ZESZYTY
NAUKOWE
POLITECHNIKI
ŚLĄSKIEJ

ANDRZEJ WILK

MODELE KOLEJKOWE Z SEMAFORAMI

MODELOWANIE PROTOKOŁU KOMUNIKACYJNEGO X.25

INFORMATYKA

P. 4193/91

Z. 17
GLIWICE
1991

POLITECHNIKA ŚLĄSKA

ZESZYTY NAUKOWE

Nr 1141

ANDRZEJ WILK

MODELE KOLEJKOWE Z SEMAFORAMI

MODELOWANIE PROTOKOŁU KOMUNIKACYJNEGO X.25

GLIWICE 1991

OPINIODAWCY
Prof. zw. dr hab. Andrzej Grzywak
Prof. dr hab. inż. Jan Węglarz

KOLEGIUM REDAKCYJNE

REDAKTOR NACZELNY — Prof. dr hab. inż. Jan Bandrowski
REDAKTOR DZIAŁU — Dr inż. Marcin Skowronek
SEKRETARZ REDAKCJI — Mgr Elżbieta Leśko

REDAKCJA
Mgr Roma Łoś

REDAKCJA TECHNICZNA
Alicja Nowacka

Wydano za zgodą
Rektora Politechniki Śląskiej

PL ISSN 0208-7286

Wydawnictwo Politechniki Śląskiej
ul. Kujawska 3, 44-100 Gliwice

Nakł. 170+55 **Ark. wyd.** 8,25 **Ark. druk.** 6,25 **Papier offset.** kl.III 70x100, 70g
Oddano do druku 1.10.91 **Podpis. do druku** 1.10.91 **Druk ukończ.** w październiku 1991
Zam. 437/91 **Cena** zł 11 600,—

Folokopie, druk i oprawę
wykonano w Zakładzie Graficznym Politechniki Śląskiej w Gliwicach

Spis treści

Str.

1. Wprowadzenie 7

1.1. Modelowanie systemów informatycznych ... 7
1.2. Cel i zakres pracy ... 18
1.3. Pojęcie stacji semafora w modelu kolejkowym ... 20
1.4. Analiza stacji semafora ... 25

2. Model kolejkowy z semaforami 31

2.1. Wprowadzenie ... 31
2.2. Analiza modelu ... 33
2.3. Ocena dokładności metody analizy ... 36
2.4. Podsumowanie ... 44

3. Sieci kolejek z semaforami 46

3.1. Wprowadzenie ... 46
3.2. Model ... 47
3.3. Agregacja podsieci kolejek ... 49
3.4. Zamknięta sieć kolejek ... 52
3.5. Otwarta sieć kolejek ... 57
3.6. Podsumowanie ... 65

4. Modelowanie protokołu X.25 66

4.1. Wprowadzenie ... 66
4.2. Prezentacja systemu ... 66
4.3. Analiza modelu ... 70
4.4. Obliczenia ... 75
4.5. Symulacja ... 78
4.6. Podsumowanie ... 79

Zakończenie 80

Literatura 85

Streszczenia 94

Contents

Page

1. Introduction 7

1.1. Computer systems modelling 7
1.2. Obiective and scope of study 18
1.3. Notion of semaphore station in network of queues 20
1.4. Analysis of semaphore station 25

2. Queueing network model with semaphores 31

2.1. Introduction 31
2.2. Model analysis 33
2.3. Method validation 36
2.4. Summary 44

3. Networks of queues with semaphores 46

3.1. Introduction 46
3.2. Model 47
3.3. Queueing subnetwork aggregation 49
3.4. Closed queueing network 52
3.5. Open queueing network 57
3.6. Summary 65

4. X.25 protocol modelling 66

4.1. Introduction 66
4.2. System presentation 66
4.3. Model analysis 70
4.4. Calculations 75
4.5. Simulation 78
4.6. Summary 79

Conclusions 80

Literature 85

Abstracts 94

Содержание

Стр.

1. **Введение** 7

1.1. Моделирование информационных систем 7
1.2. Цель и объём работы 18
1.3. Понятие станции семафора в модели системы массового обслуживания 20
1.4. Анализ станции семафора 25

2. **Модель системы массового обслуживания с семафорами** 31

2.1. Введение 31
2.2. Анализ модели 33
2.3. Оценка точности метода анализа 36
2.4. Резюме 44

3. **Сети очередей с семафорами** 46

3.1. Введение 46
3.2. Модель 47
3.3. Соединение подсетей очередей 49
3.4. Закрытая сеть очередей 52
3.5. Открытая сеть очередей 57
3.6. Резюме 65

4. **Моделирование протокола X.25** 66

4.1. Введение 66
4.2. Представление системы 66
4.3. Анализ модели 70
4.4. Расчёты 75
4.5. Симуляция 78
4.6. Резюме 79

Заключение 80

Литература 85

Короткие содержания 94

Rozdział 1

WPROWADZENIE

1.1. MODELOWANIE SYSTEMÓW INFORMATYCZNYCH

Modelowanie systemów informatycznych od lat stało się uznanym narzędziem ich konfigurowania i analizy prawidłowego funkcjonowania. Celem modelowania jest analiza ilościowa lub jakościowa dostarczająca informacji co do zachowania się systemu z punktu widzenia jego efektywności. W tak ujętym zagadnieniu wyszczególnia się dwa aspekty badań operujące zgoła odmiennymi językami opisu problemu i tym samym dające jakościowo różne wyniki modelowania. Są to:

- badania dotyczące własności struktury analizowanego systemu;
- badania dotyczące ilościowej oceny pracy systemu.

Sieci Petri [BRAM83] nie uwzględniające zależności czasowych bardzo dobrze nadają się do opisu logicznego zachowania się analizowanego systemu. Dostarczają one klarownego modelu graficznego odzwierciedlającego operacje zachodzące w systemie, dzięki czemu są szczególnie przydatne przy opisie zjawisk konkurencji, konfliktu i synchronizacji. Za pomocą tego typu modeli otrzymuje się zwykle analizę jakościową badanego systemu przedstawianą w kategoriach *żywy*, *bezpieczny* itp., pozwalającą stosować te modele np. do oceny niezawodności. Ostatnie prace nad tymi modelami poświęcone były uwzględnieniu czynnika czasu [NATK85], [CHRE84], [HOLL85]. Sieci Petri z czynnikiem czasu pozwalają już na pewną ilościową analizę wydajności. W tym zakresie dużego znaczenia nabrały stochastyczne sieci Petri, których parametry zadawane, w tym czasowe, wyrażone są za pomocą zmiennych losowych [MOLL82], [BALB85], [AJMO87], [MAGO89].

Alternatywą modelowania za pomocą sieci Petri są modele kolejkowe [KLEI75] uwzględniające aspekt ilościowej oceny pracy badanych systemów informatycznych. Metody

analizy oparte na modelach kolejkowych pozwalają, jak dotąd, najlepiej estymować parametry wydajnościowe badanych systemów, stąd są z pewnością najczęściej używane i od ponad dwudziestu lat uwaga badaczy jest głównie im poświęcona. Ponadto, istotnym argumentem przemawiającym za modelami kolejkowymi są związane z nimi szybkie i oszczędne techniki analizy oparte głównie na obliczeniach analitycznych. Z drugiej strony klasa modeli kolejkowych, które można rozwiązać w sposób dokładny, jest mocno ograniczona i nie pozwala uwzględnić wielu często spotykanych własności badanych systemów (architektury wieloprocesorowe, systemy rozproszone, obliczenia współbieżne ...). Można tu wspomnieć o jednoczesnym posiadaniu kilku zasobów, blokowaniu, konkurencji w dostępie do wspólnych zasobów lub zależności pomiędzy elementami sieci kolejek (stanowiskami obsługi) lub krążącymi po niej klientami (synchronizacja, generacja lub integracja klientów). Te zależności wynikają z interakcji występującej między różnymi elementami badanego systemu i są trudne do odzwierciedlenia w modelu kolejkowym. Z kolei ignorowanie tych własności ogranicza zakres stosowalności modeli kolejkowych lub prowadzi do otrzymania nieprecyzyjnych wyników.

Mimo występujących trudności metody kolejkowe są stosowane coraz powszechniej do estymacji ilościowej parametrów wydajnościowych badanych systemów. Aktualne podejścia zmierzają w kierunku opracowania i wykorzystania metod przybliżonych, których jakość weryfikuje się poprzez symulację. Zagadnieniu temu poświęcona jest również niniejsza praca. W celu wykorzystania mniejszej złożoności obliczeniowej modeli kolejkowych i większych możliwości opisowych sieci Petri podejmowane są również próby łączenia modeli kolejkowych z sieciami Petri [BALB86], [BALB88].

