 \varphi_1^2 (1 + ch_1)^4}{2(1 - \rho)^2} - \frac{a^4 \beta_1 \beta_3 \varphi_1^2 (1 + ch_1)^4}{3(1 - \rho)^2} + \\
& + \frac{a(\alpha_{21}(1 + ch_1) + 2\alpha_{11}cg_{11} + \beta_1cg_{21})}{(1 - \rho)^2} - \frac{a^3 \beta_1 \beta_2 (1 + ch_1)^2 (\varphi_2(1 + ch_1) + 2c\varphi_1g_{11})}{2(1 - \rho)^2} + \\
& + \frac{a^3 \varphi_1^2 (1 + ch_1)(\beta_3(1 + ch_1)^2 + 3\beta_2ch_2)}{3(1 - \rho)^2} + \frac{a^4 \beta_1^3 c \varphi_1 (1 + ch_1)(g_{12}(1 + ch_1) - cg_{11}h_2)}{(1 - \rho)^2} + \\
& + \frac{a^4 \beta_1^2 c^2 \varphi_1^2 h_2^2}{2(1 - \rho)^2} + \frac{a^3 \alpha_{11}(1 + ch_1)(\beta_2 \varphi_1 (1 + ch_1)^2 + \beta_1 c(2\beta_1(g_{11}(1 + ch_1)) - \varphi_1 h_2))}{(1 - \rho)^2} - \\
& - \frac{a^2(-\beta_2 \varphi_2 + 8\alpha_{11} \beta_1 cg_{11} + 4\alpha_{21} \beta_1 (1 + ch_1)^2 - 2\alpha_{12} \varphi_1 (1 + ch_1)^2)}{2(1 - \rho)^2} + \\
& + \frac{a^2 c(-4\beta_1^2 g_{21}(1 + ch_1) + \beta_2(\varphi_2 h_1(2 + ch_1) + 2\varphi_1 g_{11}(1 + ch_1)) + 2\alpha_{11} \varphi_1 h_2)}{2(1 - \rho)^2} + \\
& + \frac{a^2 c \beta_1 (2\varphi_1 g_{12} - 8\alpha_{11} cg_{11} h_1 + \varphi_2 h_2)}{2(1 - \rho)^2} + \\
& + \frac{a^3 \beta_1^2 (2\alpha_{21}(1 + ch_1)^3 - c(\varphi_2(1 + ch_1)h_2 + \varphi_1(4g_{12}(1 + ch_1) - 2cg_{11}h_2)))}{2(1 - \rho)^2} + \frac{a^3 \beta_1 c \varphi_1^2 h_3}{3(1 - \rho)}
\end{aligned} \tag{5.35}$$

- **Schemat 2.** Oznaczmy przez  $\eta_1(t)$  liczbę procesów wewnętrznych przybyłych od początku obsługi procesu zewnętrznego do chwili  $t$ . Jest jasne, że  $P\{\eta_1(y) = l\} = \frac{(cy)^l}{l!} e^{-cy}$ . Wówczas na mocy wzoru na prawdopodobieństwo całkowite dla schematu 2 otrzymujemy następującą równość:

$$\begin{aligned}
D(x) &= p_0 + \sum_{k=1}^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} P\{\sigma < x \mid \eta = k, \eta_1(y) = l, \nu = 0, \xi_{(0)}^* = y\} \times \\
&\quad \times P\{\eta_1(y) = l\} p_k(0, y) dy + \\
&+ \sum_{k=1}^{\infty} \int_0^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} P\{\sigma < x \mid \eta = k, \eta_1(y) = l, \nu = 1, \xi_{(0)}^* = y, \xi_{(1)}^* = u\} \times \\
&\quad \times P\{\eta_1(y) = l\} p_k(1, y, u) dy du.
\end{aligned} \tag{5.36}$$

Zauważmy, że

$$P\{\sigma < x \mid \eta = k, \eta_1(y) = l, \nu = 0, \xi_{(0)}^* = y\} = L_*^{(k-1)} * E_y^{(0)} * R_*^{(l)}(x)$$

oraz

$$\begin{aligned} P\{\sigma < x \mid \eta = k, \eta_1(y) = l, \nu = 1, \xi_{(0)}^* = y, \xi_{(1)}^* = u\} = \\ = L_*^{(k-1)} * E_y^{(0)} * E_u^{(1)} * R_*^{(l)}(x). \end{aligned}$$

Stąd wyrażenie (5.36) przyjmuje postać:

$$\begin{aligned} D(x) = p_0 + \sum_{k=1}^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} L_*^{(k-1)} * E_y^{(0)} * R_*^{(l)}(x) \frac{(cy)^l}{l!} e^{-cy} p_k(0, y) dy + \\ + \sum_{k=1}^{\infty} \int_0^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} L_*^{(k-1)} * E_y^{(0)} * E_u^{(1)} * R_*^{(l)}(x) \frac{(cy)^l}{l!} e^{-cy} p_k(1, y, u) dy du. \end{aligned} \quad (5.37)$$

Przechodząc we wzorze (5.37) do PLS, otrzymujemy po obliczeniach:

$$\begin{aligned} \delta(s) = p_0 + \frac{1}{\varphi(s)} \int_0^{\infty} e^{-(1-r(s))cy} p_{(0)}(\varphi(s), y) e_y^{(0)}(s) dy + \\ + \frac{1}{r(s)\varphi(s)} \int_0^{\infty} \int_0^{\infty} e^{-(1-r(s))cy} e_y^{(0)}(s) e_u^{(1)}(s) p_{(1)}(\varphi(s), y, u) dy du. \end{aligned} \quad (5.38)$$

Wykorzystując wzory (5.29)–(5.30) oraz wykonując odpowiednie przekształcenia, otrzymujemy ostatecznie:

$$\begin{aligned} \delta(s) = p_0 \left\{ 1 + \frac{\varphi(s) - \alpha(s, c - cr(s) + \kappa(\varphi(s)))}{\beta(\kappa(\varphi(s))) - \varphi(s)} \times \right. \\ \left. \times \frac{ar(s)(1 - \varphi(s)) + c(r(s) - g(s, a - a\varphi(s)))}{r(s)(c - cr(s) + \kappa(\varphi(s)))} \right\}. \end{aligned} \quad (5.39)$$

Z uzyskanego wzoru można otrzymać formuły pozwalające obliczać pierwsze dwa momenty objętości sumarycznej w warunkach stacjonarnych:

$$\begin{aligned} \delta_1 = a[\alpha_{11}(1 + ch_1) + c\beta_1(g_{11} - h_1r_1)] + \frac{ac\beta_2r_1(1 + ch_1)}{2} + \\ + \frac{a^2\varphi_1[\beta_2(1 + ch_1)^2 + c\beta_1h_2]}{2[1 - a\beta_1(1 + ch_1)]}. \end{aligned} \quad (5.40)$$

$$\begin{aligned} \delta_2 = \delta''(0) = \frac{a^3\beta_1^3cg_{21}(1 + ch_1)^2}{(1 - \rho)^2} - \frac{2a^3\alpha_{12}\beta_1\varphi_1(1 + ch_1)^3}{(1 - \rho)^2} + \frac{a^4\alpha_{12}\beta_1^2\varphi_1(1 + ch_1)^4}{(1 - \rho)^2} - \\ - \frac{a^4\alpha_{11}\beta_1\beta_2\varphi_1(1 + ch_1)^4}{(1 - \rho)^2} + \frac{a^4\beta_2^2\varphi_1^2(1 + ch_1)^4}{2(1 - \rho)^2} - \frac{a^4\beta_1\beta_3\varphi_1^2(1 + ch_1)^4}{3(1 - \rho)^2} + \frac{a^4\beta_1^2c^2\varphi_1^2h_2^2}{2(1 - \rho)^2} + \\ + \frac{a(\alpha_{21}(1 + ch_1) + 2\alpha_{11}cg_{11} + \beta_1cg_{21})}{(1 - \rho)^2} - \frac{a^3\beta_1\beta_2(1 + ch_1)^2(\varphi_2(1 + ch_1) + 2c\varphi_1g_{11})}{2(1 - \rho)^2} + \end{aligned}$$

$$\begin{aligned}
& + \frac{a^3 \varphi_1^2 (1 + ch_1) (\beta_3 (1 + ch_1)^2 + 3\beta_2 ch_2)}{3(1 - \rho)^2} + \frac{a^4 \beta_1^3 c \varphi_1 (1 + ch_1) (g_{12} (1 + ch_1) - c g_{11} h_2)}{(1 - \rho)^2} + \\
& + \frac{a^3 \alpha_{11} (1 + ch_1) (\beta_2 \varphi_1 (1 + ch_1)^2 + \beta_1 c (2\beta_1 (g_{11} (1 + ch_1)) - \varphi_1 h_2))}{(1 - \rho)^2} - \\
& - \frac{a^2 (-\beta_2 \varphi_2 + 8\alpha_{11} \beta_1 c g_{11} + 4\alpha_{21} \beta_1 (1 + ch_1)^2 - 2\alpha_{12} \varphi_1 (1 + ch_1)^2)}{2(1 - \rho)^2} + \\
& + \frac{a^2 c (-4\beta_1^2 g_{21} (1 + ch_1) + \beta_2 (\varphi_2 h_1 (2 + ch_1) + 2\varphi_1 g_{11} (1 + ch_1)) + 2\alpha_{11} \varphi_1 h_2)}{2(1 - \rho)^2} + \\
& + \frac{a^2 c \beta_1 (2\varphi_1 g_{12} - 8\alpha_{11} c g_{11} h_1 + \varphi_2 h_2)}{2(1 - \rho)^2} - \frac{a^2 \beta_3 c (1 + ch_1)^2 r_1 (-2\varphi_1 + \beta_1 c r_1)}{3(1 - \rho)} + \\
& + \frac{a^3 \beta_1^2 (2\alpha_{21} (1 + ch_1)^3 - c (\varphi_2 (1 + ch_1) h_2 + \varphi_1 (4g_{12} (1 + ch_1) - 2c g_{11} h_2)))}{2(1 - \rho)^2} + \\
& + \alpha_{12} c (1 + ch_1) r_1 + \frac{a^3 c \varphi_1 ((3\beta_2^2 - 4\beta_1 \beta_3) (1 + ch_1)^3 + 6\beta_1^2 h_2) r_1}{6(1 - \rho)} - \\
& + \frac{a c r_1 (3\beta_2 c g_{11} - 6\alpha_{11} h_1 + c (\beta_3 (1 + ch_1) - 3\beta_2 h_1) r_1 - 6\beta_1 (g_{11} - h_1 r_1))}{3(1 - \rho)} - \\
& - \frac{a^2 c r_1 (\beta_2 \varphi_1 (2h_1 (1 + ch_1) - ch_2) + 4\beta_1^2 (1 + ch_1) (-g_{11} + h_1 r_1))}{2(1 - \rho)} - \\
& - \frac{2a^2 c r_1 \beta_1 (-2\alpha_{11} h_1 (1 + ch_1) + \varphi_1 h_2 + c \beta_2 (1 + ch_1) (g_{11} - h_1 r_1))}{2(1 - \rho)} + \\
& + \frac{1}{2} a c (\beta_2 (1 + ch_1) - 2\beta_1 h_1) r_2 + \frac{a^3 \beta_1 c \varphi_1^2 h_3}{3(1 - \rho)}. \tag{5.41}
\end{aligned}$$

Jak widać, otrzymane wyniki analityczne są dość skomplikowane, a ich wyrażenie w terminach PLS powoduje duże trudności obliczeniowe już na poziomie wyznaczania momentów objętości sumarycznej w warunkach stacjonarnych. W obu schematach wartość  $p_0$  można wyznaczyć z warunku  $\delta(0) = 1$ . Jest również oczywiste, że wstawiając w badanym systemie  $c = 0$ , otrzymamy model zwykłego systemu  $M/G/1/\infty$ .

## 5.4 Realizacje techniczne analizowanego modelu

Rozważany w tym rozdziale model jest uogólnieniem modelu  $M/G/1/\infty$  z czasem obsługi zależnym od objętości zgłoszenia i nieograniczoną objętością sumaryczną. Model tego typu był po raz pierwszy analizowany w pracy [16]. Model analizowany w tym rozdziale różni się jednak tym, że zgłoszenia (procesy) są podzielone na dwa typy procesów o różnych priorytetach obsługi. Praktyczne zastosowanie takich modeli można znaleźć w systemach, w których uruchamiane procesy można podzielić na dwa typy, z których jeden ma większy priorytet obsługi. Oto przykład takiego systemu.

- Model współdzielonego procesora z przełączaniem zadań

Założmy, że mamy dany system komputerowy składający się z jednego komputera wyposażonego w jeden procesor. Rozmiar pamięci operacyjnej jest ograniczony wielkością  $V$ . Użytkownik obsługujący komputer uruchamia procesy (aplikacje) na komputerze. Podczas obsługi takich aplikacji procesor może inicjalizować wykonanie dodatkowych procesów, których priorytet jest większy i które są potrzebne do obsługi procesu inicjalizowanego przez użytkownika. W trakcie obsługi procesów inicjalizowanych przez procesor obsługa procesu użytkownika jest zatrzymywana i kontynuowana dopiero po zakończeniu obsługi aplikacji inicjalizowanej przez procesor. Praca procesora polega więc na przełączaniu zadań pomiędzy obsługą procesów użytkownika a obsługą procesów dodatkowych inicjalizowanych przez ten procesor. Omawiana realizacja techniczna może być analizowana w oparciu o wyniki z punktów 5.2–5.3, a głównym celem jest wyznaczenie charakterystyk utrat informacji. Takie charakterystyki mogą być potem wykorzystane do odpowiedniego doboru rozmiaru pamięci współdzielonej procesora (pamięci RAM). Podstawowe założenia takiego modelu są opisane w pracy [19].

## 5.5 Wyznaczenie charakterystyk utrat informacji w szczególnych przypadkach

Wykorzystując uzyskane w poprzednim punkcie wzory na obliczanie pierwszych dwóch momentów objętości sumarycznej możemy oszacować charakterystyki utrat informacji w systemie analogicznym, lecz z ograniczoną objętością sumaryczną. Wykorzystamy w tym celu nierówności (4.15) oraz (4.16) oraz aproksymacje występujących w tych wzorach funkcji z wykorzystaniem funkcji gamma.

Poniżej zebrano wyniki analityczne oraz symulacyjne dla pewnych przypadków szczególnych analizowanego systemu. Aproksymacje prawdopodobieństw utrat zostały wyznaczone w oparciu o obliczenia wykonane w środowisku **Mathematica 8.0**. Natomiast w rozdziale **Dodatki** (punkt 7.1 oraz 7.2) znajdują się omówienia symulacyjnych programów komputerowych, w oparciu o które (lub ich pewne modyfikacje) uzyskano wyniki dla wszystkich danych zawartych w poniższych tabelach.

### 5.5.1 Schemat 1. Rozmiar procesów i czas obsługi procesów niezależne.

Założmy, że rozmiary procesów zewnętrznych i wewnętrznych mają rozkład wykładniczy z tym samym parametrem  $f$ , czasy obsługi procesów wewnętrznych i zewnętrznych nie zależą od rozmiarów tych procesów i mają rozkład wykładniczy z tym samym parametrem  $\mu$ . Wówczas pierwsze dwa momenty objętości sumarycznej procesów w warunkach stacjonarnych, w oparciu o wzory (5.34)–(5.35), okazują się równe:

$$\delta_1 = \frac{a(ac^2 - \mu^2(2c + \mu))}{f\mu^2(a(c + \mu) - \mu^2)}. \quad (5.42)$$

$$\delta_2 = \frac{2a(a^3c^2 - 3ac^2\mu^2 + a^2c^2(c + \mu) + \mu^4(3c + \mu))}{f^2\mu^2(\mu^2 - a(c + \mu))^2}. \quad (5.43)$$

Wykorzystując wzory (5.42)–(5.43) możemy oszacować charakterystyki utrat informacji w oparciu o wzory (4.15)–(4.16) i wykorzystując aproksymacje funkcją gamma. W tabeli 5.1 zebrane są wyniki przy następujących parametrach:  $a = 1$ ,  $f = 1$ ,  $c = 1$ ,  $\mu = 2$  (wówczas  $\rho = 0.75$ ), natomiast w tabeli 5.2 - przy parametrach  $a = 1$ ,  $f = 1$ ,  $c = 1$ ,  $\mu = 10$  (wówczas  $\rho = 0.11$ ). Odpowiednie wykresy odpowiadające tym obliczeniom są przedstawione na rysunkach 5.1-5.2. Dodatkowo w tym szczególnym przypadku można wyznaczyć dystrybucję objętości sumarycznej procesów w warunkach stacjonarnych w postaci jawnej. Obliczając oszacowania charakterystyk utrat informacji nie musimy więc wykorzystywać aproksymacji funkcją gamma. Wyniki



uzyskane z wykorzystaniem pakietu Mathematica pokazują jednak, że wyniki uzyskane bez użycia aproksymacji różnią się nieznacznie od tych uzyskanych z jej wykorzystaniem.

**Tabela 5.1.** Charakterystyki utrat w przypadku  $\rho = 0.75$  (czas obsługi i rozmiar procesu niezależne)

$V$	$p_u^{zSIM}$	$Q^{zSIM}$	$p_u^{wSIM}$	$Q^{wSIM}$	$p_u^{SIM}$	$Q^{SIM}$	$p^*$	$Q^*$
1	0.452853	0.787778	0.607895	0.885532	<b>0.486190</b>	<b>0.808800</b>	<b>0.806592</b>	<b>0.921361</b>
2	0.269524	0.554872	0.376589	0.684678	<b>0.298196</b>	<b>0.589634</b>	<b>0.653007</b>	<b>0.785886</b>
3	0.182746	0.398013	0.242533	0.503171	<b>0.200082</b>	<b>0.428488</b>	<b>0.529224</b>	<b>0.653506</b>
4	0.131815	0.293061	0.162885	0.365588	<b>0.141220</b>	<b>0.314994</b>	<b>0.429155</b>	<b>0.537962</b>
5	0.098522	0.221301	0.114275	0.267471	<b>0.103417</b>	<b>0.235642</b>	<b>0.348141</b>	<b>0.440650</b>
6	0.075176	0.169841	0.083108	0.199124	<b>0.077680</b>	<b>0.179100</b>	<b>0.282497</b>	<b>0.359850</b>
7	0.057729	0.131257	0.061765	0.149944	<b>0.059022</b>	<b>0.137248</b>	<b>0.229279</b>	<b>0.293235</b>
8	0.045037	0.102888	0.046634	0.113398	<b>0.045553</b>	<b>0.106280</b>	<b>0.186116</b>	<b>0.238561</b>
9	0.035535	0.081065	0.036268	0.088430	<b>0.035774</b>	<b>0.083461</b>	<b>0.151099</b>	<b>0.193829</b>
10	0.028031	0.064169	0.028434	0.069306	<b>0.028163</b>	<b>0.065850</b>	<b>0.122684</b>	<b>0.157320</b>

**Tabela 5.2.** Charakterystyki utrat w przypadku  $\rho = 0.11$  (czas obsługi i rozmiar procesu niezależne)

$V$	$p_u^{zSIM}$	$Q^{zSIM}$	$p_u^{wSIM}$	$Q^{wSIM}$	$p_u^{SIM}$	$Q^{SIM}$	$p^*$	$Q^*$
1	0.382355	0.744932	0.586853	0.875514	<b>0.394261</b>	<b>0.752532</b>	<b>0.412734</b>	<b>0.753767</b>
2	0.152965	0.427311	0.323365	0.635248	<b>0.166234</b>	<b>0.443490</b>	<b>0.172234</b>	<b>0.439893</b>
3	0.062848	0.221799	0.167829	0.406640	<b>0.071830</b>	<b>0.237610</b>	<b>0.072045</b>	<b>0.229063</b>
4	0.026136	0.108728	0.083693	0.239918	<b>0.031240</b>	<b>0.120364</b>	<b>0.030172</b>	<b>0.112469</b>
5	0.010964	0.051543	0.040163	0.131806	<b>0.013593</b>	<b>0.058778</b>	<b>0.012645</b>	<b>0.053268</b>
6	0.004640	0.023906	0.018758	0.068750	<b>0.005913</b>	<b>0.027960</b>	<b>0.005302</b>	<b>0.024636</b>
7	0.001982	0.010947	0.008423	0.034153	<b>0.002566</b>	<b>0.013054</b>	<b>0.002224</b>	<b>0.011208</b>
8	0.000830	0.004850	0.003825	0.017043	<b>0.001102</b>	<b>0.005956</b>	<b>0.000933</b>	<b>0.005040</b>
9	0.000346	0.002127	0.001731	0.008255	<b>0.000472</b>	<b>0.002684</b>	<b>0.000392</b>	<b>0.002247</b>
10	0.000152	0.000969	0.000688	0.003357	<b>0.000201</b>	<b>0.001185</b>	<b>0.000164</b>	<b>0.000953</b>

### 5.5.2 Schemat 1. Czas obsługi proporcjonalny do rozmiaru procesu.

Załóżmy teraz, że rozmiary procesów zewnętrznych i wewnętrznych mają rozkład wykładniczy z tym samym parametrem  $f$ , czasy obsługi procesów wewnętrznych i zewnętrznych są proporcjonalne do rozmiarów tych procesów z tym samym współczynnikiem proporcjonalności  $l$ . Wówczas pierwsze dwa momenty objętości sumarycznej procesów w warunkach stacjonarnych są równe:

$$\delta_1 = \frac{al(2f^3 - 3acfl^2 - 3ac^2l^3 + f^2l(4c - a))}{f^3(f^2 - al(f + cl))}. \quad (5.44)$$

$$\begin{aligned} \delta_2 = & \frac{2(4a^2 + 6ac - 3c^2)}{c^2f^2} - \frac{2a(3a + 2c)l}{cf^3} + \frac{6a(a + 2c)l^2}{f^4} + \frac{6a^2cl^3}{f^5} + \\ & + \frac{2a^2c^2l^4}{f^6} + \frac{2(a^3f^2 + 3a^2cf^2 - ac^2f^2 + c^3f^2 - a^4fl - 4a^3cfl - a^2c^2fl + 2ac^3fl)}{c^3(ac^2l^2 + afl - f^2)^2} - \\ & - \frac{2(-a^3f - 7a^2cf - 5ac^2f + 2c^3f + 5a^2cl + 11a^2c^2l - ac^3l)}{c^3f(ac^2l^2 + afl - f^2)}. \end{aligned} \quad (5.45)$$

Wykorzystując wzory (5.44)–(5.45) możemy oszacować charakterystyki utrat informacji w oparciu o wzory (4.15)–(4.16) i wykorzystując aproksymację funkcją gamma. W tabeli 5.3 zebrane są wyniki przy następujących parametrach:  $a = 1$ ,  $f = 2$ ,  $c = 1$ ,  $l = 1$  (wówczas  $\rho = 0.75$ ), natomiast w tabeli 5.4 - przy parametrach  $a = 0.2$ ,  $f = 1$ ,  $c = 1$ ,  $l = 1$  (wówczas  $\rho = 0.4$ ). Odpowiednie wykresy są przedstawione na rysunkach 5.3–5.4.

**Tabela 5.3.** Charakterystyki utrat w przypadku  $\rho = 0.75$  (czas obsługi proporcjonalny do rozmiaru procesu)

$V$	$p_u^{zSIM}$	$Q^{zSIM}$	$p_u^{wSIM}$	$Q^{wSIM}$	$p_u^{SIM}$	$Q^{SIM}$	$p^*$	$Q^*$
1	0.230317	0.507360	0.459095	0.758919	<b>0.275519</b>	<b>0.557043</b>	<b>0.735541</b>	<b>0.819277</b>
2	0.120311	0.268125	0.189841	0.409397	<b>0.138928</b>	<b>0.305987</b>	<b>0.517601</b>	<b>0.619025</b>
3	0.074179	0.165265	0.083015	0.202785	<b>0.076781</b>	<b>0.176320</b>	<b>0.359480</b>	<b>0.452767</b>
4	0.046486	0.104892	0.041110	0.105866	<b>0.044824</b>	<b>0.105193</b>	<b>0.247887</b>	<b>0.320561</b>
5	0.029667	0.067813	0.022257	0.058507	<b>0.027311</b>	<b>0.064854</b>	<b>0.170155</b>	<b>0.221470</b>
6	0.018985	0.043824	0.013111	0.034746	<b>0.017085</b>	<b>0.408879</b>	<b>0.116422</b>	<b>0.150412</b>
7	0.012359	0.028672	0.008051	0.021568	<b>0.010951</b>	<b>0.026350</b>	<b>0.079466</b>	<b>0.100918</b>
8	0.008070	0.018601	0.005068	0.013457	<b>0.007082</b>	<b>0.016907</b>	<b>0.054140</b>	<b>0.067110</b>
9	0.005230	0.012086	0.003252	0.008769	<b>0.004576</b>	<b>0.010990</b>	<b>0.036830</b>	<b>0.044319</b>
10	0.003430	0.007997	0.002088	0.005658	<b>0.002985</b>	<b>0.007222</b>	<b>0.025024</b>	<b>0.029110</b>

**Tabela 5.4.** Charakterystyki utrat w przypadku  $\rho = 0.4$  (czas obsługi proporcjonalny do rozmiaru procesu)

$V$	$p_u^{zSTM}$	$Q^{zSTM}$	$p_u^{wSTM}$	$Q^{wSTM}$	$p_u^{STM}$	$Q^{STM}$	$p^*$	$Q^*$
1	0.386072	0.746121	0.696282	0.927971	<b>0.448979</b>	<b>0.782985</b>	<b>0.639594</b>	<b>0.838745</b>
2	0.183267	0.456355	0.457276	0.759278	<b>0.279769</b>	<b>0.563058</b>	<b>0.435453</b>	<b>0.620079</b>
3	0.110608	0.286556	0.285607	0.562728	<b>0.183462</b>	<b>0.401597</b>	<b>0.300745</b>	<b>0.451965</b>
4	0.077060	0.193465	0.172150	0.387522	<b>0.119530</b>	<b>0.280101</b>	<b>0.209264</b>	<b>0.329809</b>
5	0.055986	0.137401	0.101341	0.254020	<b>0.076983</b>	<b>0.191413</b>	<b>0.146296</b>	<b>0.239570</b>
6	0.040616	0.098854	0.059379	0.160836	<b>0.049506</b>	<b>0.128218</b>	<b>0.102611</b>	<b>0.172098</b>
7	0.028658	0.070468	0.034871	0.100619	<b>0.031650</b>	<b>0.084987</b>	<b>0.072145</b>	<b>0.121862</b>
8	0.020382	0.050887	0.020970	0.063151	<b>0.020668</b>	<b>0.056856</b>	<b>0.050820</b>	<b>0.085002</b>
9	0.013828	0.035270	0.012657	0.039778	<b>0.013253</b>	<b>0.037483</b>	<b>0.035852</b>	<b>0.058456</b>
10	0.009531	0.024465	0.007819	0.024813	<b>0.008686</b>	<b>0.024637</b>	<b>0.025324</b>	<b>0.039690</b>

### 5.5.3 Schemat 2. Rozmiar procesów i czas obsługi procesów niezależne.

Tutaj, podobnie jak w punkcie 5.5.1 rozmiary procesów zewnętrznych i wewnętrznych mają rozkład wykładniczy z tym samym parametrem  $f$ , czasy obsługi procesów wewnętrznych i zewnętrznych nie zależą od rozmiarów tych procesów i mają rozkład wykładniczy z tym samym parametrem  $\mu$ . Wówczas pierwsze dwa momenty objętości sumarycznej procesów w warunkach stacjonarnych w oparciu o wzory (5.40)–(5.41) okazują się równe:

$$\delta_1 = \frac{a(ac^2(c+2\mu) - \mu^2(c+\mu)^2)}{f\mu^3(a(c+\mu) - \mu^2)}. \quad (5.46)$$

$$\begin{aligned} \delta_2 = & \frac{2a^2c^3}{f^2\mu^5} + \frac{2(2a^2c^2 + ac^3)}{f^2\mu^4} + \frac{2(a^2c + 2ac^2)}{f^2\mu^3} - \frac{2(a^2 - ac + c^2)}{cf^2\mu} + \\ & + \frac{2(a^4 + 2a^3c + a^2c^2 + 4a^2c\mu)}{f^2(\mu^2 - a\mu - ac)^2}. \end{aligned} \quad (5.47)$$

Wykorzystując wzory (5.46)–(5.47) możemy oszacować charakterystyki utrat informacji w oparciu o wzory (4.15)–(4.16) i wykorzystując aproksymację funkcją gamma. W tabeli 5.5 zebrane są wyniki przy następujących parametrach:  $a = 1$ ,  $f = 1$ ,  $c = 1$ ,  $\mu = 2$  (wówczas  $\rho = 0.75$ ), natomiast w tabeli 5.6 - przy parametrach  $a = 1$ ,  $f = 1$ ,  $c = 1$ ,  $\mu = 10$  (wówczas  $\rho = 0.11$ ). Odpowiednie wykresy znajdują się na rysunkach 5.5-5.6.