A. Model kolejkowy pojedynczego stanowiska obsługi

Podstawowym elementem modelu kolejkowego jest stanowisko obsługi składające się z bufora i kanałów obsługi. Do stanowiska w kolejnych chwilach czasu przybywają klienci, którzy są obsługiwani w kanałach obsługi (jeden klient w jednym kanale). Klienci, którzy zastaną wszystkie kanały obsługi zajęte, mogą tworzyć kolejkę w buforze. Sposób ustawienia klientów oczekujących w kolejce odbywa się według reguł określonych regulaminem szeregowania. Zwykle klienci ustawiają się jeden za drugim w kolejności, w jakiej nadeszli do stanowiska. Taki regulamin szeregowania zwany jest naturalnym albo *FIFO*.

Wobec powyższego stanowisko obsługi charakteryzują następujące parametry:

- rozkład losowy odstępów czasu t_i pomiędzy kolejnymi nadejściami klientów zwany dalej strumieniem przybyć klientów;

- rozkład losowy czasów obsługi τ_i kolejnych klientów zwany dalej strumieniem obsługi klientów;
- liczba kanałów obsługi k;
- regulamin szeregowania klientów w kolejce.

Zakłada się, że odstępy czasu t_i (analogicznie τ_i) są niezależnymi zmiennymi losowymi opisanymi tym samym rozkładem. Przedstawione funkcjonowanie stanowiska wyznacza proces $n(t)$ liczby klientów na stanowisku w chwili t. Jeśli stanowisko nie jest przeciążone, tzn., że jest ono w stanie obsłużyć wszystkich przybywających klientów, a ich liczba nie rośnie wraz z czasem do nieskończoności, to rozwiązanie analityczne stanowiska obsługi daje wartości prawdopodobieństw $p(n) = \lim_{t\to\infty} p(n(t) = n)$ stanu stacjonarnego znalezienia n, $n = 0, \ldots, \infty$, klientów na stanowisku.

Załóżmy, że strumień przybyć klientów, jak i strumień obsługi mają charakter poissonowski (rozkład wykładniczy):

$$\begin{aligned} p(t_i < t) &= A(t) = 1 - e^{-\lambda t}, \\ p(\tau_i < t) &= B(t) = 1 - e^{-\mu t} \end{aligned}$$

oraz że stanowisko posiada tylko jeden kanał obsługi, a klienci w kolejce szeregowani są według regulaminu naturalnego. Przy takich założeniach prawdopodobieństwa stanu stacjonarnego wyrażają zależności:

$$\begin{aligned} p(n) &= \rho^n(1-\rho) \qquad n > 0, \\ p(0) &= 1 - \rho, \end{aligned}$$

z $\rho = \lambda/\mu$ i warunkiem istnienia stanu stacjonarnego (nieprzeciążenia stanowiska): $\rho < 1$.

Prawdopodobieństwa $p(n)$ pozwalają obliczyć:

— średnią liczbę klientów na stanowisku:

$$\mathcal{L} = \sum_{n=0}^{\infty} n \cdot p(n) = \frac{\rho}{1-\rho},$$

— czas reakcji stanowiska (średni czas od momentu przybycia klienta do stanowiska do momentu opuszczenia stanowiska. W czas ten wliczony jest średni czas oczekiwania w kolejce). Czas reakcji liczy się przy użyciu formuły Little'a [LITT61]:

$$\mathcal{R} = \frac{\mathcal{L}}{\lambda},$$

— średni czas oczekiwania w kolejce do kanałów obsługi:

$$W = \mathcal{R} - \frac{1}{\mu}.$$

B. Sieć kolejek

Zbiór stanowisk obsługi wraz z marszrutami klientów od stanowiska do stanowiska wyznaczającymi sieć powiązań stanowisk nazywany jest siecią kolejek. Sieć może być otwarta – istnieje strumień wejściowy i wyjściowy klientów – lub zamknięta – gdy w sieci krążą wciąż ci sami klienci.

Sieć kolejek zdefiniowana jest strumieniem przybywających klientów (jeśli istnieje), rozkładami losowymi czasów obsługi na kolejnych stanowiskach obsługi, regulaminami szeregowania klientów w kolejkach oraz drogami przejść klientów od stanowiska do stanowiska. Klienci krążący w sieci kolejek mogą się dzielić na klasy. Poszczególne klasy klientów mogą mieć różne czasy obsługi na stanowiskach, mogą też różnić się drogą w sieci. Dla jednych klas klientów sieć kolejek może być siecią otwartą z zewnętrznym strumieniem przybywających klientów, dla innych – zamkniętą. Dobierając odpowiednią konfigurację sieci kolejek, można zamodelować niemal każdy system masowej obsługi. Poszczególne stanowiska sieci mogą reprezentować, np. procesory, pamięci, terminale, a cyrkulujący po nich klienci – programy, zgłoszenia, przerwania, transmitowane bloki danych itp.

Niech $\mathbf{N} = (N_1, \ldots, N_M)$ wyraża stan sieci kolejek złożonej z M stanowisk obsługi. Poszczególne $N_i = (n_{i1}, \ldots, n_{iK})$, $i = 1, \ldots, M$, reprezentują populacje klientów n_{ik} poszczególnych klas k, $k = 1, \ldots, K$, na stanowisku i modelu. Jeśli żadne ze stanowisk nie jest przeciążone, to rozwiązanie analityczne modelu kolejkowego pozwala otrzymać prawdopodobieństwa $p(\mathbf{N})$ stanu stacjonarnego sieci. Prawdopodobieństwa te służą następnie do wyznaczenia jej miar wydajności, do których należą:

- średni czas przejścia przez każde ze stanowisk sieci, a tym samym sumaryczne czasy przejścia przez kilka stanowisk lub od punktu do punktu sieci;
- średnie liczby klientów na stanowiskach;
- wykorzystanie każdego ze stanowisk rozumiane jako procent czasu, w którym na stanowisku przebywał minimum jeden klient;
- przepustowość każdego ze stanowisk rozumiana jako średnia liczba klientów opuszczających obiekt w jednostce czasu.

C. Sieć kolejek BCMP

Nie dla każdej sieci kolejek można znaleźć dokładne rozwiązanie w jawnej (nieuwikłanej) postaci algebraicznej. Jak dotąd najogólniejszym modelem kolejkowym posiadającym takie rozwiązanie jest sieć kolejek spełniająca założenia teorii *BCMP* [BASK75] (z pewnymi szczególnymi rozszerzeniami, które tu nie będą poruszane). Idea teorii *BCMP* bazuje na założeniu o niezależności rozwiązań poszczególnych stanowisk sieci kolejek, pozwalającej przedstawić prawdopodobieństwa stanu stacjonarnego sieci w postaci tzw. formy iloczynowej:

$$p(\mathbf{N} = (N_1, \ldots, N_M)) = \frac{1}{\mathbf{H}} \cdot \Lambda(\mathbf{N}) \cdot \mathbf{V}(N_1) \cdot \ldots \cdot \mathbf{V}(N_M) \ ,$$

gdzie: $\mathbf{H}$ — jest stałą normalizacyjną sprawiającą, że prawdopodobieństwa wszystkich stanów sumują się do 1;

$\Lambda(\mathbf{N})$ — jest wyrażeniem ujmującym strumienie przybywających do sieci klientów. Dla zamkniętej sieci kolejek: $\Lambda(\mathbf{N}) = 1$;

$\mathbf{V}(N_i)$ — jest wyrażeniem ujmującym czas obsługi klienta na stanowisku i przeskalowany w proporcji wizyt klientów na tym stanowisku w stosunku do innych stanowisk sieci.

Ogólnie rzecz biorąc, spełnienie założenia o niezależności wymaga zachowania charakteru strumieni wejściowego i wyjściowego klientów do/z poszczególnych stanowisk obsługi modelu kolejkowego oraz wzajemnej niezależności parametrów obsługi poszczególnych stanowisk. Sieć kolejek spełniająca te założenia ma następujące cechy:

- Strumień klientów przybywających do sieci ma rozkład wykładniczy z parametrem λ. Parametr λ może być zależny od łącznej populacji klientów w sieci bądź, jeśli dotyczy klientów klasy k, od populacji klientów tej klasy w sieci.

- Stanowiska obsługi sieci kolejek *BCMP* mogą być jednym z czterech następujących typów:

  1. *FIFO* — Regulamin szeregowania klientów jest naturalny. Strumień obsługi klientów ma rozkład wykładniczy z parametrem μ, który może zależeć od populacji klientów na stanowisku. Strumień ten jest identyczny dla wszystkich klas klientów przechodzących przez stanowisko.
  2. *PS* — z podziałem procesora. Każdy klient przybywający do stanowiska jest obsługiwany natychmiast. Szybkość obsługi maleje proporcjonalnie do liczby

klientów na stanowisku. Czasy obsługi mają rozkład Coxa i mogą być różne dla różnych klas klientów.