**Tabela 5.5.** Charakterystyki utrat w przypadku  $\rho = 0.75$  (czas obsługi i rozmiar procesu niezależne)

$V$	$p_u^{zSTM}$	$Q^{zSTM}$	$p_u^{wSTM}$	$Q^{wSTM}$	$p_u^{STM}$	$Q^{STM}$	$p^*$	$Q^*$
1	0.458559	0.789596	0.636026	0.897100	<b>0.496369</b>	<b>0.812504</b>	<b>0.810891</b>	<b>0.922106</b>
2	0.282146	0.561112	0.424435	0.717782	<b>0.319753</b>	<b>0.602513</b>	<b>0.660002</b>	<b>0.788988</b>
3	0.198672	0.408275	0.293173	0.548283	<b>0.225692</b>	<b>0.448328</b>	<b>0.537763</b>	<b>0.659280</b>
4	0.149086	0.307425	0.209556	0.413523	<b>0.167140</b>	<b>0.339094</b>	<b>0.438422</b>	<b>0.545835</b>
5	0.115116	0.237308	0.154056	0.312771	<b>0.127067</b>	<b>0.260474</b>	<b>0.357571</b>	<b>0.449809</b>
6	0.090070	0.186122	0.115941	0.239022	<b>0.098159</b>	<b>0.202664</b>	<b>0.291711</b>	<b>0.369559</b>
7	0.071248	0.147318	0.089183	0.184750	<b>0.076937</b>	<b>0.159179</b>	<b>0.238032</b>	<b>0.302941</b>
8	0.056772	0.117740	0.068994	0.143555	<b>0.060688</b>	<b>0.126014</b>	<b>0.194262</b>	<b>0.247886</b>
9	0.045179	0.093876	0.054094	0.112780	<b>0.048059</b>	<b>0.099984</b>	<b>0.158563</b>	<b>0.202542</b>
10	0.036267	0.075403	0.043030	0.089488	<b>0.038466</b>	<b>0.079983</b>	<b>0.129438</b>	<b>0.165296</b>

**Tabela 5.6.** Charakterystyki utrat w przypadku  $\rho = 0.11$  (czas obsługi i rozmiar procesu niezależne)

$V$	$p_u^{zSTM}$	$Q^{zSTM}$	$p_u^{wSTM}$	$Q^{wSTM}$	$p_u^{STM}$	$Q^{STM}$	$p^*$	$Q^*$
1	0.382671	0.745045	0.593955	0.878851	<b>0.394957</b>	<b>0.752837</b>	<b>0.412742</b>	<b>0.757343</b>
2	0.154188	0.428408	0.336364	0.647135	<b>0.168415</b>	<b>0.445505</b>	<b>0.172625</b>	<b>0.439941</b>
3	0.063942	0.222798	0.181182	0.423925	<b>0.073963</b>	<b>0.239973</b>	<b>0.072406</b>	<b>0.229217</b>
4	0.027216	0.110552	0.094076	0.255567	<b>0.033149</b>	<b>0.123425</b>	<b>0.030413</b>	<b>0.112672</b>
5	0.011637	0.052627	0.047215	0.144799	<b>0.014838</b>	<b>0.060921</b>	<b>0.012785</b>	<b>0.053461</b>
6	0.005153	0.025077	0.023383	0.079082	<b>0.006800</b>	<b>0.029959</b>	<b>0.005378</b>	<b>0.024789</b>
7	0.002250	0.011589	0.010944	0.040354	<b>0.003039</b>	<b>0.014198</b>	<b>0.002631</b>	<b>0.011316</b>
8	0.001004	0.005302	0.005194	0.020561	<b>0.001385</b>	<b>0.006689</b>	<b>0.000953</b>	<b>0.005109</b>
9	0.000450	0.002448	0.002416	0.010113	<b>0.000628</b>	<b>0.003144</b>	<b>0.000401</b>	<b>0.002289</b>
10	0.000207	0.001156	0.001135	0.004942	<b>0.000291</b>	<b>0.001501</b>	<b>0.000169</b>	<b>0.001020</b>

#### 5.5.4 Schemat 2. Czas obsługi proporcjonalny do rozmiaru procesu.

Założmy teraz, analogicznie jak w punkcie 5.5.2, że rozmiary procesów zewnętrznych i wewnętrznych mają rozkład wykładniczy z tym samym parametrem  $f$ , czasy obsługi procesów wewnętrznych i zewnętrznych są proporcjonalne do rozmiarów tych procesów z tym samym współczynnikiem proporcjonalności  $l$ . Wówczas pierwsze dwa momenty objętości sumarycznej procesów w warunkach stacjonarnych są równe:

$$\delta_1 = \frac{al(2f^4 - 4ac^2fl^3 - ac^3l^4 + f^3(4cl - al) + cf^2(cl^2 - 3al^2))}{f^4(f^2 - afl - acl^2)}. \quad (5.48)$$

$$\delta_2 = \frac{6(a^2 + ac - c^2)}{c^2f^2} - \frac{2(2a^2 + c^2)l}{cf^3} + \frac{2a(2a + 5c)l^2}{f^4} + \frac{2ac(4a + 5c)l^3}{f^5} + \frac{2ac^2(3a + c)l^4}{f^6} + \frac{2a^2c^3l^5}{f^7} +$$

$$\begin{aligned}
& + \frac{2(a^3 f^2 + 3a^2 c f^2 - a c^2 f^2 + c^3 f^2 - a^4 f l - 4a^3 c f l - a^2 c^2 f l + 2a c^3 f l)}{c^3 (a c l^2 + a f l - f^2)^2} \\
& - \frac{2(-a^3 f - 6a^2 c f - 2a c^2 f + 2c^3 f + 4a^3 c l + 7a^2 c^2 l - 3a c^3 l + c^4 l)}{c^3 f (a c l^2 + a f l - f^2)}. \quad (5.49)
\end{aligned}$$

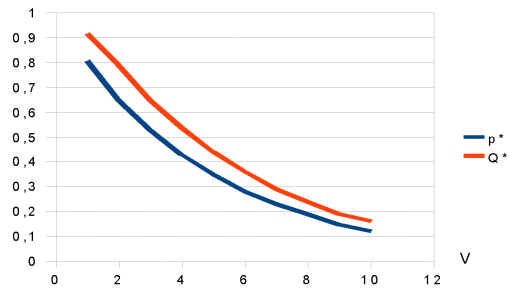
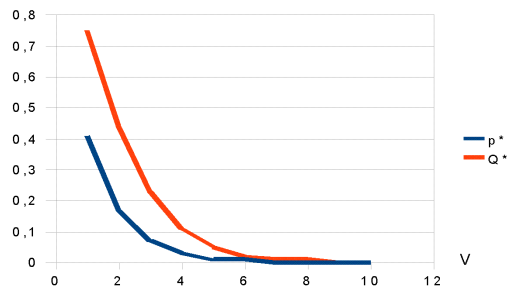
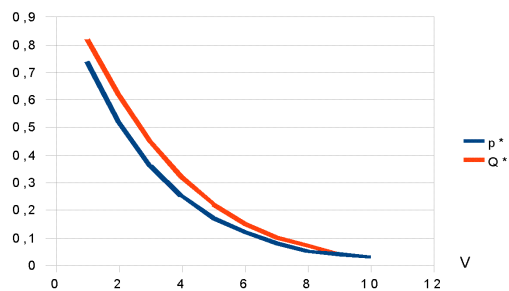
Wykorzystując wzory (5.48)–(5.49) możemy oszacować charakterystyki utrat informacji w oparciu o wzory (4.15)–(4.16) i wykorzystując aproksymację funkcją gamma. W tabeli 5.7 zebrane są wyniki przy parametrach:  $a = 1$ ,  $f = 2$ ,  $c = 1$ ,  $l = 1$  (wówczas  $\rho = 0.75$ ), natomiast w tabeli 5.8 - przy parametrach  $a = 0.2$ ,  $f = 1$ ,  $c = 1$ ,  $l = 1$  (wówczas  $\rho = 0.4$ ). Odpowiednie wykresy odpowiadające tym obliczeniom są przedstawione na rysunkach 5.7-5.8.

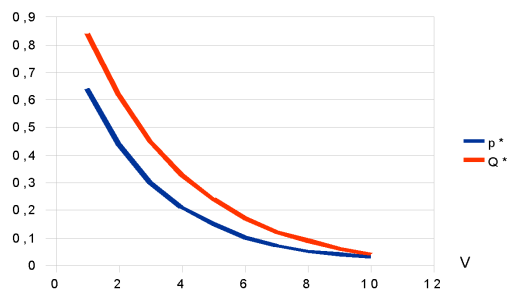
**Tabela 5.7.** Charakterystyki utrat w przypadku  $\rho = 0.75$  (czas obsługi proporcjonalny do rozmiaru procesu)

$V$	$p_u^{zSIM}$	$Q^{zSIM}$	$p_u^{wSIM}$	$Q^{wSIM}$	$p_u^{SIM}$	$Q^{SIM}$	$p^*$	$Q^*$
1	0.234189	0.509203	0.481052	0.771891	<b>0.282794</b>	<b>0.560929</b>	<b>0.737611</b>	<b>0.819722</b>
2	0.132378	0.276429	0.237101	0.455941	<b>0.160008</b>	<b>0.323828</b>	<b>0.523395</b>	<b>0.622074</b>
3	0.087134	0.175523	0.130131	0.258982	<b>0.099363</b>	<b>0.199246</b>	<b>0.367175</b>	<b>0.458839</b>
4	0.056386	0.112730	0.076447	0.151241	<b>0.062245</b>	<b>0.123985</b>	<b>0.255991</b>	<b>0.328440</b>
5	0.035695	0.071385	0.046517	0.090706	<b>0.038887</b>	<b>0.077082</b>	<b>0.177766</b>	<b>0.229769</b>
6	0.022425	0.044678	0.028439	0.054878	<b>0.024207</b>	<b>0.047701</b>	<b>0.123098</b>	<b>0.158167</b>
7	0.013903	0.027749	0.017358	0.033458	<b>0.014930</b>	<b>0.029446</b>	<b>0.085065</b>	<b>0.107633</b>
8	0.008562	0.017135	0.010774	0.020584	<b>0.009220</b>	<b>0.018161</b>	<b>0.058687</b>	<b>0.072626</b>
9	0.005295	0.010639	0.006625	0.012569	<b>0.005691</b>	<b>0.011214</b>	<b>0.040436</b>	<b>0.048687</b>
10	0.003230	0.006500	0.004018	0.007611	<b>0.003465</b>	<b>0.006831</b>	<b>0.027831</b>	<b>0.032471</b>

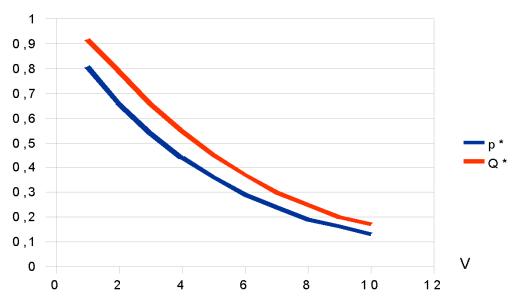
**Tabela 5.8.** Charakterystyki utrat w przypadku  $\rho = 0.4$  (czas obsługi proporcjonalny do rozmiaru procesu)

$V$	$p_u^{zSIM}$	$Q^{zSIM}$	$p_u^{wSIM}$	$Q^{wSIM}$	$p_u^{SIM}$	$Q^{SIM}$	$p^*$	$Q^*$
1	0.386749	0.746491	0.705541	0.931002	<b>0.451247</b>	<b>0.783797</b>	<b>0.624345</b>	<b>0.837731</b>
2	0.186496	0.458207	0.498088	0.783852	<b>0.295993</b>	<b>0.572631</b>	<b>0.439787</b>	<b>0.617416</b>
3	0.117262	0.290308	0.357662	0.623320	<b>0.217090</b>	<b>0.428499</b>	<b>0.317556</b>	<b>0.451354</b>
4	0.086468	0.199235	0.260673	0.480501	<b>0.163974</b>	<b>0.324422</b>	<b>0.232246</b>	<b>0.334968</b>
5	0.067651	0.146988	0.191564	0.362510	<b>0.124716</b>	<b>0.246185</b>	<b>0.171225</b>	<b>0.251625</b>
6	0.053620	0.112699	0.142032	0.271695	<b>0.095159</b>	<b>0.187458</b>	<b>0.126945</b>	<b>0.189874</b>
7	0.042185	0.087528	0.105450	0.201834	<b>0.072337</b>	<b>0.141991</b>	<b>0.094508</b>	<b>0.143102</b>
8	0.033053	0.068647	0.078933	0.151035	<b>0.055184</b>	<b>0.108362</b>	<b>0.070586</b>	<b>0.107388</b>
9	0.025454	0.052695	0.059702	0.113626	<b>0.042107</b>	<b>0.082328</b>	<b>0.052855</b>	<b>0.080145</b>
10	0.019606	0.040571	0.045329	0.085394	<b>0.032197</b>	<b>0.062512</b>	<b>0.039661</b>	<b>0.059476</b>

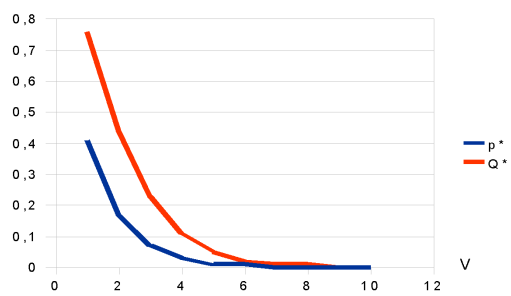
Rysunek 5.1. Prawdopodobieństwa utrat w przypadku  $\rho = 0.75$ Rysunek 5.2. Prawdopodobieństwa utrat w przypadku  $\rho = 0.11$ Rysunek 5.3. Prawdopodobieństwa utrat w przypadku  $\rho = 0.75$



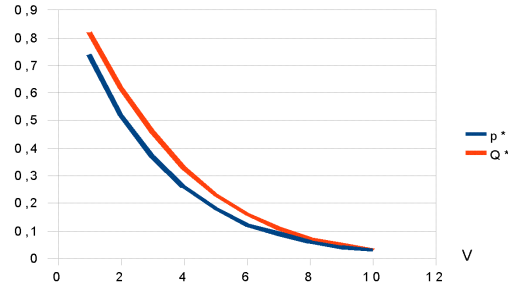
Rysunek 5.4. Prawdopodobieństwa utrat w przypadku  $\rho = 0.4$



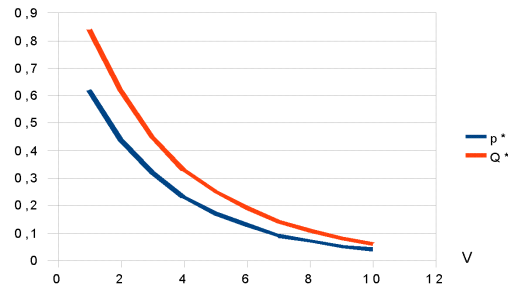
Rysunek 5.5. Prawdopodobieństwa utrat w przypadku  $\rho = 0.75$



Rysunek 5.6. Prawdopodobieństwa utrat w przypadku  $\rho = 0.11$



Rysunek 5.7. Prawdopodobieństwa utrat w przypadku  $\rho = 0.75$



Rysunek 5.8. Prawdopodobieństwa utrat w przypadku  $\rho = 0.4$

Uzyskane wyniki pokazują, że omawiana aproksymacja jest korzystna jedynie w przypadku małych prawdopodobieństw utrat, jeśli charakterystyki utrat są duże, wówczas aproksymacja pokazuje dużo większe wartości charakterystyk utrat (porównanie  $p^*$ ,  $Q^*$  oraz  $p_u^{SIM}$ ,  $Q^{SIM}$ ), co prowadzi do sytuacji, w której wybieramy potrzebny rozmiar pamięci z dużym nadmiarem. Możemy zatem w przypadku małych prawdopodobieństw utrat stosować wspomniane oszacowania dla wyznaczenia charakterystyk utrat informacji, w przypadku przeciwnym natomiast (gdy prawdopodobieństwa utrat są duże) skuteczniejsza i bardziej efektywna jest metoda symulacyjna.