3. *IS* – Liczba kanałów obsługi tego stanowiska jest dostatecznie duża, aby przyjąć wszystkich przybywających klientów (w szczególności nieskończona). Rozkład czasu obsługi klientów na stanowisku jest typu Coxa i może zależeć od klasy klientów.

4. *LCFS-PR* — regulamin stosowy z wywłaszczeniem. Nadchodzący klient wstrzymuje obsługę klienta w kanale obsługi i zajmuje jego miejsce. Obsługa wstrzymanego klienta jest kontynuowana po zwolnieniu kanału obsługi. Rozkład czasu obsługi klientów w kanale obsługi jest typu Coxa i może zależeć od klasy klientów.

Duży nakład obliczeniowy związany zwłaszcza z obliczaniem stałej normalizacyjnej przyczynił się do opracowania kilku efektywnych metod pozwalających znaleźć miary wydajności analizowanej sieci kolejek *BCMP* bez konieczności obliczania wszystkich prawdopodobieństw stanu. Najbardziej popularną metodą jest metoda wartości średnich *MVA* [REIS80] odnosząca się do zamkniętych sieci kolejek. Idea metody opiera się na stwierdzeniu, że czas reakcji $\mathcal{R}_i(N)$ stanowiska i w sieci z N klientami jest równy sumie czasu obsługi $\frac{1}{\mu_i}$ oraz czasowi czekania (nie dotyczy to stanowiska *IS*) wyrażonemu przez $\frac{1}{\mu_i} \cdot \mathcal{L}_i(N-1)$, gdzie $\mathcal{L}(N-1)$ jest średnią liczbą klientów obecnych na stanowisku w przypadku, gdy w sieci krąży $N-1$ klientów:

$$R_i(N) = \frac{1}{\mu_i}(1 + \mathcal{L}_i(N-1)) .$$

Wzór powyższy wraz z formułą Little'a [LITT61] wiążącą czas reakcji $\mathcal{R}$, średnią liczbę klientów $\mathcal{L}$ i przepustowość $\mathcal{T}$:

$$\mathcal{T} = \frac{\mathcal{L}}{\mathcal{R}}$$

pozwala otrzymać parametry wydajnościowe sieci kolejek w sposób rekurencyjny dla stopniowo rosnącej populacji N klientów w sieci.

D. Sieci kolejek nie spełniające założeń BCMP

Założenia teorii *BCMP* zapewniając rozwiązanie dokładne wnoszą istotne ograniczenia co do postaci analizowanego modelu kolejkowego. W sieci kolejek nie dopuszcza się innych (np. priorytetowego) regulaminów szeregowania klientów ani żadnej interakcji pomiędzy

klientami lub stanowiskami obsługi. Tymczasem dla większości systemów informatycznych (sieci komputerowe, systemy rozproszone, maszyny wieloprocesorowe) trudno jest zbudować model spełniający te założenia. Zadania wykonywane w tych systemach, korzystając z zasobów sprzętowych i programowych systemu, współdziałają i komunikują się ze sobą przy zachowaniu określonych reguł synchronizacji. Wprowadza to zależności między elementami składowymi systemu i konieczność ich uwzględnienia w procesie modelowania. Wynikające z dzielenia zasobów, najczęściej spotykane cechy systemu to:

— *równoczesne posiadanie kilku zasobów.* Przykładowo, zaangażowanie procesora do wykonania programu wymaga uprzedniego przydzielenia programowi pamięci operacyjnej.

— *blokowanie.* Proces, korzystając z kolejnych zasobów systemu, może zostać wstrzymany, jeśli następny potrzebny mu do wykonania zasób jest jeszcze zajęty. Dla przykładu, w sieci komputerowej przesyłany pakiet może być przytrzymany w jednym z węzłów sieci, jeśli następny węzeł chwilowo (np. z braku wolnych buforów) nie może tego pakietu przyjąć.

— *wzajemne wykluczanie.* Przykładowo, operacja zapisu do rekordu bazy danych nie może być jednocześnie wykonywana przez więcej niż jeden program. Operacja zapisu jest zatem sekcją krytyczną wymagającą indywidualnego dostępu.

Dla sieci kolejek uwzględniających powyższe właściwości nie ma rozwiązań dokładnych. Aktualne podejścia zmierzają w kierunku opracowania i wykorzystywania metod przybliżonych, których jakość weryfikuje się poprzez symulację.

Równoczesne posiadanie kilku zasobów — Temu specyficznemu problemowi zostało poświęconych wiele publikacji w aspekcie analizy wydajności. Jacobson i Lazowska [JACO82] zaproponowali metodę przybliżoną, którą stosuje się do zespołu rozłącznych zasobów: zasobu pierwotnego i zasobu wtórnego. Klient, otrzymawszy element zasobu pierwotnego, jest obsługiwany i ma prawo wystąpić o przydział zasobu wtórnego. Element zasobu pierwotnego jest tak długo w posiadaniu klienta, dopóki ten nie uwolni zasobu wtórnego. Koncepcja metody bazuje na rozdzieleniu modelu kolejkowego na dwie składowe reprezentujące:

1. oczekiwanie na przydział zasobu wtórnego łącznie z czasem obsługi przy przydziale zasobu pierwotnego;
2. oczekiwanie na przydział zasobu pierwotnego.

Dla dwóch składowych zaproponowano dwa uzupełniające się modele kolejkowe. W każdym z modeli druga składowa jest reprezentowana przez równoważne stanowisko obsługi typu *IS*. Przybliżone rozwiązanie otrzymuje się w sposób iteracyjny, analizując na przemian oba modele.

Freund i Bexfield [FREU83] zaproponowali metodę przybliżoną bazującą na dekompozycji i pozwalającą studiować system z jednoczesnym posiadaniem zasobów za pomocą równoważnej zamkniętej sieci kolejek. Model wyjściowy składa się z podsieci Q_1 spełniającej założenia teorii *BCMP* i Q_2 reprezentującej przydzielane zasoby. Podsieć Q_2 jest zastępowana stanowiskiem obsługi o wykładniczym strumieniu obsługi zależnym od stanu (od liczby klientów na stanowisku). Otrzymana sieć kolejek posiada już rozwiązanie w formie iloczynowej i może być rozwiązana przy użyciu znanych metod dokładnych.

W [SAUE81b] Sauer rozszerzył metodę przybliżoną analizy sieci kolejek z jednoczesnym posiadaniem zasobów i jedną klasą klientów na sieci tego samego typu, ale dla wielu klas klientów. Rozwiązanie opiera się również na transformacji sieci kolejek G w sieć kolejek G^* w sposób zachowujący równoważność strumieni przybywających klientów. Sieć G^* posiada stanowisko odpowiadające zespołowi stanowisk S w sieci G reprezentującemu przydzielane zasoby. Dla dwóch klas klientów strumień obsługi stanowiska zastępczego $\mu(n_1, n_2)$ jest równy przepustowości wyizolowanego, zamkniętego zespołu stanowisk S z n_1 klientami klasy 1 i n_2 klientami klasy 2. Przy założeniu, że rozwiązanie stanowiska zastępczego da się przedstawić w postaci formy iloczynowej, sieć kolejek G^* można rozwiązać znanymi sposobami. Dokładność metody wydaje się być zadowalająca, ale tylko dla pewnych typów zastosowań.

Blokowanie — Model kolejkowy uwzględniający mechanizm blokowania jest przedstawiony w postaci szeregu M stanowisk obsługi ustawionych jeden za drugim (w tandemie). Z wyjątkiem pierwszego stanowiska modelu, maksymalna liczba klientów, którzy mogą przebywać w każdym z następnych stanowisk, jest ograniczona. Klient po zakończonej obsłudze może przejść do następnego stanowiska pod warunkiem, że jest ono w stanie go przyjąć, tzn. liczba klientów tego stanowiska nie osiągnęła nałożonego ograniczenia. W przeciwnym przypadku klient zmuszony jest do czekania i blokuje aktualnie zajmowany kanał obsługi. Rozwiązanie modelu polega głównie na znalezieniu maksymalnej jego przepustowości lub, co jest równoważne, warunku istnienia stanu stacjonarnego.

Jacobson i Lazowska w [JACO83] zaproponowali rozwiązanie modelu kolejkowego, którego fragment reprezentuje powyższy mechanizm blokowania. Technika

rozwiązania obejmuje 3 fazy. W pierwszej fazie analizowana sieć kolejek jest przekształcana w sieć równoważną zawierającą w miejsce powyższego fragmentu sieci podsystem o limitowanej populacji klientów. Ideą metody jest markowanie klasy klienta wizytującego podsystem. Wszystkie ograniczenia na tym zasobie mogą być zatem wyrażone ograniczoną liczbą klientów tej klasy przyjętych do podsystemu. Następnie używa się metody przybliżonej do rozwiązania otrzymanej równoważnej sieci kolejek. W ostatniej fazie otrzymane wyniki służą do wyznaczenia miar wydajności modelu pierwotnego.