## Układy systemów połączonych poprzez wspólną pamięć (model serwera)

### 6.1 Wstęp

W poniższym rozdziale została zaprezentowana analiza układu systemów informatycznych połączonych poprzez wspólną pamięć. Dokładniej rzecz ujmując zbadany jest pewien przypadek szczególny – układ systemów typu  $M/M/n/m$  połączonych wspólną ograniczoną pamięcią. Model badany w tym rozdziale jest modelem serwera obsługującego żądania wielu użytkowników, z których każdy dysponuje systemem informatycznym. Dobrym przykładem jest tutaj serwer pocztowy czy serwer www. Dla takiego układu systemów wyznaczone są stacjonarne prawdopodobieństwa odmowy w każdym z systemów oraz stacjonarny rozkład liczby żądań znajdujących się w pamięci serwera. Analizowany model opiera się o modele analizowane m.in w pracach [16], [44]. Wyniki uzyskane w trakcie analizy zostały częściowo zaprezentowane w pracach [45] oraz [48].

Rozważmy układ (kombinację) systemów informatycznych połączonych poprzez wspólną pamięć. W takim układzie dwa lub więcej systemów korzysta z jednej wspólnej pamięci, do sytuacji takiej dochodzi często w praktyce, kiedy kilka systemów informatycznych korzysta ze wspólnych ograniczonych zasobów pamięci. Wprowadzenie wspólnej ograniczonej pamięci prowadzi do tego, że systemy takie stają się zależne, tzn. z uwagi na to, że każdy z systemów układu korzysta z tej samej pamięci, zachowanie się każdego z systemów zależy od zachowania się pozostałych systemów. W dalszym ciągu przeprowadzimy analizę układu dwóch systemów obsługi typu  $M/M/n/m$  połączonych poprzez wspólną ograniczoną pamięć oraz dokonamy uogólnienia na większą ilość takich systemów. Dla takich kombinacji wyznaczymy też pewne istotne ze względów praktycznych charakterystyki liczbowe – stacjonarne prawdopodobieństwa odmowy w każdym z systemów oraz stacjonarny rozkład liczby żądań pochodzących z każdego systemu.

Rozpatrzmy zatem dwa działające niezależnie od siebie klasyczne systemy obsługi typu  $M/M/n/m$ , oznaczając je jako  $M/M/n_i/m_i$ , gdzie  $i = 1, 2$ ,  $1 \leq n_i \leq \infty$ ,  $0 \leq m_i \leq \infty$ . Niech  $a_i$ ,  $\mu_i$  będą parametrami wejściowego (najprostszego) strumienia żądań oraz czasu obsługi dla  $i$ -tego systemu. Dodatkowo zakładamy, że każde żądanie w  $i$ -tym systemie posiada pewien losowy rozmiar  $\zeta_i$ . Oznaczmy przez  $L_i(x)$  dystrybucję zmiennej losowej  $\zeta_i$  oraz przez  $\sigma(t)$  objętość sumaryczną wszystkich żądań obecnych w pamięci serwera pochodzących od obu systemów w chwili  $t$ . Załóżmy teraz, że czas obsługi żądań nie zależy od ich rozmiaru oraz, że oba systemy są połączone poprzez wspólną ograniczoną wielkością  $V$  pamięć. Z powodu ostatniego założenia systemy te oczywiście stają się zależne. Dla takiego układu systemów wyznaczmy wspomniane wyżej charakterystyki.

## 6.2 Analiza modelu

Niech  $\eta_i(t)$  będzie liczbą żądań w  $i$ -tym systemie w chwili czasu  $t$ ,  $i = 1, 2$ . Oznaczmy przez  $\sigma_j^i(t)$  rozmiar  $j$ -tego żądania obecnego w  $i$ -tym systemie w chwili czasu  $t$ ,  $j = \overline{1, \eta_i(t)}$ . Wówczas analizowany układ dwóch systemów może być opisany przez poniższy proces markowowski

$$\left( \eta_1(t), \eta_2(t), \sigma_j^1(t), j = \overline{1, \eta_1(t)}, \sigma_j^2(t), j = \overline{1, \eta_2(t)} \right), \quad (6.1)$$

gdzie  $\sum_{i=1}^2 \sum_{j=1}^{\eta_i(t)} \sigma_j^i(t) = \sigma(t)$ .

Proces (6.1) możemy charakteryzować poprzez poniższe funkcje:

$$G(k_1, k_2, x, t) = P\{\eta_1(t) = k_1, \eta_2(t) = k_2, \sigma(t) < x\}; \quad (6.2)$$

$$P(k_1, k_2, t) = P\{\eta_1(t) = k_1, \eta_2(t) = k_2\} = G(k_1, k_2, V, t), \quad (6.3)$$

$$k_i = \overline{0, n_i + m_i}, k_1 + k_2 \geq 1, i = 1, 2;$$

$$P_0(t) = P(0, 0, t) = P\{\eta_1(t) = 0, \eta_2(t) = 0\}. \quad (6.4)$$

Jest jasne, że warunki stacjonarne dla badanego układu systemów przy wprowadzonych założeniach są spełnione, jeśli  $V$  jest wielkością skończoną. Wtedy  $\eta_i(t) \Rightarrow \eta_i$ ,  $\sigma(t) \Rightarrow \sigma$  w sensie zbieżności według rozkładu, gdzie  $\eta_i$ ,  $\sigma$  są odpowiednio stacjonarną liczbą żądań  $i$ -tego systemu oraz stacjonarną objętością sumaryczną. Zatem istnieją poniższe granice:

$$g(k_1, k_2, x) = \lim_{t \rightarrow \infty} G(k_1, k_2, x, t) = P\{\eta_1 = k_1, \eta_2 = k_2, \sigma < x\}; \quad (6.5)$$

$$p(k_1, k_2) = \lim_{t \rightarrow \infty} P(k_1, k_2, t) = P\{\eta_1 = k_1, \eta_2 = k_2\} = g(k_1, k_2, V), \quad (6.6)$$

$$k_i = \overline{0, n_i + m_i}, k_1 + k_2 \geq 1, i = 1, 2;$$

$$p_0 = p(0, 0) = \lim_{t \rightarrow \infty} P_0(t) = P\{\eta_1 = 0, \eta_2 = 0\}. \quad (6.7)$$

Wyznamy teraz rozkład stacjonarny liczby żądań każdego z systemów.

Niech  $\delta_{i,k}$  będzie symbolem Kroneckera:  $\delta_{i,k} = \begin{cases} 0, & i \neq k, \\ 1, & i = k. \end{cases}$  Dalej będziemy używać następujących oznaczeń:

$\kappa_i = \min(k_i, n_i)$ ,  $\gamma_i = \min(k_i + 1, n_i)$ . Analizując zachowanie się procesu (6.1), można zauważyć, że funkcje (6.2)–(6.4) spełniają następujący układ równań różniczkowych:

$$P'_0(t) = -(a_1 L_1(V) + a_2 L_2(V)) P_0(t) + \mu_1 P(1, 0, t) + \mu_2 P(0, 1, t); \quad (6.8)$$

$$P'(0, 1, t) = a_2 P(0, 0, t) L_2(V) - a_1 \int_0^V G(0, 1, V - x, t) dL_1(x) -$$

$$- a_2 \int_0^V G(0, 1, V - x, t) dL_2(x) - \mu_2 P(0, 1, t) + \mu_1 P(1, 1, t) + 2\mu_2 P(0, 2, t); \quad (6.9)$$

$$P'(1, 0, t) = a_1 P(0, 0, t) L_1(V) - a_1 \int_0^V G(1, 0, V - x, t) dL_1(x) -$$

$$- a_2 \int_0^V G(1, 0, V - x, t) dL_2(x) - \mu_1 P(1, 0, t) + 2\mu_1 P(2, 0, t) + \mu_2 P(1, 1, t); \quad (6.10)$$

$$P'(k_1, k_2, t) = (1 - \delta_{0,k_1}) a_1 \int_0^V G(k_1 - 1, k_2, V - x, t) dL_1(x) +$$

$$+ (1 - \delta_{0,k_2}) a_2 \int_0^V G(k_1, k_2 - 1, V - x, t) dL_2(x) -$$

$$- (1 - \delta_{m_1+n_1, k_1}) a_1 \int_0^V G(k_1, k_2, V - x, t) dL_1(x) -$$

$$- (1 - \delta_{m_2+n_2, k_2}) a_2 \int_0^V G(k_1, k_2, V - x, t) dL_2(x) -$$

$$- (\kappa_1 \mu_1 + \kappa_2 \mu_2) P(k_1, k_2, t) + (1 - \delta_{m_1+n_1, k_1}) \gamma_1 \mu_1 P(k_1 + 1, k_2, t) +$$

$$+ (1 - \delta_{m_2+n_2, k_2}) \gamma_2 \mu_2 P(k_1, k_2 + 1, t), k_i = \overline{0, m_i + n_i}, k_1 + k_2 \geq 2, i = 1, 2. \quad (6.11)$$

W przypadku istnienia warunków stacjonarnych, z równań (6.8)–(6.11) przy przejściu granicznym  $t \rightarrow \infty$  otrzymujemy poniższy układ równań algebraicznych dla funkcji (6.5)–(6.7):

$$0 = -(a_1 L_1(V) + a_2 L_2(V)) p_0 + \mu_1 p(1, 0) + \mu_2 p(0, 1); \quad (6.12)$$

$$0 = a_2 p(0, 0) L_2(V) - a_1 \int_0^V g(0, 1, V-x) dL_1(x) - a_2 \int_0^V g(0, 1, V-x) dL_2(x) - \mu_2 p(0, 1) + \mu_1 p(1, 1) + 2\mu_2 p(0, 2); \quad (6.13)$$

$$0 = a_1 p(0, 0) L_1(V) - a_1 \int_0^V g(1, 0, V-x) dL_1(x) - a_2 \int_0^V g(1, 0, V-x) dL_2(x) - \mu_1 p(1, 0) + 2\mu_1 p(2, 0) + \mu_2 p(1, 1); \quad (6.14)$$

$$\begin{aligned} 0 = & (1 - \delta_{0, k_1}) a_1 \int_0^V g(k_1 - 1, k_2, V-x) dL_1(x) + \\ & + (1 - \delta_{0, k_2}) a_2 \int_0^V g(k_1, k_2 - 1, V-x) dL_2(x) - \\ & - (1 - \delta_{m_1 + n_1, k_1}) a_1 \int_0^V g(k_1, k_2, V-x) dL_1(x) - \\ & - (1 - \delta_{m_2 + n_2, k_2}) a_2 \int_0^V g(k_1, k_2, V-x) dL_2(x) - \\ & - (\kappa_1 \mu_1 + \kappa_2 \mu_2) p(k_1, k_2) + (1 - \delta_{m_1 + n_1, k_1}) \gamma_1 \mu_1 p(k_1 + 1, k_2) + \\ & + (1 - \delta_{m_2 + n_2, k_2}) \gamma_2 \mu_2 p(k_1, k_2 + 1), \quad k_i = \overline{0, m_i + n_i}, \quad k_1 + k_2 \geq 2, \quad i = 1, 2. \end{aligned} \quad (6.15)$$

Oznaczmy przez  $L_i^{(k)}(x)$  splot według Stieltjesa  $k$ -tego rzędu funkcji  $L_i(x)$ , natomiast dla splotu według Stieltjesa dystrybuant  $F_i(x)$ ,  $i = \overline{1, r}$  nieujemnych zmiennych losowych będziemy używać następującej notacji:

$$F_1 * \dots * F_r(x) = \underset{i=1}{*}^r F_i(x).$$

Wprowadźmy również następujące oznaczenie:

$$N_i(k) = \begin{cases} \frac{(n_i \rho_i)^k}{k!}, & k = \overline{0, n_i}, \\ \frac{n_i^{n_i} \rho_i^k}{n_i!}, & k = \overline{n_i + 1, n_i + m_i}, \end{cases} \quad (6.16)$$

gdzie  $\rho_i = a_i / (n_i \mu_i)$ .

Poprzez bezpośrednie podstawienie można sprawdzić, że rozwiązanie układu (6.12)–(6.15) ma postać:

$$g(k_1, k_2, x) = p_0 N_1(k_1) N_2(k_2) L_1^{(k_1)} * L_2^{(k_2)}(x), \quad k_1 + k_2 \geq 1, \quad (6.17)$$

skąd otrzymujemy

$$p(k_1, k_2) = g(k_1, k_2, V) = p_0 N_1(k_1) N_2(k_2) L_1^{(k_1)} * L_2^{(k_2)}(V), \quad k_i = \overline{0, m_i + n_i}. \quad (6.18)$$

Wykorzystując teraz warunek normalizacyjny  $\sum_{k_1=0}^{n_1+m_1} \sum_{k_2=0}^{n_2+m_2} p(k_1, k_2) = 1$ , otrzymujemy równość:

$$p_0 = \left[ \sum_{k_1=0}^{n_1+m_1} \sum_{k_2=0}^{n_2+m_2} N_1(k_1) N_2(k_2) L_1^{(k_1)} * L_2^{(k_2)}(V) \right]^{-1}. \quad (6.19)$$

### 6.3 Wyznaczenie prawdopodobieństw odmowy

Stacjonarne prawdopodobieństwo odmowy  $p_u^i$  żądania w  $i$ -tym systemie ( $i = 1, 2$ ) może być uzyskane z równania równowagi – równanie to ma następujący sens probabilistyczny – w warunkach stacjonarnych średnia liczba żądań przybyłych do systemu w ciągu jednostki czasu, które nie zostały odrzucone (tj. przyjętych do pamięci serwera) jest równa średniej liczbie żądań obsłużonych w ciągu jednostki czasu. W przypadku  $i = 1$  równanie równowagi przyjmuje postać:

$$a_1 (1 - p_u^1) = \mu_1 \sum_{k_1=1}^{n_1-1} k_1 p_{k_1}^1 + n_1 \mu_1 \left( 1 - \sum_{k_1=1}^{n_1-1} p_{k_1}^1 \right), \quad (6.20)$$

gdzie  $p_{k_1}^1 = \sum_{k_2=0}^{n_2+m_2} p(k_1, k_2)$ ,  $p_{k_2}^2 = \sum_{k_1=0}^{n_1+m_1} p(k_1, k_2)$ .

Z równania (6.20) oraz analogicznego równania w przypadku  $i = 2$  otrzymujemy równość:

$$p_u^i = 1 - (n_i \rho_i)^{-1} \sum_{k_i=1}^{n_i-1} k_i p_{k_i}^i - \rho_i^{-1} \left( 1 - \sum_{k_i=0}^{n_i-1} p_{k_i}^i \right). \quad (6.21)$$

### 6.4 Uogólnienie badanego modelu oraz pewne przypadki szczególne

Dokonyamy teraz uogólnienia otrzymanych wyników na dowolną liczbę systemów obsługi. Załóżmy, że mamy układ  $r$ ,  $r = 1, 2, \dots$  systemów informatycznych typu  $M/M/n/m$  połączonych poprzez wspólną ograniczoną pamięć. Niech  $a_i, \mu_i$  będą parametrami strumienia wejściowego oraz czasu obsługi dla  $i$ -tego systemu odpowiednio,  $i = \overline{1, r}$ ,  $L_i(x)$  - dystrybuantą objętości zgłoszenia w  $i$ -tym systemie,  $V$  niech oznacza objętość pamięci. Wówczas dla prawdopodobieństw stacjonarnych  $p(k_1, \dots, k_r) = P\{\eta_1 = k_1, \dots, \eta_r = k_r\}$  otrzymujemy poniższe wyrażenie:

$$p(k_1, \dots, k_r) = p_0 \prod_{i=1}^r N_i(k_i) *_{i=1}^r L_i^{(k_i)}(V),$$

gdzie funkcja  $N_i(k)$  jest określona wzorem (6.16), w którym  $i = \overline{1, r}$ . Wyrażenie (6.19) dla  $p_0$  w tym przypadku przyjmuje następującą postać:

$$p_0 = \left[ \sum_{k_1, \dots, k_r} \prod_{i=1}^r N_i(k_i) *_{i=1}^r L_i^{(k_i)}(V) \right]^{-1},$$

gdzie sumowanie dla każdego wskaźnika odbywa się od 0 do  $n_i + m_i$ .