Agrawal i Buzen w [AGRA83] przedstawili metodę, w której przejścia przez kolejne stanowiska modelu są reprezentowane za pomocą równoważnych stanowisk obsługi. Powiększone czasy obsługi stanowisk równoważnych uwzględniają czasy blokowania klientów. Współczynniki powiększenia czasu obsługi są obliczane iteracyjnie.

Wzajemne wykluczenie — Zasób wymagający indywidualnego dostępu nazywany jest sekcją krytyczną. Sieć kolejek uwzględniająca ten mechanizm synchronizacji jest przedstawiona np. przez Smitha i Browne'a w [SMIT80]. Oczekiwanie klienta przed sekcją krytyczną jest reprezentowane przez dodatkowe stanowisko. Klient przybywający do sekcji krytycznej z prawdopodobieństwem, że sekcja krytyczna jest wolna, wchodzi do niej, natomiast z prawdopodobieństwem zajętości sekcji krytycznej jest kierowany do dodatkowego stanowiska. Prawdopodobieństwo, że dla przybywającego klienta sekcja krytyczna jest wolna, jest tożsame z prawdopodobieństwem, że wszyscy pozostali klienci są poza nią w pozostałej części sieci. Czas obsługi na dodatkowym stanowisku jest obliczany iteracyjnie przy zachowaniu średniej liczby klientów sekcji krytycznej równej jeden. Metoda daje poprawne wyniki jedynie w przypadku silnego obciążenia sekcji krytycznej.

Modelowaniem oczekiwania przed sekcją krytyczną zajęli się również Thomasian i Bay, którzy w [THOM83] zaproponowali dwie metody analityczne bazujące na iteracji i dekompozycji. Zaprezentowana technika iteracji jest podobna do tej zaproponowanej przez Smitha i Browne'a w pracy [SMIT80]. Aby uwzględnić opóźnienie spowodowane oczekiwaniem na przejście przez sekcję krytyczną, zostały użyte stanowiska z mechanizmem semaforowania. W metodzie iteracyjnej oblicza się kolejne przybliżenia czasu czekania przed sekcją krytyczną. Metoda iteracyjna została zweryfikowana przez symulację dla małych obciążeń sekcji krytycznej i trzech różnych sposobów wyznaczania prawdopodobieństwa zajętości sekcji krytycznej. Metoda dekompozycji obejmuje dwa etapy. W pierwszym etapie liczy się przepustowości wy-

izolowanych fragmentów modelu. Etap drugi dotyczy analizy łańcucha Markowa stanów systemu, w którym wartości strumieni przejść pomiędzy stanami odpowiadają obliczonym przepustowościom z etapu 1. Metoda dekompozycji dostarcza bardziej satysfakcjonujących rezultatów, ale wymaga czasochłonnych obliczeń.

E. Modelowanie sieci komputerowych

W typowej sieci komputerowej zarówno lokalnej, jak i rozległej jednostki danych (np. pakiety, ramki), zanim osiągną miejsce przeznaczenia, przechodzą przez kilka poziomów mechanizmów sterowania przepływem za pomocą okna (window flow control mechanism). Wirtualna droga przepływu w takiej sieci była dla przykładu badana przez Schwartza w [SCHW82] za pomocą zamkniętej sieci kolejek w postaci tzw. systemu ze stratami. System taki zakłada, że pakiety nadmiarowe, których mechanizm sterowania przepływem nie może przepuścić ze względu na przekroczenie rozmiaru okna, ulegają straceniu. Model kolejkowy stanowi zespół stanowisk obsługi ustawionych w linii, w którym każde ze stanowisk reprezentuje węzeł wirtualnej drogi przepływu danych w sieci komputerowej. Wpływ obcych strumieni danych dopływających i odpływających do poszczególnych węzłów jest uwzględniony przez odpowiednie zredukowanie szybkości obsługi w odpowiadającym węzłowi stanowisku modelu.

Jeszcze wcześniej Reiser w [REIS79] zamodelował system komunikacyjny z wieloma różnymi drogami i strumieniami przepływu danych, ale bez uwzględnienia konieczności przesłania potwierdzeń odbioru. Jego model kolejkowy stanowi wieloklasowa zamknięta sieć stanowisk obsługi również w postaci systemu ze startami. Każdej klasie klientów odpowiada niezależna droga przepływu przez sieć. Do analizy sieci Reiser zaproponował efektywny przybliżony algorytm bazujący na metodzie analizy średnich, przeznaczony dla dużych zamkniętych wieloklasowych modeli kolejkowych.

W rzeczywistych sieciach komputerowych pakiety nadmiarowe przekraczające rozmiar okna zwykle nie są tracone, lecz przez mechanizm sterowania przepływem ustawiane w kolejkę oczekujących na przesłanie. W [REIS81b] Reiser podjął próbę uwzględnienia tego czasu oczekiwania stosując model kolejkowy typu otwartego. Z modelem tym związana jest pula W żetonów, gdzie W odpowiada rozmiarowi okna przepływu. Klient może przejść przez sieć, jeśli posiada żeton. Jeśli przybywający klient zastaje pustą pulę żetonów, jest zmuszony do czekania w kolejce wejściowej do sieci. Gdy klient po przejściu przez sieć opuszcza ją, zwraca posiadany żeton, który po pewnej zwłoce pojawia się ponownie w puli umożliwiając przejście przez sieć następnemu klientowi oczekującemu na jej

wejściu. Zwłoka w ponownym pojawieniu się żetonu w puli odpowiada czasowi przesłania w sieci komputerowej potwierdzenia odbioru pakietu. Model ten jest następnie w sposób przybliżony redukowany do równoważnego stanowiska zastępczego, co znacznie upraszcza jego analizę.

Varghese, Chou i Nilsson w [VARG83] używając tej samej przybliżonej metody zanalizowali podobny model kolejkowy, ale bez uwzględnienia czasu przesłania potwierdzenia odbioru pakietu. Model ich dopuszcza jednak bardziej ogólny rozkład Coxa czasu obsługi klientów na stanowiskach. Dobry przegląd analitycznych metod analizy systemów komunikacyjnych zawarty jest w [REIS82]. Bardziej współczesna praca Gihra i Kuehna [GIHR85] poświęcona jest analizie wielowarstwowego systemu protokołów polegającej na wykorzystaniu hierarchicznej dekompozycji i agregacji. W pracy tej mechanizm sterowania przepływem za pomocą okna został zamodelowany w sposób podobny, jak to zrobił Reiser w [REIS81b]. Czas przesłania potwierdzenia jest tu jednak uwzględniony w sposób uproszczony za pomocą równoważnego stanowiska typu *IS*.

Wraz z postępem prac normalizacyjnych związanych z sieciami komputerowymi dokonywane są próby analizy wydajności mechanizmów sterowania przepływem uwzględniającej w istotnym stopniu istniejące rozwiązania sieciowe (np. [LIM87] – ALOHA, [CHEN88] – TOKEN-RING, [GANZ88] – TOKEN-BUS, [BOSC88] – architektura MAP). Związany z tym wzrost złożoności modelu okupiony jest zwykle uproszczeniami w określeniu wpływu rozmiaru okna i czasu przesłania potwierdzenia odbioru pakietu na parametry wydajnościowe systemu. Stosowane są tu np. modele w postaci systemu ze stratami, modele nie uwzględniające czasu przesłania potwierdzenia bądź zależności heurystyczne.

Problem analizy wydajnościowej mechanizmu sterowania przepływem za pomocą okna może być sformułowany w postaci jedno- lub wieloklasowego modelu kolejkowego typu zamkniętego lub otwartego z podsieciami o ograniczonej populacji klientów. Modele takie pierwotnie były rozwijane z myślą o wieloprogramowości. Wobec potrzeb analizy sieci komunikacyjnych nabrały one jakościowo nowego znaczenia zachowując podobne metody analizy. W modelach tych zwykle zakłada się, że jeśli nie uwzględniać mechanizmów synchronizacji, to pozostała po ich usunięciu sieć kolejkowa spełnia założenia modelu *BCMP* [BASK75] o rozwiązaniu dokładnym w postaci formy iloczynowej. Limitowanie populacji klientów jest wyrażone w sposób następujący. W przypadku jednej klasy klientów, tzn. przy założeniu, że klienci nie są rozróżnialni, całkowita liczba klientów w określonym fragmencie (podsieci) modelu kolejkowego nie może przekroczyć określonej liczby. Ewentualni klienci nadmiarowi oczekują w kolejce wejściowej do tej podsieci. W przypadku wielu klas klientów dla każdej z klas określone jest niezależne ograniczenie i przydzielona

niezależna kolejka wejściowa. Modele takie były już analizowane przez wielu autorów metodami przybliżonymi opisanymi w początkowej części tego rozdziału. Przegląd tych metod można znaleźć np. w [THOM84] i [KRZE85].