Prawdopodobieństwo odmowy żądania dla  $i$ -tego systemu jest określone przez równość (6.21), przy czym

$$p_{k_i}^i = \sum_{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_r} p(k_1, \dots, k_{i-1}, k_i, k_{i+1}, \dots, k_r), \quad k_i = \overline{0, n_i + m_i}.$$

Zbadamy teraz pewien szczególny przypadek – mianowicie układ dwóch systemów jednoliniowych połączonych wspólną pamięcią ( $r = 2$ ). Uogólnienie na większą ilość systemów jest oczywiste. Ten przypadek odpowiada w praktyce modelowi serwera obsługującego dwa lub więcej pojedynczych komputerów wysyłających żądania do serwera (korzystających wspólnie z jego pamięci). Jeśli  $n_i = 1$ , to z równania (6.16) otrzymujemy, że  $N_i(k) = \rho_i^k$ ,  $i = 1, 2$ . Wówczas, jak to wynika z równań (6.18)–(6.21),

$$p_0 = \left[ \sum_{k_1=0}^{m_1+1} \sum_{k_2=0}^{m_2+1} \rho_1^{k_1} \rho_2^{k_2} L_1^{(k_1)} * L_2^{(k_2)}(V) \right]^{-1}, \quad (6.22)$$

$$p(k_1, k_2) = p_0 \rho_1^{k_1} \rho_2^{k_2} L_1^{(k_1)} * L_2^{(k_2)}(V), \quad k_i = \overline{0, m_i + 1}, \quad (6.23)$$

$$p_u^i = 1 - \frac{1}{\rho_i} (1 - p_0^i), \quad i = 1, 2. \quad (6.24)$$

Otrzymane wyrażenia nie są wygodne do obliczeń z uwagi na trudności wyznaczenia splotów według Stieltjesa. Sploty te mogą być przybliżane z wykorzystaniem np. metody Monte Carlo. Natomiast w pewnych szczególnych przypadkach można otrzymać precyzyjne wyniki. Na przykład, niech rozmiary żądań w  $i$ -tym systemie spełniają rozkład gamma o poniższej gęstości

$$l_i(x) = \gamma_f(\alpha_i, x) = \frac{f}{\Gamma(\alpha_i)} x^{\alpha_i-1} e^{-fx}, \quad \alpha_i > 0, \quad f > 0, \quad i = 1, 2,$$

(parametr  $f$  jest taki sam dla obu systemów). Jest jasne, że w tym przypadku splot  $L_1^{(k_1)} * L_2^{(k_2)}(x)$  przedstawia dystrybucję rozkładu gamma  $Y_f(\alpha, x)$  z parametrami  $f$  oraz  $\alpha = k_1 \alpha_1 + k_2 \alpha_2$ , wówczas mamy

$$p(k_1, k_2) = p_0 \rho_1^{k_1} \rho_2^{k_2} Y_f(k_1 \alpha_1 + k_2 \alpha_2, V),$$

$$p_0 = \left[ \sum_{k_1=0}^{m_1+1} \sum_{k_2=0}^{m_2+1} \rho_1^{k_1} \rho_2^{k_2} Y_f(k_1 \alpha_1 + k_2 \alpha_2, V) \right]^{-1}.$$

Wzór na prawdopodobieństwo odmowy w pierwszym systemie przyjmuje wówczas postać:

$$p_u^1 = 1 - \frac{1}{\rho_1} \left[ 1 - \sum_{k_2=0}^{m_2+1} \rho_2^{k_2} Y_f(k_2 \alpha_2, V) \right].$$

Założmy teraz, że  $L_1(x) = L_2(x) = L(x)$  (rozkład objętości zgłoszeń w obu systemach jest taki sam). Wówczas ze wzorów (6.18) oraz (6.19) otrzymujemy

$$p(k_1, k_2) = p_0 N_1(k_1) N_2(k_2) L^{(k_1+k_2)}(V), \quad k_i = \overline{0, m_i + n_i}, \quad i = 1, 2;$$

$$p_0 = \left[ \sum_{k_1=0}^{n_1+m_1} \sum_{k_2=0}^{n_2+m_2} N_1(k_1) N_2(k_2) L^{(k_1+k_2)}(V) \right]^{-1}.$$

Założmy teraz dodatkowo, że  $n_1 = n_2 = 1$ ,  $m_1 = m_2 = \infty$  oraz objętości zgłoszeń mają rozkład wykładniczy z parametrem  $f$ :  $L(x) = 1 - e^{-fx}$ . Dodatkowo przypuścmy, że  $\rho_1 \neq 1$ ,  $\rho_2 \neq 1$  oraz  $\rho_1 \neq \rho_2$ . Wówczas ze wzorów (6.18), (6.19) otrzymujemy

$$p(k_1, k_2) = p_0 \rho_1^{k_1} \rho_2^{k_2} \left[ 1 - e^{-fV} \sum_{j=0}^{k_1+k_2-1} \frac{(fV)^j}{j!} \right], \quad k_i = 0, 1, \dots, \quad i = 1, 2;$$

$$p_0 = \left\{ \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \rho_1^{k_1} \rho_2^{k_2} \left[ 1 - e^{-fV} \sum_{j=0}^{k_1+k_2-1} \frac{(fV)^j}{j!} \right] \right\}^{-1}. \quad (6.25)$$

Obliczmy sumę

$$\begin{aligned} & \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \rho_1^{k_1} \rho_2^{k_2} \left[ 1 - e^{-fV} \sum_{j=0}^{k_1+k_2-1} \frac{(fV)^j}{j!} \right] = \\ & = e^{-fV} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \rho_1^{k_1} \rho_2^{k_2} \left[ e^{fV} - \sum_{j=0}^{k_1+k_2-1} \frac{(fV)^j}{j!} \right] = \\ & = e^{-fV} \sum_{j=0}^{\infty} \frac{(fV)^j}{j!} \sum_{k_1=0}^j \rho_1^{k_1} \sum_{k_2=0}^{j-k_1} \rho_2^{k_2} = \\ & = \frac{e^{-fV}}{1-\rho_2} \sum_{j=0}^{\infty} \frac{(fV)^j}{j!} \sum_{k_1=0}^j \rho_1^{k_1} (1 - \rho_2^{j-k_1+1}), \end{aligned} \quad (6.26)$$

gdzie

$$\sum_{k_1=0}^j \rho_1^{k_1} \rho_2^{j-k_1+1} = \rho_2^{j+1} \sum_{k_1=0}^j \left(\frac{\rho_1}{\rho_2}\right)^{k_1} = \frac{\rho_2}{\rho_2 - \rho_1} (\rho_2^{j+1} - \rho_1^{j+1}),$$

$$\sum_{k_1=0}^j \rho_1^{k_1} = \frac{1 - \rho_1^{j+1}}{1 - \rho_1},$$

ostatecznie

$$\sum_{j=0}^{\infty} \frac{(fV)^j}{j!} \sum_{k_1=0}^j \rho_1^{k_1} = \sum_{j=0}^{\infty} \frac{(fV)^j}{j!} \cdot \frac{1 - \rho_1^{j+1}}{1 - \rho_1} = \frac{e^{fV} - \rho_1 e^{\rho_1 fV}}{1 - \rho_1},$$

$$\sum_{j=0}^{\infty} \frac{(fV)^j}{j!} \sum_{k_1=0}^j \rho_1^{k_1} \rho_2^{j-k_1+1} = \frac{\rho_2}{\rho_2 - \rho_1} \sum_{j=0}^{\infty} \frac{(fV)^j}{j!} (\rho_2^{j+1} - \rho_1^{j+1}) =$$

$$= \frac{\rho_2 (\rho_2 e^{\rho_2 fV} - \rho_1 e^{\rho_1 fV})}{\rho_2 - \rho_1}.$$

Wstawiając wartość obliczonej sumy do wzoru (6.26) po prostych przekształceniach otrzymujemy

$$p_0 = \left[ \frac{1}{(1 - \rho_1)(1 - \rho_2)} - \frac{\rho_1^2 e^{-(1-\rho_1)fV}}{(1 - \rho_1)(\rho_1 - \rho_2)} - \frac{\rho_2^2 e^{-(1-\rho_2)fV}}{(1 - \rho_2)(\rho_2 - \rho_1)} \right]^{-1}. \quad (6.27)$$

Prawdopodobieństwo odmowy wyznaczamy ze wzoru (6.24), gdzie dla  $i = 1$  otrzymujemy

$$p_0^1 = \sum_{j=0}^{\infty} p(0, j) = \sum_{j=0}^{\infty} p_0 \rho_2^j \left[ 1 - e^{-fV} \sum_{i=0}^{j-1} \frac{(fV)^i}{i!} \right] =$$

$$= p_0 e^{-fV} \sum_{j=0}^{\infty} \rho_2^j \left[ e^{fV} - \sum_{i=0}^{j-1} \frac{(fV)^i}{i!} \right] = p_0 e^{-fV} \sum_{j=0}^{\infty} \rho_2^j \sum_{i=j}^{\infty} \frac{(fV)^i}{i!} =$$

$$= p_0 e^{-fV} \sum_{i=0}^{\infty} \frac{(fV)^i}{i!} \sum_{j=0}^i \rho_2^j = p_0 e^{-fV} \sum_{i=0}^{\infty} \frac{(fV)^i}{i!} \cdot \frac{1 - \rho_2^{j+1}}{1 - \rho_2} =$$

$$= \frac{p_0 e^{-fV}}{1 - \rho_2} (e^{fV} - \rho_2 e^{\rho_2 fV}),$$

Ostatecznie na mocy (6.24) mamy

$$p_u^1 = 1 - \frac{1}{\rho_1} \left[ 1 - \frac{p_0 (1 - \rho_2 e^{-(1-\rho_2)fV})}{1 - \rho_2} \right]. \quad (6.28)$$



W analogiczny sposób otrzymujemy

$$p_u^2 = 1 - \frac{1}{\rho_2} \left[ 1 - \frac{p_0 (1 - \rho_1 e^{-(1-\rho_1)fV})}{1 - \rho_1} \right].$$

Można łatwo pokazać, że  $p_u^1 = p_u^2$ .

Jeśli założymy, że  $\rho_1 = \rho_2 = \rho \neq 1$ , to otrzymamy

$$p_u^1 = p_u^2 = \frac{(\rho - 1)^2 + p_0(-1 + e^{f(\rho-1)V})}{\rho(\rho - 1)},$$

gdzie  $p_0 = \left[ \frac{1 + e^{(\rho-1)fV} \rho(-2 + \rho + f(\rho-1)\rho V)}{(\rho-1)^2} \right]^{-1}$ .

Numeryczne obliczenia w tym przypadku są przedstawione w tabeli 6.1.

**Tabela 6.1.** Prawdopodobieństwa odmowy w przypadku ( $\rho_1 = \rho_2 = 2$ )

	$f = 1$			$f = 2$		
$V$	$p_u^{an}$	$p_u^{1sim}$	$p_u^{2sim}$	$p_u^{an}$	$p_u^{1sim}$	$p_u^{2sim}$
1	0.686832	0.686795	0.686833	0.614603	0.614647	0.614602
2	0.614603	0.614588	0.614515	0.561857	0.561810	0.562041
3	0.580923	0.580899	0.580654	0.541611	0.541680	0.541745
4	0.561857	0.561779	0.561724	0.531244	0.531494	0.531350
5	0.549815	0.549978	0.549948	0.524999	0.525055	0.524902
6	0.541611	0.541553	0.541606	0.520833	0.521165	0.521116
7	0.535697	0.535677	0.535696	0.517857	0.517851	0.517885
8	0.531244	0.531021	0.531105	0.515625	0.515622	0.515614
9	0.527776	0.527945	0.527817	0.513889	0.513766	0.513638
10	0.524999	0.524907	0.525007	0.512500	0.512583	0.512445

Jeśli  $\rho_1 = 1$  oraz  $\rho_2 \neq 1$ , to wówczas

$$p_u^1 = p_u^2 = 1 - \frac{1 - (1 + fV)p_0}{\rho_2},$$

gdzie  $p_0 = \left[ \frac{1 + fV + \rho_2(-2 + e^{f(\rho_2-1)V} \rho_2 - fV)}{(\rho_2-1)^2} \right]^{-1}$ .

Jeśli  $\rho_1 \neq 1$  oraz  $\rho_2 = 1$ , to

$$p_u^1 = p_u^2 = 1 - \frac{1 - (1 + fV)p_0}{\rho_1},$$

gdzie  $p_0 = \left[ \frac{1 + fV + \rho_1(-2 + e^{f(\rho_1-1)V} \rho_1 - fV)}{(\rho_1-1)^2} \right]^{-1}$ .

Jeśli wreszcie  $\rho_1 = 1$  oraz  $\rho_2 = 1$ , to w tym wypadku otrzymujemy

$$p_u^1 = p_u^2 = p_0(1 + fV),$$

gdzie  $p_0 = \frac{2}{2+fV(4+fV)}$ .

Wyniki dla tego przypadku są przedstawione w tabeli 6.2.