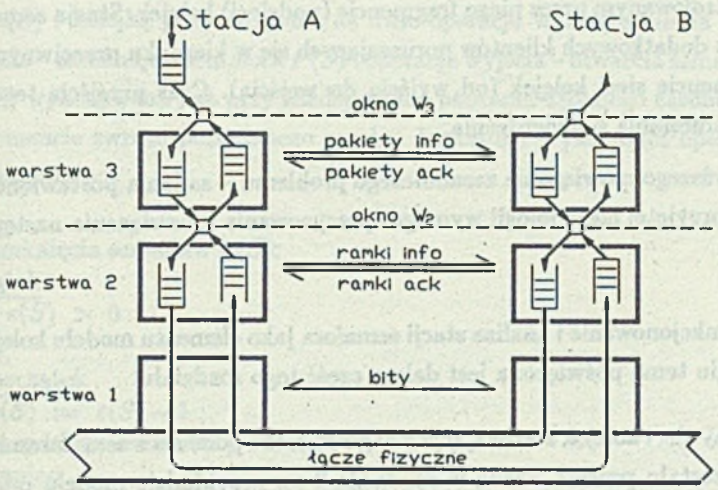
1.2. CEL I ZAKRES PRACY

Przegląd literatury ukazuje zainteresowanie, z jakim usiłuje się uwzględnić w modelach kolejkowych mechanizmy synchronizacji badanych systemów, podczas gdy są one z łatwością reprezentowane w sieciach Petri. Ponadto żadna z proponowanych metod, choć jest metodą przybliżoną, nie jest uniwersalna. Każda z nich obejmuje wąską dziedzinę zastosowań i często wykorzystuje technikę iteracji, gdzie trzeba sprawdzać kryteria zbieżności. Jednakże ze względów, o których była już mowa, metody te nadal są rozwijane, stanowiąc poważną konkurencję ekonomiczną w stosunku do symulacji i znajdując u jej boku zastosowanie przy rozwiązywaniu dużych modeli hybrydowych. Metody kolejkowe są wykorzystywane przy badaniach wydajnościowych rzeczywistych systemów informatycznych, w tym również sieci komputerowych.

W niniejszej pracy pragnę przedstawić oryginalne rozwiązania poszerzające zakres zastosowań modeli kolejkowych w dziedzinie modelowania własności systemów wykraczających poza założenia teorii *BCMP*. Chodzi tu o metody rozwiązywania sieci kolejek zawierających semafory – elementy typowe dla sieci Petri – elementy, które szczególnie nadają się do modelowania wieloprogramowości, jednoczesnego dostępu do wspólnych zasobów (np. baz danych), a zwłaszcza protokołów komunikacyjnych z tzw. sterowaniem przepływu za pomocą okna (Window flow control) [ISO82], [CCITT84]. Właśnie na przykładzie modelowania protokołu komunikacyjnego X.25 chcę pokazać zastosowanie opracowanych rozwiązań.

Inspiracją do tej pracy, jej problemem przewodnim jest próba obliczenia wydajności mechanizmu sterowania przepływem za pomocą okna stosowanego w pakietowych sieciach komunikacyjnych pracujących np. z wykorzystaniem protokołu X.25 [CCITT84]. Opracowany w 1982 roku przez ISO model odniesienia [ISO82] dekomponuje interfejs komunikacyjny na 7 poziomów–warstw protokołów. Filozofia otrzymanej w ten sposób architektury dla każdej z warstw precyzyjnie definiuje zakres usług komunikacyjnych udostępnianych warstwie wyższej oraz realizujące te usługi – protokoły. Procesy zachodzące na każdym z poziomów przebiegają w sposób asynchroniczny i tym samym tworzą kolejki transmitowanych bloków danych pomiędzy warstwami (rys.1-1). Kolejki te wynikają również z ograniczonej rozmiarem tzw. okna liczby pakietów i ramek, które można wysłać bez

uzyskania potwierdzenia ich odbioru.



Rys.1-1. Trzy pierwsze warstwy interfejsu komunikacyjnego dla X.25

Fig.1-1. First three communication layers for X.25

W modelu kolejkowym przedstawionego systemu kolejnym warstwom komunikacyjnym odpowiadają kolejne zagnieżdżone w sobie podsieci stanowisk obsługi. Populacja klientów w każdej podsieci reprezentujących przesyłane pakiety lub ramki jest ograniczona rozmiarem okna danej warstwy. Specyfiką modelu systemu komunikacyjnego jest konieczność uwzględnienia czasu przesłania potwierdzenia odbioru pakietu lub ramki. Oznacza to, że klient oczekujący na wejście do podsieci nie dostaje się do niej natychmiast w tym samym momencie, w którym inny już w niej przebywający opuści ją, lecz musi jeszcze czekać przez czas odpowiadający przesłaniu potwierdzenia tego faktu. Takie modele kolejkowe nie mają dokładnego rozwiązania w postaci formy iloczynowej [BASK75] i wymagają tym samym zastosowania przybliżonej metody rozwiązania.

Cytowane w literaturze metody lepiej lub gorzej nadają się do analizy wydajnościowej mechanizmów oknowego sterowania przepływem w sieciach komunikacyjnych. Żadna z nich nie uwzględnia jednak w sposób zadowalający czasu przesłania potwierdzenia odbioru przesyłanych siecią danych, które to potwierdzenie w postaci pakietu lub ramki wraca drogą powrotną przez sieć będąc poddane tym samym wielopoziomowym mechanizmom sterowania. W zasadzie zagadnienie to w sposób uproszczony zostało podjęte

jedynie przez Gihra i Kuehna w [GIHR85]. Twierdzę, że dobre rezultaty daje tu zastosowanie dodatkowych elementów – semaforów. Działanie semafora limituje populację klientów w kontrolowanym przez niego fragmencie (podsieci) kolejek. Stacja semafora jest otwierana przez dodatkowych klientów poruszających się w kierunku przeciwnym w limitowanym fragmencie sieci kolejek (od wyjścia do wejścia). Czas przejścia tego klienta modeluje czas przesłania potwierdzenia.

Wobec powyższego rozwiązanie zasadniczego problemu – zadania postawionego w tej pracy na bazie przyjętej metodologii wymaga sprecyzowania i rozwiązania następujących podproblemów:

1. Pojęcie, funkcjonowanie i analiza stacji semafora jako elementu modelu kolejkowego. Zagadnieniu temu poświęcona jest dalsza część tego rozdziału.

2. Idea analizy sieci kolejek zawierających zagnieżdżone podsieci z semaforami. Zagadnienie to zostało przedstawione w rozdziale 2 na przykładzie modelu odniesienia reprezentującego warstwową architekturę interfejsu komunikacyjnego. W rozdziale tym zamieszczona jest również ocena dokładności proponowanej metody.

3. Rozdział 3 zawiera uogólnienie i formalny algorytm proponowanej przybliżonej metody analizy dla różnego typu sieci kolejek: zamkniętych i otwartych ze stałym lub zależnym od liczby klientów w sieci strumieniem przybywających klientów.

4. Sposób wykorzystania metody do analizy wydajności mechanizmu sterowania przepływem za pomocą okna protokołu komunikacyjnego X.25. Ten przykład zastosowania opracowanej metody jest zaprezentowany w rozdziale 4.

1.3. POJĘCIE STACJI SEMAFORA W MODELU KOLEJKOWYM

Rozważmy system, którego wspólne zasoby mogą być przydzielane jednocześnie kilku użytkownikom. Maksymalna liczba użytkowników jednocześnie posiadających zasoby ograniczona jest jednak liczbą dostępnych zasobów systemu. Konkurowanie o przydział zasobu powoduje tym samym konieczność blokowania jednych użytkowników przez innych. Mechanizm przydziału efektywnie rozstrzygający tego typu konflikt dostępu jest zwykle realizowany za pomocą semafora. W dalszej części pracy mechanizm ten będzie nazywany zwięźle i jednoznacznie mechanizmem semaforowania.

Semafor [DIJK65] definiuje zmienna całkowita $e(S)$, której wartość początkowa W zwana rozmiarem semafora odpowiada liczbie udostępnianych zasobów, i kolejka $f(S)$. Ze zmienną $e(S)$ i kolejką $f(S)$ związane są dwie operacje wykonywane na semaforze: operacja wejścia – zamknięcia semafora $P(S)$ i operacja wyjścia – otwarcia semafora $V(S)$. Operacja $P(S)$ wykonywana jest przy każdej próbie pobrania żądanego zasobu, operacja $V(S)$ – w momencie zwrotu posiadanego zasobu. Procedury wykonujące operacje $P(S)$ i $V(S)$ są następujące:

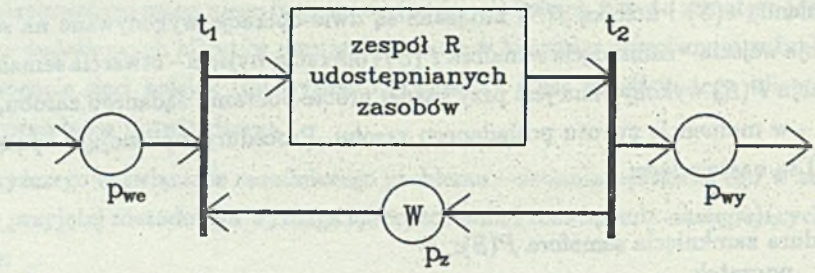
Procedura zamknięcia semafora $P(S)$:

początek
jeżeli $e(S) > 0$
wtedy
    początek
    $e(S) := e(S) - 1$;
    przydzielenie zasobu ;
    koniec
w przeciwnym razie
    wprowadzenie użytkownika do kolejki $f(S)$;
koniec

Procedura otwarcia semafora $V(S)$:

początek
$e(S) := e(S) + 1$;
jeżeli kolejka $f(S)$ niepusta wtedy
    wyprowadzenie użytkownika z kolejki $f(S)$;
    { dla wyprowadzonego użytkownika ponownie jest wykonywana
    operacja $P(S)$ }
koniec

Właściwości opisowe sieci Petri pozwalają z łatwością zamodelować mechanizm semaforowania. Otrzymany model przedstawiony jest na rysunku 1-2. W modelu tym operację zamknięcia semafora $P(S)$ wykonuje przejście t_1, natomiast operację jego otwarcia $V(S)$ - przejście t_2. Miejsca sieci p_{we}, p_{wy} i p_z reprezentują kolejno: kolejkę wejściową semafora - $f(S)$, miejsce, dokąd udają się użytkownicy po zwrocie posiadanego zasobu i liczbę dostępnych zasobów – $e(S)$.

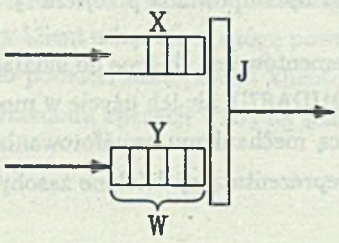


Rys.1-2. Reprezentacja mechanizmu semaforowania za pomocą sieci Petri

Fig.1-2. Semaphore mechanism representation by means of Petri Nets

Punkt reprezentujący użytkownika przybywający do miejsca p_{we} sieci może zapalić (otworzyć) przejście t_1 pod warunkiem, że miejsce p_z sieci jest niepuste. Jeśli warunek ten jest spełniony, palące się przejście konsumuje po jednym punkcie z miejsc p_{we} oraz p_z i generuje punkt, który przechodzi do fragmentu R sieci reprezentującego zespół udostępnianych zasobów. W przeciwnym wypadku przybywający punkt jest zmuszony zaczekać w miejscu p_{we}. Takie działanie odpowiada ściśle wykonaniu operacji $P(S)$ na semaforze. Liczba punktów, które mogą przedostać się do fragmentu R sieci, ograniczona jest liczbą punktów znajdujących się w miejscu p_z. Liczba ta odpowiada wartosci zmiennej $e(S)$ semafora. Przejście t_2 zapalane jest w momencie, gdy jakiś punkt opuszcza fragment R sieci. Przejście to generuje po jednym punkcie do miejsc p_{wy} oraz p_z sieci. Pierwszy z nich reprezentuje użytkownika, który zwraca posiadany zasób i kontynuuje swą dalszą drogę już bez niego. Drugi punkt zasila pulę punktów w miejscu p_z, umożliwiając ewentualne kolejne zapalenie przejścia t_1. To działanie odpowiada z kolei wykonaniu operacji $V(S)$ na semaforze.

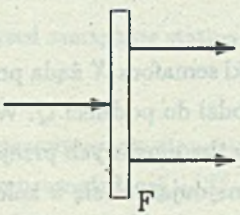
Klasyczne modele kolejkowe nie pozwalają zamodelować mechanizmu semaforowania. Aby to uczynić, należy wprowadzić do struktury modelu składającego się ze źródeł i stanowisk obsługi klientów dwa nowe obiekty. Są to: element łączący (rys.1-3) i element rozłączający (rys.1-4).



Rys.1-3. Element łączący

Fig.1-3. Join element

Z elementem łączącym J skojarzone są dwie kolejki X i Y. Jeśli obie kolejki są niepuste, to element łączący eliminuje z obu kolejek po jednym kliencie i tworzy z nich jednego klienta, który udaje się w dalszą drogę. Utworzony klient przejmuje cechy klienta z kolejki X. Wyjaśniając działanie elementu łączącego innymi słowy, można powiedzieć, że element ten przy niepustej kolejce Y przepuszcza klienta z kolejki X jednocześnie eliminując jednego klienta z kolejki Y. Działanie to powtarzane jest z zerowym czasem wykonania tak długo, jak długo jednocześnie w obu kolejkach są klienci. Jeśli w jednej z kolejek zabraknie klientów, klienci drugiej kolejki zmuszeni są do czekania do chwili pojawienia się nowych klientów w opróżnionej kolejce. W porównaniu z siecią Petri z rysunku 1-2, kolejki X, Y oraz element łączący J odpowiadają odpowiednio miejscom p_{we}, p_z oraz przejściu t_1 sieci. W dalszej części pracy zespół ten nazwany jest stacją semafora, kolejka X – kolejką semafora, a kolejka Y – kolejką licznika.

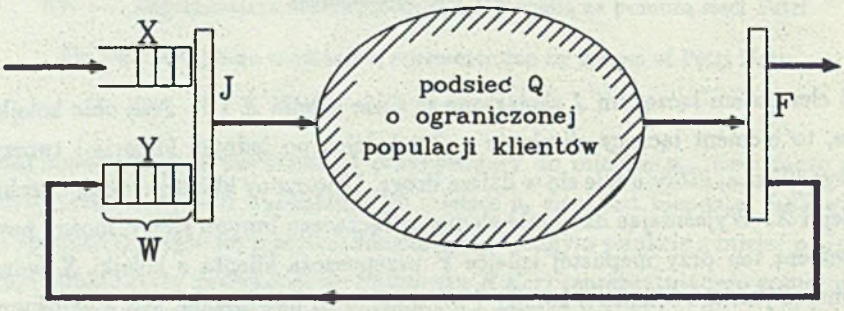


Rys.1-4. Element rozłączający

Fig.1-4. Fork element

Element rozłączający F symbolizuje generację dodatkowego klienta przez klienta przechodzącego przez ten element lub rozdwojenie się klienta. W porównaniu z siecią Petri z rysunku 1-2 element ten odpowiada przejściu t_2.

Funkcjonowanie obu elementów jest zbiżone do działania znanych z literatury elementów typu *join* i *fork* (np. [DUDA87]), ale ich użycie w modelu kolejkowym jest odmienne. Modelowanie za ich pomocą mechanizmu semaforowania ograniczającego populację klientów w podsieci kolejek reprezentującej dzielone zasoby przedstawione jest na rysunku 1-5.



Rys.1-5. Ilustracja wykorzystania semafora do limitowania populacji klientów podsieci
J — element łączący, F — element rozłączający,
X — kolejka semafora, Y — kolejka licznika, W — rozmiar semafora

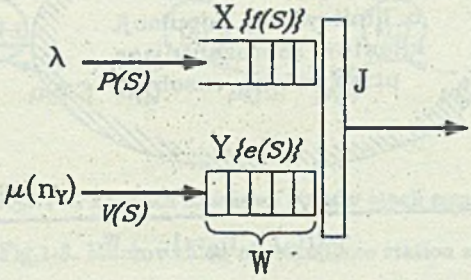
Fig.1-5. Illustration of the use of a semaphore to limit the customer population within a subnetwork
J — join element, F — fork element,
X — semaphore queue, Y — counter queue, W — semaphore size

Klient przybywający do kolejki semafora X żąda przepuszczenia. Jeśli jest to możliwe, opuszcza kolejkę semafora i wchodzi do podsieci Q. W przeciwnym wypadku jest blokowany i zmuszony do czekania. Liczba możliwych przejść przez element łączący J wyznaczona jest przez liczbę klientów znajdujących się w kolejce licznika Y, wstępnie ustawioną na W, gdzie W, zwane rozmiarem semafora, określa maksymalną liczbę klientów mogących przebywać w podsieci. Dopóki w kolejce licznika są klienci, przybywający do kolejki semafora klient jest przepuszczany natychmiast i przechodząc przez element łączący J ruguje z kolejki licznika jednego klienta, co odpowiada wykonaniu operacji $P(S)$ na semafo-

rze. W przeciwnym wypadku przybywający klient czeka w kolejce semafora do momentu, w którym w kolejce licznika pojawi się ponownie jakiś klient. Klient, który po przejściu przez podsieć Q opuszcza ją, przechodzi przez element rozłączający F, gdzie generuje dodatkowego klienta. Ten dodatkowy klient udaje się w drogę powrotną i przybywając do kolejki licznika umożliwia wejście do podsieci następnemu klientowi z kolejki semafora. Ta ostatnia czynność odpowiada wykonaniu operacji $V(S)$ na semaforze.