**Tabela 6.2.** Prawdopodobieństwa odmowy w przypadku ( $\rho_1 = \rho_2 = 1$ )

	$f = 1$			$f = 2$		
$V$	$p_u^{an}$	$p_u^{1sim}$	$p_u^{2sim}$	$p_u^{an}$	$p_u^{1sim}$	$p_u^{2sim}$
1	0.571429	0.571275	0.571358	0.428571	0.428492	0.428357
2	0.428571	0.428538	0.428451	0.294118	0.294076	0.294357
3	0.347826	0.347784	0.348007	0.225806	0.225838	0.225987
4	0.294118	0.294001	0.294107	0.183673	0.183590	0.183621
5	0.255319	0.254936	0.255082	0.154930	0.155001	0.155152
6	0.225806	0.225657	0.225452	0.134021	0.134111	0.134160
7	0.202532	0.202312	0.202552	0.118110	0.117945	0.117995
8	0.183673	0.183616	0.183433	0.105590	0.105270	0.105623
9	0.168067	0.168141	0.168265	0.095477	0.095482	0.095491
10	0.154930	0.154811	0.154734	0.087137	0.086818	0.086976

Wzory (6.27),(6.28) mogą być uogólnione na przypadek  $r$  systemów (gdzie  $\rho_i \neq \rho_j$  dla  $i \neq j$  oraz  $\rho_i \neq 1$  dla  $i = \overline{1, r}$ ). W tym przypadku wzór (6.27) przyjmuje postać

$$p_0 = \left[ \prod_{i=1}^r \frac{1}{1 - \rho_i} - \sum_{i=1}^r \frac{\rho_i^r e^{-(1-\rho_i)fV}}{(1 - \rho_i) \prod_{j \neq i} (\rho_i - \rho_j)} \right]^{-1}.$$

Prawdopodobieństwa odmowy  $p_u^i$  mają te same wartości i są wyrażone poprzez wzór (6.24), gdzie  $i = \overline{1, r}$  oraz

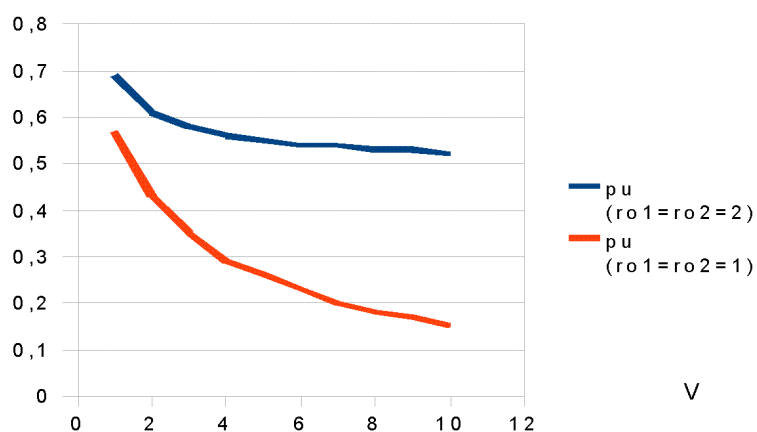
$$p_0^i = p_0 \left[ \prod_{j \neq i} \frac{1}{1 - \rho_j} - \sum_{j \neq i} \frac{\rho_j^{r-1} e^{-(1-\rho_j)fV}}{(1 - \rho_j) \prod_{k \neq j, k \neq i} (\rho_j - \rho_k)} \right].$$

Podczas analizy możemy zauważyć, że badana kombinacja systemów jednoliniowych z wykładniczym rozkładem rozmiaru żądań i parametrem  $f$  jednakowym dla wszystkich systemów może należeć do jednej z klas:

- **stan stabilny** ( $\rho_i \leq 1$  dla wszystkich  $i = \overline{1, r}$ ). Wtedy dla każdego systemu mamy  $p_u \rightarrow 0$ , gdy  $V \rightarrow \infty$ ;

- **stan przeładowania** ( $\rho_i > 1$  przynajmniej dla jednego  $i$ ,  $i = \overline{1, r}$ ). Wtedy dla każdego systemu mamy  $p_u \rightarrow 1 - \min\{\frac{1}{\rho_i}, i = \overline{1, r}\}$ , gdy  $V \rightarrow \infty$ .

Fakt ten pokazują obliczenia zawarte w tabelach 6.1 oraz 6.2 oraz rys. 6.1.



**Rysunek 6.1.** Porównanie prawdopodobieństw utrat w przypadku  $\rho_1 = \rho_2 = 2$  oraz  $\rho_1 = \rho_2 = 1$

Jeśli parametry wykładniczego rozmiaru żądania nie są jednakowe w każdym systemie, to wówczas prawdopodobieństwa utrat w każdym systemie są różne.

Wyniki symulacji zawarte w tabeli 6.3 potwierdzają ten fakt.

**Tabela 6.3.** Prawdopodobieństwa odmowy w przypadku  $\rho_1 = 1, \rho_2 = 1$  (wykładniczy rozkład rozmiarów procesów,  $f_1 = 1, f_2 = 2$ )

V	$p_u^{1sim}$	$p_u^{2sim}$
1	0.599983	0.399990
2	0.454333	0.272754
3	0.368260	0.210088
4	0.310260	0.172009
5	0.267856	0.146252
6	0.236325	0.127624
7	0.210847	0.112379
8	0.190707	0.100984

Na koniec zbadamy jeszcze jeden szczególny przypadek. Załóżmy, że dana jest kombinacja dwóch systemów jednoliniowych ze skończoną liczbą miejsc oczekiwania. Dodatkowo założmy, że rozmiary żądań mają jednakowy rozkład jednostajny w przedziale  $[a, b]$  ( $0 \leq a < b$ ).

Wykorzystując odwrócone przekształcenie Laplace'a możemy znaleźć wzór określający splot według Stieltjesa  $k$ -tego rzędu  $k$  niezależnych zmiennych losowych, z których każda ma rozkład jednostajny na odcinku  $[a, b]$

$$L^{(k)}(x) = \left( \frac{-1}{b-a} \right)^k \sum_{l=0}^k \frac{(-1)^l ((b-a)l - bk + x)^k H((b-a)l - bk + x)}{l!(k-l)!}, \quad (6.29)$$

gdzie  $H(x)$  jest funkcją Heaviside'a skoku jednostkowego.

Wówczas wykorzystując wzór (6.24) oraz obliczając odpowiednie sumy skończone, możemy otrzymać numeryczne wartości prawdopodobieństw  $p_u^1$  oraz  $p_u^2$ . Wyniki dla tego przypadku są przedstawione w tabelach 6.4 oraz 6.5.

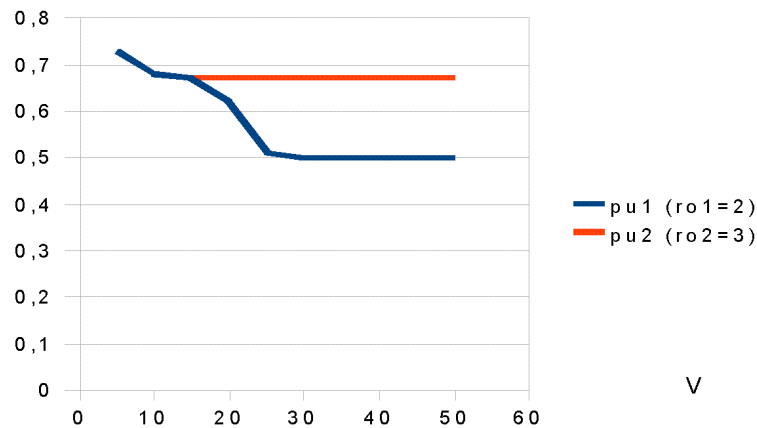
**Tabela 6.4.** Prawdopodobieństwa odmowy w przypadku  $\rho_1 = 2, \rho_2 = 3$  (jednostajny rozkład rozmiaru żądań,  $a = 1, b = 3$ )

$V$	$p_u^{1sim}$	$p_u^{1an}$	$p_u^{2sim}$	$p_u^{2an}$
5	0.743848	0.725126	0.743849	0.725126
10	0.687949	0.681329	0.687947	0.681329
15	0.673332	0.670894	0.673261	0.670945
20	0.652558	0.616383	0.668862	0.667861
25	0.548879	0.510477	0.667036	0.666757
30	0.508135	0.500290	0.666488	0.666669
35	0.501514	0.500009	0.666820	0.666667
40	0.500208	0.500001	0.666781	0.666667

**Tabela 6.5.** Prawdopodobieństwa odmowy w przypadku  $\rho_1 = \frac{1}{2}, \rho_2 = \frac{1}{3}$  (jednostajny rozkład rozmiaru żądań,  $a = 1, b = 3$ )

$V$	$p_u^{1sim}$	$p_u^{2sim}$
5	0.224913	0.225005
10	0.044114	0.044294
15	0.008617	0.008629
20	0.001590	0.001643
25	0.000248	0.000251
30	0.000248	0.000022
35	0.000243	0.000006
40	0.000239	0.000004

Możemy zauważyć, że prawdopodobieństwa odmowy  $p_u^i$  nie są jednakowe i są zbieżne do 0 w stanie stabilnym oraz do  $1 - \frac{1}{\rho_i}$ , ( $i = 1, 2$ ) w stanie przeładowania, co ma związek ze skończoną liczbą miejsc oczekiwania (rys. 6.2).



**Rysunek 6.2.** Porównanie prawdopodobieństw utrat w przypadku  $\rho_1 = 2$  oraz  $\rho_2 = 3$

W rozdziale **Dodatki** - punkt 7.3 oraz 7.4 znajdują się omówienia programów komputerowych, w oparciu o które (oraz ich modyfikacje) uzyskano symulacyjne wyniki dla wszystkich danych zawartych w powyższych tabelach.

## 6.5 Realizacje techniczne omawianego modelu

W punkcie tym zaprezentujemy kilka przykładów technicznych realizacji analizowanego modelu. Jest jasne, że model serwera obsługującego żądania wielu użytkowników jest pewnym uogólnieniem systemu obsługi zgłoszeń niejednorodnych  $M/M/n/(m, V)$ , w którym czasy obsługi zgłoszeń i ich objętości są niezależne. Początkowo modele takiego typu były używane do wyznaczania pamięci buforowej systemów telekomunikacyjnych. Problemy takiego typu były poruszane już w pracach [27] oraz [28] w oparciu o klasyczne modele teorii obsługi masowej. Następnie zauważono, że klasyczne modele teorii obsługi masowej nie wystarczą i należy wprowadzić istotne uogólnienia. Wyniki związane z analizą systemu  $M/M/n/(m, V)$  pojawiły się przede wszystkim w pracy [16], a ich techniczne realizacje na przykład w pracy [44].

Jeżeli chodzi o techniczne realizacje modelu analizowanego w tym rozdziale, to możemy ich poszukiwać przede wszystkim w sieciach komputerowych. Wszędzie tam, gdzie mamy do czynienia z serwerem przechowującym dane pochodzące od wielu użytkowników sieci komputerowej analizowany model ma swoje praktyczne uzasadnienie. Przejdziemy teraz do przedstawienia kilku interesujących przypadków zastosowania analizowanego modelu.

### 1. Serwer FTP

Rozważamy serwer FTP, na którym użytkownicy umieszczają różnorodne pliki. Dodatkowo zakładamy, że każdy klient serwera FTP może zamieszczać pliki o maksymalnym możliwym w danej chwili rozmiarze, to znaczy ograniczenie jest związane tylko z maksymalną dostępną w danym momencie pamięcią serwera. Całkowita dostępna pamięć serwera ma rozmiar  $V$ , kolejka zamieszczanych plików jest nieograniczona dla każdego klienta. Klienci wysyłają losowe żądania zamieszczenia plików na serwerze - czas między sąsiednimi chwilami inicjalizacji żądań ma rozkład wykładniczy z parametrem  $a_i$  dla  $i$ -tego klienta. Czas obsługi żądania może być tutaj rozumiany jako czas przechowywania określonego pliku na serwerze. Zakładamy, że czas ten ma rozkład wykładniczy z parametrem  $\mu_i$  dla  $i$ -tego klienta i jest niezależny od rozmiaru żądania, co ma swoje praktyczne uzasadnienie, gdyż czas przechowywania plików na serwerze FTP zwykle nie zależy od ich rozmiaru. W oparciu o analizę modelu serwera przedstawioną w punktach 6.2–6.4 możemy projektować architekturę serwera FTP w taki sposób, aby prawdopodobieństwa utrat żądań nie były przekroczone (odpowiedni dobór pamięci serwera). Jest to problem dość ważny, firmom udostępniającym konta FTP zależy bowiem na stabilnym zarządzaniu takimi kontami i zadowoleniu klientów.

### 2. Serwer mailowy

Tutaj sytuacja jest bardzo podobna. Serwer poczty mailowej jest wyspecjalizowanym komputerem lub siecią komputerów przechowującą wiadomości mailowe pochodzące od wielu użytkowników. Czas obsługi wiadomości mailowej może być rozumiany jako czas przechowywania wiadomości na serwerze do momentu jej skasowania. Tutaj również możemy założyć, że czas obsługi wiadomości nie zależy od jej rozmiaru. W tej realizacji technicznej właścicielom serwerów pocztowych zależy przede wszystkim na niezawodności komunikacji mailowej, stąd wyznaczanie rozmiaru potrzebnej pamięci jest bardzo ważnym zadaniem przy projektowaniu takiego serwera.

### 3. Portale aukcyjne, sklepy internetowe

W dzisiejszych czasach rola handlu elektronicznego znacznie wzrosła. Udział procentowy handlu w internecie oraz generowane zyski stale się powiększają. Sukces portali aukcyjnych oraz sklepów internetowych

zależy między innymi od odpowiednio zaprojektowanych serwerów, na których zamieszczane są dane związane z przeprowadzonymi transakcjami. Użytkownicy, oprócz ogłoszeń (będących plikami w ogólnym przypadku tekstowymi), zamieszczają zdjęcia, filmy i inne pliki o dużych rozmiarach. Jako czas obsługi żądania generowanego w portalu aukcyjnym lub sklepie internetowym rozumiemy czas transakcji do momentu usunięcia związanych z nią danych z pamięci serwera, na którym dany portal jest zamieszczony. Założenie o niezależności czasu transakcji od rozmiaru danych związanych z transakcją ma również swoje uzasadnienie. Serwery obsługujące takie portale są również dobrym przykładem modelu analizowanego w tym rozdziale, a głównym problemem jest również odpowiedni dobór rozmiaru pamięci, tak aby portale działały stabilnie, w przeciwnym wypadku klienci mogą wycofywać się z transakcji i w najgorszych przypadkach tracić zaufanie do danego portalu.

#### 4. Portale społecznościowe

Portale społecznościowe są oparte na serwerach o dużych pamięciach umożliwiających zamieszczanie przez użytkowników różnorodnego rodzaju plików (zdjęcia, filmy). Czas obsługi żądania generowanego przez użytkownika jest rozumiany jako czas przechowywania danego pliku na portalu do momentu jego usunięcia, tutaj też czas ten zwykle nie zależy od rozmiaru zamieszczanego pliku. Z uwagi na działalność reklamową portalom społecznościowym zależy na utrzymaniu klientów, co ma związek z zapewnieniem stabilności działania takich portali, związanych niewątpliwie z zapewnieniem odpowiedniego rozmiaru pamięci dla serwerów obsługujących dany portal.

#### 5. Obsługa urządzeń końcowych

Założmy, że do systemu komputerowego posiadającego pamięć o rozmiarze  $V$  podłączonych jest kilka urządzeń końcowych, które generują żądania obsługi, każde takie żądanie zajmuje pamięć systemu i ma losowy rozmiar. Czas obsługi żądania to czas potrzebny na wykonanie operacji na danym urządzeniu, po wykonaniu takiej operacji pamięć jest zwalniana. Zagadnieniem związanym z taką techniczną realizacją jest zapewnienie stabilności działania urządzeń końcowych poprzez odpowiedni dobór pamięci wydzielonej dla obsługi żądań w oparciu o znane charakterystyki wejściowych strumieni żądań, ich rozmiarów oraz czasów obsługi. Analizowany odpowiednik techniczny jest pewną modyfikacją modelu przedstawionego w pracy [7].