1.4. ANALIZA STACJI SEMAFORA

W modelu kolejkowym z rysunku 1-5 suma klientów przebywających w podsieci Q o limitowanej populacji oraz w kolejce licznika Y jest stała i równa rozmiarowi W semafora. W efekcie znajomość stanu kolejki Y jednoznacznie określa liczbę pobranych zasobów, tj. liczbę klientów przebywających w podsieci. Analiza modelu pozwalająca wyznaczyć między innymi średnią liczbę używanych zasobów może być zatem sprowadzona jedynie do analizy stacji semafora (rys.1-6).



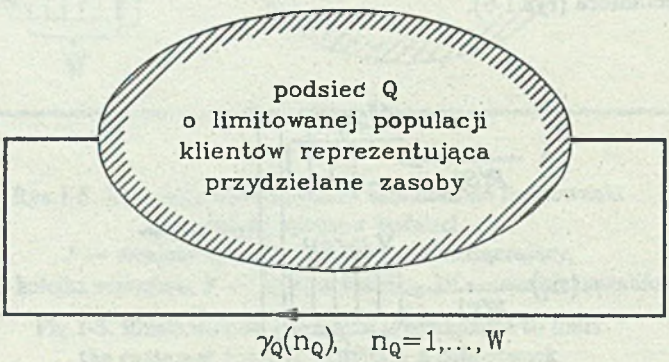
Rys.1-6. Analizowana stacja semafora

Fig.1-6. Analysed semaphore station

Stacja semafora jest całkowicie zdeterminowana stanem kolejki X reprezentującej $f(S)$ oraz stanem kolejki licznika Y reprezentującej $e(S)$. Stan kolejki licznika jest inicjowany liczbą W klientów odpowiadającą liczbie istniejących zasobów. Strumień przybywających klientów żądających przejścia, tj. wykonania operacji $P(S)$ na semaforze, jest Poissonowski z parametrem λ. Żądania przybywających klientów kontrolowane są przez element łączący J, który, jeśli spełnione są warunki, przepuszcza je z zerowym

czasem obsługi. Rozkład odstępów czasu między wykonaniami operacji $V(S)$ zwolnienia zasobu, tj. przybyciami klientów do kolejki licznika Y, jest nieznany, gdyż zależy od stanu podsieci Q kontrolowanej przez stację semafora. Szeregowanie klientów w kolejce semafora X odbywa się według regulaminu naturalnego *FIFO*. Znajomość sposobu szeregowania klientów w kolejce licznika Y nie jest konieczna, gdyż zakłada się, że przydzielane zasoby nie są rozróżnialne.

Do kolejki licznika Y przybywają klienci wygenerowani przez element rozłączający F w chwilach przejścia przez niego klientów opuszczających podsieć Q (rys.1-5). Gdy podsieć Q zostanie opróżniona z klientów, ustanie również strumień klientów przybywających do kolejki Y, pozostawiając ją w stanie pełnym z W klientami. Z drugiej strony maksymalna liczba klientów mogących przebywać w podsieci Q nie może przekroczyć wartości W rozmiaru semafora. Wynika stąd, że nieznany parametr, jakim jest strumień $\mu(n_Y)$ klientów przybywających do kolejki licznika Y, może być otrzymany w funkcji n_Q klientów w wyniku analizy zamkniętej podsieci kolejek Q z liczbą n_Q klientów zmieniającą się od 1 do W (rys.1-7).



Rys.1-7. Analiza strumienia zwrotów zasobów

Fig.1-7. Analysis of resource release stream

Niech $\mathcal{R}_Q(n_Q)$ będzie obliczonym średnim czasem reakcji podsieci Q zawierającej n_Q klientów. Stumień $\gamma_Q(n_Q)$ klientów przechodzących przez podsieć, równy przepustowości podsieci Q, wyraża zależność:

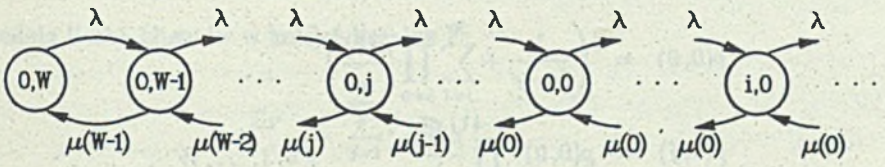
$$\gamma_Q(n_Q) = \frac{n_Q}{\mathcal{R}_Q(n_Q)} \quad . \tag{1-1}$$

Jeśli w podsieci Q przebywa n_Q klientów, to w kolejce licznika Y znajduje się $n_Y = W - n_Q$ klientów. Stąd nieznany parametr, jakim jest strumień $\mu(n_Y)$ klientów przybywających do kolejki Y, otrzymuje się w sposób następujący:

$$\mu(n_Y) = \gamma_Q(W-n_Y) = \frac{W - n_Y}{\mathcal{R}_Q(W-n_Y)} \tag{1-2}$$

Celem uproszczenia analizy, w dalszej części pracy zakłada się, że strumień $\mu(n_Y)$ ma charakter Poissonowski. Należy jednak zauważyć, że założenie to nie zawsze jest prawdziwe i zależy od rodzaju podsieci Q.

Stan stacji semafora jest scharakteryzowany przez parę (i,j), gdzie i reprezentuje liczbę klientów w kolejce X, a j – w kolejce Y. Możliwe stany stacji wraz z przejściami między nimi są przedstawione na rysunku 1-8.



Rys.1-8. Łańcuch Markowa stanów stacji semafora

Fig.1-8. Markov chain of semaphore station states

Proces liczby klientów w stacji semafora posiada stan stacjonarny, jeśli ich liczba nie rośnie wraz z czasem do nieskończoności i w granicy (przy $t \to \infty$) istnieją niezerowe prawdopodobieństwa każdego z osiągalnych stanów stacji. Kryterium istnienia stanu stacjonarnego dla stacji semafora spełniającej założenia teorii Lavenberga [LAVE78] wyraża zależność: $\lambda < \mu(0)$. Strumień klientów $\mu(0)$ równy jest maksymalnemu strumieniowi $\gamma_Q(W)$ klientów przechodzących przez podsieć Q, jaki może istnieć przy stale niepustej kolejce semafora X.

Z przedstawionych na rysunku 1-8 przejść pomiędzy stanami stacji semafora wynikają następujące równania równowagi globalnej:

$$\begin{aligned}
\lambda \cdot p(0,W) &= p(0,W-1) \cdot \mu(W-1) \ , \\
[\mu(j)+\lambda] \cdot p(0,j) &= p(0,j+1) \cdot \lambda + p(0,j-1) \cdot \mu(j-1) \qquad j = W-1,\ldots,1 \ , \\
[\mu(0)+\lambda] \cdot p(0,0) &= p(1,0) \cdot \mu(0) + p(0,1) \cdot \lambda \ , \\
[\mu(0)+\lambda] \cdot p(i,0) &= p(i-1,0) \cdot \lambda + p(i+1,0) \cdot \mu(0) \qquad i = 1,\ldots,\infty \ , \qquad (1\text{-}3)
\end{aligned}$$

gdzie: $p(i,j)$ jest prawdopodobieństwem stanu stacjonarnego znalezienia i klientów w kolejce semafora X oraz j klientów w kolejce licznika Y.

Proste sumowanie równań równowagi globalnej pokazuje, że spełnione są również równania równowagi lokalnej, które można zapisać w sposób następujący:

$$\begin{aligned}
\lambda \cdot p(0,j) &= p(0,j-1) \cdot \mu(j-1) \qquad j = W,\ldots,1 \ , \\
\lambda \cdot p(i,0) &= p(i+1,0) \cdot \mu(0) \qquad i = 0,\ldots,\infty \ . \qquad (1\text{-}4)
\end{aligned}$$

Niech $\rho = \lambda/\mu(0)$. Jeśli $\rho < 1$, tj. spełnione jest kryterium istnienia stanu stacjonarnego, to biorąc pod uwagę, że suma wszystkich prawdopodobieństw $p(i,j)$ równa jest jedności, rozwiązanie równań równowagi lokalnej daje następujące wyrażenia na wartość prawdopodobieństw poszczególnych stanów stacji semafora:

$$p(0,0) = \left(\frac{1}{1-\rho} + \sum_{j=1}^{W} \prod_{k=0}^{j-1} \frac{\mu(k)}{\lambda} \right)^{-1} \ , \qquad (1\text{-}5)$$

$$p(0,j) = p(0,0) \cdot \prod_{k=0}^{j-1} \frac{\mu(k)}{\lambda} \qquad j = 1,\ldots,W \ , \qquad (1\text{-}6)$$

$$p(i,0) = p(0,0) \cdot \rho^i \qquad i = 1,\ldots,\infty \ . \qquad (1\text{-}7)$$