## Dodatki

W rozdziale tym omówione zostało działanie programów symulacyjnych dla modeli pamięci operacyjnej komputera rozważanych w rozdziale 5 oraz modeli serwera obsługującego żądania wielu użytkowników rozważanych w rozdziale 6 tej pracy. Programy są napisane w języku C w oparciu o algorytm symulacji działania systemu obsługi zgłoszeń niejednorodnych z wykorzystaniem metody następnego zdarzenia ([1]). Główna idea tej metody polega na przechodzeniu między kolejnymi zmianami stanów analizowanego systemu z równoczesnym zbieraniem pośrednich charakterystyk utrat, które służą do finalnych obliczeń związanych z wyznaczaniem prawdopodobieństw utrat dla analizowanych modeli. Zmianami stanów są tutaj zmiany związane z liczbą zgłoszeń (procesów) obecnych w systemie czyli najczęściej momenty przybywania zgłoszeń oraz momenty zakończenia obsługi zgłoszeń. W każdym programie wykorzystujemy standardowy generator liczb pseudolosowych z biblioteki `<stdlib.h>`, który generuje liczby z przedziału  $(0, 1)$  w oparciu o rozkład jednostajny, a następnie w oparciu o metodę odwracania dystrybuanty uzyskujemy realizacje interesujących nas rozkładów zmiennych losowych wykorzystywanych w symulacji np. odstępów czasu między sąsiednimi chwilami przybywania zgłoszeń czy czasów obsługi poszczególnych zgłoszeń, a także rozmiarów zgłoszeń (procesów). Przejdziemy teraz do omówienia poszczególnych programów.

### 7.1 Omówienie działania programu symulacyjnego dla modelu RAM - schemat 1

Celem programu symulacyjnego dla modelu pamięci operacyjnej komputera jest wyznaczenie charakterystyk utrat takich jak np. prawdopodobieństwa utrat procesów wewnętrznych i zewnętrznych oraz prawdopodobieństwa utrat wszystkich procesów oraz wyznaczenie prawdopodobieństw utrat jednostki informacji dla danej klasy procesów. Dane wejściowe do programu to

przede wszystkim : parametr  $a$  wejściowego najprostszego strumienia procesów zewnętrznych, parametr  $c$  wejściowego najprostszego strumienia procesów wewnętrznych, parametr  $\mu$  wykładniczego czasu obsługi procesów, parametr  $f$  rozmiaru procesów (w przypadku, gdy czas obsługi jest proporcjonalny do rozmiaru procesu wystarczy podać tylko parametr  $f$ , wtedy  $\mu = lf$ , gdzie  $l$  jest współczynnikiem proporcjonalności), rozmiar pamięci  $V$  oraz czas symulacji  $TM$ . Wprowadzamy również zmienne przechowujące czasy lokalne zmian liczby procesów obecnych w systemie. Zmienna  $DMT1$  przechowuje ostatni czas przybycia procesu zewnętrznego, zmienna  $DMT2$  - ostatni czas zakończenia obsługi procesu zewnętrznego,  $DMT4$  - ostatni czas przybycia procesu wewnętrznego, a zmienna  $DMT5$  - ostatni czas zakończenia obsługi procesu wewnętrznego. Program działa na zasadzie wyboru minimum spośród wartości analizowanych zmiennych  $DMT1$ – $DMT4$  oraz  $TM$  i wykonywaniu odpowiednich cząstkowych obliczeń w trakcie przechodzenia między zmianami liczby procesów obecnych w systemie. Dokładniej program działa w następujący sposób:

1. Przydzielamy wartości początkowe zmiennym  $DMT1$ – $DMT4$ :  
 $DMT1$  przyjmuje wylosowaną w oparciu o rozkład wykładniczy z parametrem  $a$  wartość, zmienne  $DMT2$ – $DMT4$  przyjmują wartość większą od czasu modelowania  $TM$ .
2. Realizujemy pętlę wyborów minimum spośród wartości zmiennych  $DMT1$ ,  $DMT2$ ,  $DMT4$ ,  $DMT5$  oraz  $TM$ . W pierwszym przebiegu pętli wybrane będzie oczywiście  $DMT1$ , w następnych może być wybrana każda ze zmiennych.
  - Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT1$ , to wykonujemy następujące czynności :
    - generujemy losowy rozmiar przybywającego procesu zewnętrznego  $x$ ;
    - zwiększamy sumaryczne liczby oraz objętości przybywających do systemu procesów oraz procesów zewnętrznych odpowiednio o 1 oraz o  $x$ ;
    - sprawdzamy czy przybywający proces będzie przyjęty do pamięci; jeśli tak to zwiększamy bieżącą liczbę procesów zewnętrznych o 1, zapisujemy na końcu dynamicznej tablicy rozmiar tego procesu, zwiększamy objętość sumaryczną procesów obecnych w systemie o wartość  $x$ , dodatkowo jeśli kolejka procesów jest pusta, to generujemy czas obsługi dla tego procesu i ustalamy nową wartość dla zmiennej  $DMT2$ , a także generujemy czas przybycia procesu wewnętrznego i

ustalamy nową wartość dla zmiennej  $DMT4$ , w przeciwnym wypadku zwiększamy liczbę procesów obecnych w kolejce o 1; jeżeli przybywający proces nie zostanie przyjęty do systemu, to zwiększamy liczby i objętości odpowiednio utraconych procesów oraz ich rozmiarów;

- generujemy kolejną chwilę przybycia procesu zewnętrznego i zmieniamy wartość zmiennej  $DMT1$ .

- Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT2$ , to wykonujemy poniższe operacje:

- zmieniamy wartość zmiennej  $DMT4$  na wartość większą od czasu modelowania  $TM$ ;

- zmniejszamy wartość sumarycznej objętości procesów o wartość pierwszego elementu dynamicznej tablicy rozmiarów procesów;

- zmniejszamy liczbę procesów zewnętrznych o 1;

- jeżeli kolejka procesów jest pusta to dodatkowo ustalamy wartość zmiennej  $DMT2$  na wartość większą od czasu modelowania  $TM$ ; w przeciwnym wypadku zmniejszamy liczbę procesów obecnych w kolejce o 1, generujemy czas obsługi dla kolejnego procesu, a także kolejny czas przybycia procesu wewnętrznego i ustalamy nowe wartości dla zmiennych  $DMT2$  oraz  $DMT4$ .

- Jeżeli kolejnym minimum jest wartość zmiennej  $DMT4$ , to akcje będą następujące:

- generujemy rozmiar procesu wewnętrznego  $y$ ;

- zwiększamy sumaryczne liczby oraz objętości przybywających do systemu procesów oraz procesów wewnętrznych odpowiednio o 1 oraz o  $y$ ;

- sprawdzamy czy przybywający proces będzie przyjęty do pamięci; jeśli tak, to generujemy czas obsługi dla tego procesu oraz ustalamy nową wartość dla zmiennych  $DMT5$  oraz  $DMT2$ , zmieniamy wartość zmiennej  $DMT4$  na wartość większą od czasu modelowania  $TM$  oraz zwiększamy objętość sumaryczną procesów obecnych w systemie o wartość  $y$ ; w przeciwnym wypadku zwiększamy liczby i objętości odpowiednio utraconych procesów oraz ich rozmiarów, generujemy kolejną chwilę przybycia procesu wewnętrznego oraz zmieniamy

wartość zmiennej  $DMT4$ .

- Jeżeli minimum to wartość zmiennej  $DMT5$ , to wykonujemy poniższe czynności:
    - zmniejszamy sumaryczną objętość procesów o  $y$ ;
    - generujemy kolejną chwilę przybycia procesu wewnętrznego;
    - zmieniamy wartość zmiennej  $DMT4$  oraz ustalamy wartość zmiennej  $DMT5$  na wartość większą od czasu modelowania.
  - Jeżeli minimum jest wartością zmiennej  $TM$ , to kończymy symulację.
3. Obliczamy charakterystyki utrat na podstawie wartości zmiennych oznaczających odpowiednio liczby oraz objętości przybywających oraz utraconych procesów.

## 7.2 Omówienie działania programu symulacyjnego dla modelu RAM - schemat 2

Program symulacyjny dla schematu 2 przebiega w ten sam sposób jak program dla schematu 1 z następującymi różnicami:

1. Wprowadzamy zmienną  $SUMA$  przechowującą objętość sumaryczną wszystkich procesów wewnętrznych, które przybywają do systemu podczas obsługi procesu zewnętrznego.
2. Jeżeli kolejnym minimum jest wartość zmiennej  $DMT4$ , to zwiększamy wartość zmiennej  $SUMA$  o wartość zmiennej  $x$ .
3. Jeżeli kolejnym minimum jest wartość zmiennej  $DMT5$ , to nie zmniejszamy sumarycznej objętości procesów, gdyż wszystkie procesy wewnętrzne, które przybyły do systemu w trakcie obsługi procesu zewnętrznego zachowują się w pamięci do momentu zakończenia obsługi procesu zewnętrznego.
4. Jeżeli kolejnym minimum jest  $DMT2$ , to sumaryczną objętość procesów zmniejszamy o wartość zmiennej  $SUMA$  i ustalamy wartość tej zmiennej na 0.

### 7.3 Omówienie działania programu dla modelu serwera - nieograniczona kolejka

Celem programu symulacyjnego dla modelu serwera jest wyznaczenie prawdopodobieństw odmowy realizacji żądań dla pierwszego i drugiego systemu. Dane wejściowe do programu to przede wszystkim : parametry  $a_1$ ,  $a_2$  wejściowych najprostszycy strumieni żądań, parametry  $\mu_1$ ,  $\mu_2$  wykładniczych czasów obsługi żądań, parametr  $f$  rozmiaru żądań, rozmiar pamięci  $V$  oraz czas symulacji  $TM$ . Wprowadzamy również zmienne przechowujące czasy lokalne zmian liczby żądań pochodzących od każdego systemu. Zmienna  $DMT1$  przechowuje ostatni czas generacji żądania pochodzącego z pierwszego systemu, zmienna  $DMT2$  - ostatni czas zakończenia obsługi żądania pochodzącego z pierwszego systemu,  $DMT4$  - ostatni czas generacji żądania pochodzącego z drugiego systemu, a zmienna  $DMT5$  - ostatni czas zakończenia obsługi żądania pochodzącego z drugiego systemu. Program działa na zasadzie wyboru minimum spośród wartości analizowanych zmiennych  $DMT1$ – $DMT4$  oraz  $TM$  i wykonywaniu odpowiednich cząstkowych obliczeń w trakcie przechodzenia między zmianami liczby żądań pochodzących z każdego systemu. Dokładniej program działa w następujący sposób:

1. Przydzielamy wartości początkowe zmiennym  $DMT1$ – $DMT4$ :  
Zmienne  $DMT1$ ,  $DMT4$  przyjmują wartości wylosowane w oparciu o rozkład wykładniczy z parametrami odpowiednio  $a_1$ ,  $a_2$ , zmienne  $DMT2$ – $DMT3$  przyjmują wartości większe od czasu modelowania  $TM$ .
2. Realizujemy pętlę wyborów minimum spośród wartości zmiennych  $DMT1$ ,  $DMT2$ ,  $DMT4$ ,  $DMT5$  oraz  $TM$ . W pierwszym przebiegu pętli wybrane będzie oczywiście  $DMT1$  lub  $DMT4$ , w następnych może być wybrana każda ze zmiennych.
  - Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT1$ , to wykonujemy następujące czynności :
    - generujemy losowy rozmiar żądania pochodzącego z pierwszego systemu  $x1$ ;
    - zwiększamy sumaryczne liczby oraz objętości żądań pochodzących z pierwszego systemu odpowiednio o 1 oraz o  $x1$ ;
    - sprawdzamy czy generowane żądanie będzie przyjęte do systemu; jeśli tak to zwiększamy bieżącą liczbę żądań pochodzących z pierwszego systemu o 1, zapisujemy na końcu dynamicznej tablicy rozmiar tego żądania, zwiększamy objętość sumaryczną żądań obecnych w systemie o wartość  $x1$ , dodatkowo jeśli kolejka żądań w pierwszym systemie

jest pusta, to generujemy czas obsługi tego żądania i ustalamy nową wartość dla zmiennej  $DMT2$ , w przeciwnym wypadku zwiększamy liczbę żądań oczekujących w kolejce w pierwszym systemie o 1; jeżeli przybywające żądanie nie zostanie przyjęte do systemu, to zwiększamy dla pierwszego systemu liczby i objętości odpowiednio utraconych żądań oraz ich rozmiarów;

- generujemy kolejną chwilę inicjalizacji żądania pochodzącego z pierwszego systemu i zmieniamy wartość zmiennej  $DMT1$ .

- Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT2$ , to wykonujemy poniższe operacje:

- zmniejszamy wartość sumarycznej objętości żądań o wartość pierwszego elementu dynamicznej tablicy rozmiarów żądań pochodzących z pierwszego systemu;

- zmniejszamy liczbę żądań pochodzących od pierwszego systemu o 1;

- jeżeli kolejka żądań w pierwszym systemie jest pusta to dodatkowo ustalamy wartość zmiennej  $DMT2$  na wartość większą od czasu modelowania  $TM$ ; w przeciwnym wypadku zmniejszamy liczbę żądań pochodzących od pierwszego systemu obecnych w kolejce o 1, generujemy czas obsługi dla kolejnego żądania pochodzącego z pierwszego systemu i ustalamy nową wartość dla zmiennej  $DMT2$ .

- Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT4$ , to wykonujemy następujące czynności :

- generujemy losowy rozmiar żądania pochodzącego z drugiego systemu  $x2$ ;

- zwiększamy sumaryczne liczby oraz objętości żądań pochodzących z drugiego systemu odpowiednio o 1 oraz o  $x2$ ;

- sprawdzamy czy generowane żądanie będzie przyjęte do systemu; jeśli tak to zwiększamy bieżącą liczbę żądań pochodzących z drugiego systemu o 1, zapisujemy na końcu dynamicznej tablicy rozmiar tego żądania, zwiększamy objętość sumaryczną żądań obecnych w systemie o wartość  $x2$ , dodatkowo jeśli kolejka żądań w drugim systemie jest pusta, to generujemy czas obsługi tego żądania i ustalamy nową wartość dla zmiennej  $DMT5$ , w przeciwnym wypadku zwiększamy liczbę żądań oczekujących w kolejce w drugim systemie o 1; jeżeli

przybywające żądanie nie zostanie przyjęte do systemu, to zwiększamy dla drugiego systemu liczby i objętości odpowiednio utraconych żądań oraz ich rozmiarów;

- generujemy kolejną chwilę inicjalizacji żądania pochodzącego z drugiego systemu i zmieniamy wartość zmiennej  $DMT4$ .

- Jeżeli uzyskanym minimum jest wartość zmiennej  $DMT5$ , to wykonujemy poniższe operacje:

- zmniejszamy wartość sumarycznej objętości żądań o wartość pierwszego elementu dynamicznej tablicy rozmiarów żądań pochodzących z drugiego systemu;

- zmniejszamy liczbę żądań pochodzących od drugiego systemu o 1;

- jeżeli kolejka żądań w drugim systemie jest pusta to dodatkowo ustalamy wartość zmiennej  $DMT5$  na wartość większą od czasu modelowania  $TM$ ; w przeciwnym wypadku zmniejszamy liczbę żądań pochodzących od drugiego systemu obecnych w kolejce o 1, generujemy czas obsługi dla kolejnego żądania pochodzącego z drugiego systemu i ustalamy nową wartość dla zmiennej  $DMT5$ .

- Jeżeli minimum jest wartością zmiennej  $TM$ , to kończymy symulację.

3. Obliczamy charakterystyki utrat na podstawie wartości zmiennych oznaczających odpowiednio liczby oraz objętości przybywających oraz utraconych żądań.

## 7.4 Omówienie działania programu dla modelu serwera - ograniczona kolejka

Idea działania programu symulacyjnego jest taka sama jak programu poprzedniego z tą różnicą, że losowe rozmiary żądań mają rozkład jednostajny, zatem wprowadzamy parametry  $a$ ,  $b$  tego rozkładu oraz mamy ograniczoną kolejkę - zatem gdy minimum będzie wartością zmiennej  $DMT1$  lub  $DMT4$ , to w przypadku, gdy sumaryczna liczba żądań niezależnie od rozmiarów przybywających żądań pochodzących z odpowiedniego serwera przekroczy liczbę  $N_i + 1$ ,  $i = 1, 2$ , to również zwiększamy dla odpowiedniego systemu liczby i objętości odpowiednio utraconych żądań oraz ich rozmiarów.





## Podsumowanie

Rozprawa doktorska jest poświęcona modelom i algorytmom wyznaczania objętości pamięci systemów informatycznych w warunkach niepewności stochastycznej.

Celem głównym pracy było opracowanie modeli i algorytmów wyznaczania objętości pamięci wybranych realnych systemów informatycznych o stochastycznym charakterze działania. W trakcie realizacji celu pracy badano dwa główne modele systemów informatycznych - model pamięci operacyjnej systemu komputerowego oraz model serwera obsługującego żądania wielu użytkowników. Poniżej przedstawiono główne rezultaty przeprowadzonych badań.

1. Zbudowano i przeanalizowano model pamięci operacyjnej RAM będący istotnym uogólnieniem modelu systemu obsługi zgłoszeń niejednorodnych  $M/G/1/\infty$ . W wyniku analizy otrzymano PLS dystrybuanty objętości sumarycznej w warunkach stacjonarnych, a następnie otrzymano formuły pozwalające obliczać pierwsze dwa momenty analizowanej zmiennej losowej.
2. W wyniku analizy modelu pamięci operacyjnej RAM zostały opracowane algorytmy pozwalające na szacowanie charakterystyk utrat informacji i w konsekwencji - szacowania rozmiaru potrzebnej pamięci w taki sposób, aby założone charakterystyki utrat nie były przekroczone.
3. Opracowane algorytmy wyznaczania rozmiaru potrzebnej pamięci są efektywne, a oszacowania rozmiaru pamięci z ich wykorzystaniem dają dokładniejsze wyniki w porównaniu z modelowaniem symulacyjnym w przypadku realnie występujących małych prawdopodobieństw utrat.
4. Uzyskane algorytmy dają się bezpośrednio stosować do pisania programów komputerowych, które w sposób numeryczny wyznaczają rozmiar potrzebnej pamięci realnych systemów informatycznych. Do obliczeń

numerycznych można również wykorzystać pakiety algebry komputerowej (Mathematica).

5. Otrzymane wyniki nie pozwalają na wyznaczanie w każdym przypadku dystrybuanty objętości sumarycznej w postaci jawnej. Jednak w niektórych przypadkach szczególnych jest to możliwe.
6. Zastosowanie modeli symulacyjnych w przypadku małych prawdopodobieństw utrat nie jest efektywne z uwagi na długi czas symulacji, który powoduje występowanie błędów związanych na przykład z okresowością generatorów losowych. Modele symulacyjne można natomiast stosować w przypadku dużych prawdopodobieństw utrat, ponieważ omawiana aproksymacja w przypadku dużych prawdopodobieństw utrat prowadzi do sytuacji, kiedy wybieramy rozmiar pamięci z dużym nadmiarem. Obie metody mogą być zatem stosowane łącznie, przy czym metoda analityczna jest efektywniejsza w przypadku małych, a symulacyjna w przypadku dużych prawdopodobieństw utrat.
7. Zbadano wiele przypadków szczególnych analizowanego modelu, w szczególności przypadek z niezależnym od rozmiaru procesu czasem obsługi oraz przypadek, gdy czas obsługi i rozmiar procesu są proporcjonalne.
8. Stworzono odpowiednie programy symulacyjne i przeprowadzono analizę porównawczą wyników otrzymanych podczas symulacji z wynikami otrzymanymi w sposób numeryczny.
9. Uzasadniono wartość praktyczną modelu pamięci RAM poprzez pokazanie technicznych realizacji omawianego modelu.
10. Zbudowano i przeanalizowano model serwera obsługującego żądania wielu użytkowników będący istotnym uogólnieniem systemu obsługi zgłoszeń niejednorodnych  $M/M/n/(m, V)$ . Uzyskano w postaci jawnej wzory określające rozkład liczby żądań.
11. Dla analizowanego modelu serwera uzyskano wzory pozwalające na obliczanie charakterystyk utrat informacji.
12. Wzory te pozwalają na obliczanie charakterystyk utrat żądań i mają bezpośrednie zastosowanie w wielu realnie występujących sytuacjach do oszacowania potrzebnego rozmiaru pamięci serwera.
13. Otrzymane wyniki można zastosować do szacowania rozmiaru pamięci buforowej w serwerach z dużym przepływem informacji (serwery www,

serwery FTP czy serwery pocztowe, sklepy internetowe, banki itp.).

14. Zbadano kilka interesujących przypadków szczególnych analizowanego modelu, w których występujące we wzorach sploty według Stieltjesa mogą być uzyskane w postaci jawnej.
15. Stworzono odpowiednie programy symulacyjne i przeprowadzono analizę porównawczą wyników otrzymanych podczas symulacji z wynikami otrzymanymi w sposób numeryczny.
16. Podczas analizy modelu serwera zauważono zależność między obciążeniami systemów komputerowych, a granicznymi wartościami charakterystyk utrat żądań.
17. Uzasadniono wartość praktyczną modelu serwera poprzez pokazanie technicznych realizacji omawianego modelu.

Cel pracy został osiągnięty.

Teza pracy została uzasadniona. Wyniki dla analizowanych modeli uzyskano wykorzystując metody teorii obsługi masowej zgłoszeń o losowej objętości, które pozwalają na zbudowanie modeli realnych systemów informatycznych. Wyniki uzyskane na podstawie analizy tych modeli mogą służyć do opracowania algorytmów wyznaczania charakterystyk utrat informacji i w konsekwencji do uzyskania metod wyznaczania objętości pamięci realnych systemów informatycznych.



---

## Literatura

- [1] APANASOVICH V.V., TIKHONENKO O.M. Modelowanie cyfrowe systemów stochastycznych. Uniwersitetskoe, Mińsk, 1986
- [2] BOROWKOW A.A. Rachunek prawdopodobieństwa. PWN, Warszawa, 1975.
- [3] BOROWKOW A.A. Erdodicity and Stability of Stochastic Processes. John Wiley and Sons, New York, 1998.
- [4] COX D.R. Renewal Theory. John Wiley and Sons, New York, 1961.
- [5] FELLER W. An Introduction to Probability Theory and its Applications. Vol. 1,2. John Wiley and Sons, New York, 1970, 1971.
- [6] GNIEDENKO B., KOWALENKO I. Wstęp do teorii obsługi masowej PWN, Warszawa, 1971.
- [7] FILIPOWICZ B. Modele stochastyczne w badaniach operacyjnych. WNT, Warszawa, 1996.
- [8] KARLIN S., TAYLOR H. A First Course in Stochastic Processes. Academic Press, New York, 1975.
- [9] KLEINROCK L. Queueing Systems. Vol. 1. John Wiley and Sons, New York, 1975.
- [10] KLIMOW G.P. Procesy obsługi masowej. WNT, Warszawa 1979.
- [11] KONIG D., STOYAN D. Metody teorii obsługi masowej. WNT, Warszawa, 1979.
- [12] LUKE Y.L. Mathematical Functions and their Approximations. Academic Press Inc. New York, 1975.

- [13] RUDIN W. Analiza rzeczywista i zespolona. PWN, Warszawa, 1998.
- [14] SMITH W.L. Renewal Theory and its Ramifications. Journal of the Royal Statistical Society, B, 20, (1958), No. 2.
- [15] TIKHONENKO O., TIKHONENKO A. Metody probabilistyczne. Wykłady i ćwiczenia dla informatyków. Oficyna Wydawnicza EWSIE, Warszawa, 2010.
- [16] ALEKSANDROV A.M., KATZ B.A. Non-homogeneous demands arrival servicing. Izv. AN SSSR. Tekhnicheskaya Kibernetika. 1973. N 2. P. 47-53 (in Russian)
- [17] APANASOVICH V.V., TIKHONENKO O.M. Distribution of the overall message length in a single channel queueing system. Eng. Cybern. 1984. Vol. 22, No.1, P. 63-67
- [18] CZACHÓRSKI T. Modele kolejkowe w ocenie efektywności pracy sieci i systemów komputerowych. Pracownia Komputerowa Jacka Skalmerskiego, Gliwice 1999
- [19] IVERSEN V.B. Teletraffic Engineering handbook. Geneva, 2005
- [20] KUROSE J.F., ROSS K.W. Computer Networking: A top-Down Approach, 4-th Edition, Addison-Wesley, 2007
- [21] PECHINKIN A.V. System with LIFO discipline and constrained total amount of items. Automation and Remote Control. 1998. Vol. 59, No. 4(2), P.545-553
- [22] PECHINKIN A.V. An system with LIFO discipline preemptive-resume discipline and bounded total customer volume. Automation and Remote Control. 2000. Vol. 60, No. 12(2). P.1767-1772
- [23] ROMM E.L., SKITOVICH V.V. About a generalization of the Erlang problem. Avtomatika Telemekhanika. 1971, No.6.P.164-167 (in Russian)
- [24] ROSS K.W. Multiservice loss models for broadband telecommunication networks. Springer-Verlag, London 1995
- [25] SENGUPTA B. The spatial requirements of M/G/1 queue or: how to design for buffer space. Lect. Notes Contr. Inf. Sci. 1984. Vol. 60. P. 547-562
- [26] SENGUPTA B., TAKINE T. Distribution of spatial requirements for a MAP/G/1 queue when space and time are dependent. Queueing Systems. 1996. Vol. 22. P. 121-127
- [27] SCHWARTZ M. Computer-communication network design and analysis. Prentice-Hall, Englewood Cliffs, New York, 1977

- [28] SCHWARTZ M. Telecommunication networks: protocols, modeling and analysis. Addison-Wesley Publishing Company, New York, 1987
- [29] SOHN E., KNESSL Ch. Storage allocation under processor sharing I: exact solutions and asymptotics. Queueing Systems Vol. 65. 2010, P.1-18
- [30] TIJMS H.C. A first course in stochastic models. Wiley and Sons, England 2003
- [31] TIKHONENKO O.M. Distribution of total message flow in single-server queueing system with group arrival. Automation and Remote Control. 1985. Vol. 46, No. 11. P. 1412-1416
- [32] TIKHONENKO O.M. The problem of determination of the summarized messages volume in queueing system and its applications. Journal of Information Processing and Cybernetics. 1987. Vol. 32, No. 7. P. 339-352
- [33] TIKHONENKO O.M. Distribution of total flow of calls in group arrival queueing systems. Automation and Remote Control. 1987. Vol. 48, No. 11. P. 1500-1507
- [34] TIKHONENKO O.M. Determining the characteristics of the total amount of calls in single-server priority queueing systems. Automation and Remote Control. 1989. Vol. 50, No. 2. P. 213-220
- [35] TIKHONENKO O.M. Queueing models in computer systems. Universitetskoe, Minsk, 1990 (in Russian)
- [36] TIKHONENKO O.M. Queueing systems of a random length with restrictions. Automation and Remote Control. 1991. Vol. 52, No.10(2). P. 1431-1437
- [37] TIKHONENKO O.M. An analog of Little's formula for queueing systems with non-homogeneous tasks. Automation and Remote Control. 1996. Vol. 57, No.1(2). P. 85-88
- [38] TIKHONENKO O.M. Districted capacity queueing systems: Determination of their characteristics. Automation and Remote Control. 1997. Vol. 58, No.6(1). P.969-972
- [39] TIKHONENKO O.M. The summarized volume distribution of demands in priority single-server systems. Intern. Conference "Distributed Computer Communication Networks".Proceedings. Nov. 4-8, 1997, Tel-Aviv(Israel). P.207-211
- [40] TIKHONENKO O.M. Distribution of the total load in a single-server system with exponential service and recurrent input flow. Automation and Remote Control. 1999. Vol. 60, No. 7(1), P. 970-973
- [41] TIKHONENKO O.M. Determination of the characteristics of the total amount of customers in queues with absolute priority. Automation and Remote Control. 1999. Vol. 60, No. 8(2). P. 1205-1210

- [42] TIKHONENKO O.M., KLIMOVICH K.G. Queueing systems for random-length arrivals with limited cumulative volume. *Problems of Information Transmission*. 2001. Vol.37, No. 1. P. 70-79
- [43] TIKHONENKO O.M. Generalized Erlang Problem for Service Systems with Finite Total Capacity. *Problems of Information Transmission*. 2005. Vol. 41, No.3. P. 243-253
- [44] TIKHONENKO O. Metody probabilistyczne analizy systemów informacyjnych. *Akademicka Oficyna Wydawnicza EXIT*, Warszawa, 2006
- [45] TIKHONENKO O., ZIÓŁKOWSKI M. Queueing Systems with Common Memory Space. *Prace naukowe Akademii im. Jana Długosza w Częstochowie zeszyt 10*, (2005) (str. 131-149)
- [46] TIKHONENKO O., ZIÓŁKOWSKI M. Queueing Systems with Non-Homogeneous Internal and External Demands, materiały konferencyjne XIII konferencji naukowej „Czech-Polish-Slovak Mathematical School” , Hlubos 2-4 czerwca 2005 (str. 66-71)
- [47] TIKHONENKO O., ZIÓŁKOWSKI M. Queueing Systems with Random Volume Internal and External Calls. *Prace naukowe Akademii im. Jana Długosza w Częstochowie zeszyt 11*, (2006)(str. 103-120)
- [48] TIKHONENKO O., ZIÓŁKOWSKI M. Combinations of Queueing Systems with Common Memory Space., materiały konferencyjne międzynarodowej konferencji naukowej „Mathematical Methods for Increasing Efficiency of Information Telecommunication Networks”, Grodno 29 stycznia -1 lutego 2007(str. 236-242)
- [49] TIKHONENKO O., GOŁA A., ZIÓŁKOWSKI M. Estimations of loss characteristics of single-server queueing systems with non-homogeneous customers. *Prace naukowe Akademii im. Jana Długosza w Częstochowie zeszyt 13*, (2008) (str. 53-66)
- [50] TIKHONENKO O., GOŁA A., ZIÓŁKOWSKI M. An influence of service discipline on characteristics of a single-server queue with non-homogenous customers. *Prace naukowe Akademii im. Jana Długosza w Częstochowie zeszyt 15*, (2010) (str. 191-200)
- [51] ZIÓŁKOWSKI M. Generalization of probability density of random variables. *Prace naukowe Akademii im. Jana Długosza w Częstochowie zeszyt 14*, (2009) (str. 165-172)