Powyższe wyrażenia pozwalają otrzymać prawdopodobieństwa graniczne dla każdej z kolejek stacji semafora z osobna:

— kolejka semafora X:

$$p_X(0) = \frac{1-\rho(1+p(0,0))}{1-\rho} \ , \qquad (1\text{-}8)$$

$$p_X(i) = p(0.0) \cdot \rho^i \qquad i = 1,\ldots,\infty \ , \qquad (1\text{-}9)$$

— kolejka licznika Y:

$$p_Y(0) = p(0,0) \cdot \frac{1}{1-\rho} \ , \qquad (1\text{-}10)$$

$$p_Y(j) = p(0,0) \cdot \prod_{k=0}^{j-1} \frac{\mu(k)}{\lambda} \qquad j = 1,\ldots,W \ . \qquad (1\text{-}11)$$

Ostatecznie poszukiwane miary wydajności dla stacji semafora i kontrolowanej przez nią podsieci Q o ograniczonej populacji klientów wyrażają zależności:

— średnia liczba klientów w kolejse semafora X:

$$\begin{aligned} \mathcal{L}_X &= \sum_{i=1}^{\infty} i \cdot p_X(i) \\ &= \frac{\rho}{(1-\rho)^2} \cdot p(0,0) \ ; \end{aligned} \tag{1-12}$$

— średni czas oczekiwania w kolejce X:

$$\mathcal{R}_X = \frac{\mathcal{L}_X}{\lambda} \ ; \tag{1-13}$$

— wykorzystanie kolejki X:

$$\begin{aligned} \mathcal{U}_X &= 1 - p_X(0) \\ &= \frac{\rho}{1-\rho} \cdot p(0,0) \ ; \end{aligned} \tag{1-14}$$

— średnia liczba klientów w kolejce licznika Y:

$$\begin{aligned} \mathcal{L}_Y &= \sum_{j=1}^{W} j \cdot p_Y(j) \\ &= p(0,0) \cdot \sum_{j=1}^{W} j \cdot \prod_{k=0}^{j-1} \frac{\mu(k)}{\lambda} \ ; \end{aligned} \tag{1-15}$$

— średnia liczba klientów w podsieci Q:

$$\mathcal{L}_Q = W - \mathcal{L}_Y \ ; \tag{1-16}$$

— średni czas reakcji podsieci Q:

$$\mathcal{R}_Q = \frac{\mathcal{L}_Q}{\lambda} \ ; \tag{1-17}$$

— wykorzystanie podsieci Q:

$$\begin{aligned} \mathcal{U}_Q &= 1 - p_Y(W) \\ &= 1 - p(0,0) \cdot \prod_{k=0}^{W-1} \frac{\mu(k)}{\lambda} \ . \end{aligned} \tag{1-18}$$

Identyczne miary wydajności otrzymuje się analizując stanowisko obsługi S, którego strumień obsługi (liczba klientów, których to stanowisko może obsłużyć na jednostkę czasu) $\mu_S(n)$ zależny od populacji n klientów na stanowisku wyraża zależność:

$$\mu_S(n) = \begin{cases} \mu(W-n) = \gamma_Q(n) & \text{dla } n < W \ , \\ \mu(0) = \gamma_Q(W) & \text{dla } n \geq W \ . \end{cases} \tag{1-19}$$

Z prawa Little'a [LITT61] czas reakcji $\mathcal{R}_S(n)$ tego stanowiska w funkcji n klientów obecnych na tym stanowisku wynosi:

$$\mathcal{R}_S(n) = \begin{cases} \mathcal{R}_Q(n) & \text{dla } n < W \ , \\ \mathcal{R}_Q(W) + \frac{n-W}{\gamma_Q(W)} & \text{dla } n \geq W \ , \end{cases} \tag{1-20}$$

gdzie: $\mathcal{R}_Q(n)$ jest czasem przejścia klientów przez zamkniętą, jak na rysunku 1-7, podsieć Q z n klientami. Taki czas reakcji posiada stanowisko obsługi z W kanałami obsługi i czasem obsługi równym $\mathcal{R}_Q(k)$, $k = 1, \ldots, W$, zależnym od liczby k klientów zajmujących kanały obsługi. Dodatkowy składnik we wzorze (1-20) dla $n \geq W$ wyraża czas oczekiwania klientów w kolejce do kanałów obsługi tego stanowiska i jest tożsamy z czasem oczekiwania w kolejce semafora X.

Przeprowadzone rozważania sugerują, iż podczas analizy sieci kolejek zawierającej stację semafora z podsiecią o limitowanej populacji klientów można dokonać uproszczenia, zastępując je stanowiskiem ze strumieniem obsługi $\mu_S(n)$. Otrzymane w wyniku analizy sieci miary wydajności dla tego stanowiska służą następnie do wyznaczenia miar wydajności stacji semafora i kontrolowanej przez nią podsieci Q.

Rozdział 2

MODEL KOLEJKOWY Z SEMAFORAMI

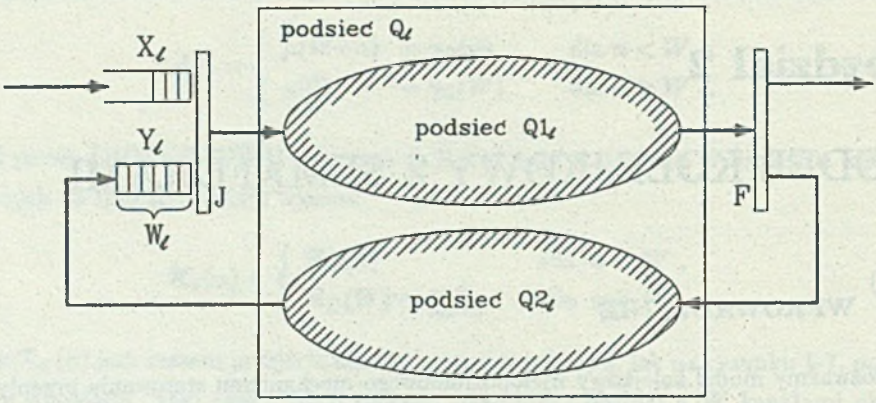
2.1. WPROWADZENIE

Rozważmy model kolejkowy wielopoziomowego mechanizmu sterowania przepływem za pomocą okna jednostek danych (np.pakietów lub ramek) w sieciach komunikacyjnych. Wielopoziomowość polega tu na tym, że fragment sieci kontrolowany jednym mechanizmem sterowania przepływem zawarty jest w innym fragmencie sieci kontrolowanym przez inny mechanizm sterowania przepływem (tzw. mechanizm wyższego poziomu), który z kolei może być częścią jeszcze większej partii sieci itd. W modelu kolejkowym tego systemu kolejnym poziomom odpowiadają kolejne zagnieżdżone w sobie podsieci stanowisk obsługi. Populacja klientów w każdej podsieci reprezentujących przesyłane jednostki danych jest ograniczona rozmiarem okna danego poziomu.

Limitowanie populacji klientów w podsieci kolejek łatwo modeluje się za pomocą mechanizmu semaforowania opisanego w poprzednim rozdziale. Zbudowany na jego podstawie model kolejkowy l-tego poziomu sterowania przepływem za pomocą okna obejmuje stację semafora z kolejką semafora X_l i kolejką licznika Y_l oraz kontrolowaną przez stację semafora podsieć Q_l stanowisk obsługi składającą się z dwóch części: podsieci $Q1_l$ i podsieci $Q2_l$ (rys.2-1).

Liczba możliwych przejść klientów przez kolejkę semafora X_l do podsieci $Q1_l$ wyznaczona jest przez liczbę dodatkowych klientów znajdujących się w kolejce licznika Y_l, wstępnie ustawioną na W_l. Wartość W_l równa jest rozmiarowi okna poziomu l badanego systemu. Klient po przejściu przez podsieć $Q1_l$ opuszcza ja (w sieci komunikacyjnej odpowiada to dotarciu np. pakietu do odbiorcy) i przechodzi przez element rozłączający F, gdzie generuje dodatkowego klienta. Ten dodatkowy klient udaje się w drogę powrotną przez podsieć $Q2_l$ (w sieci komunikacyjnej odpowiada to przesłaniu potwierdzenia odbioru pakietu) i w końcu przybywa do kolejki licznika Y_l umożliwiając wejście do podsieci $Q1_l$

następnemu klientowi z kolejki semafora. Opisany mechanizm semaforowania powoduje, że łączna liczba klientów głównych z podsieci $Q1_l$ i dodatkowych z podsieci $Q2_l$ nie przekroczy rozmiaru okna W_l.



Rys.2-1. Model kolejkowy l-tego poziomu sterowania przepływem za pomocą okna

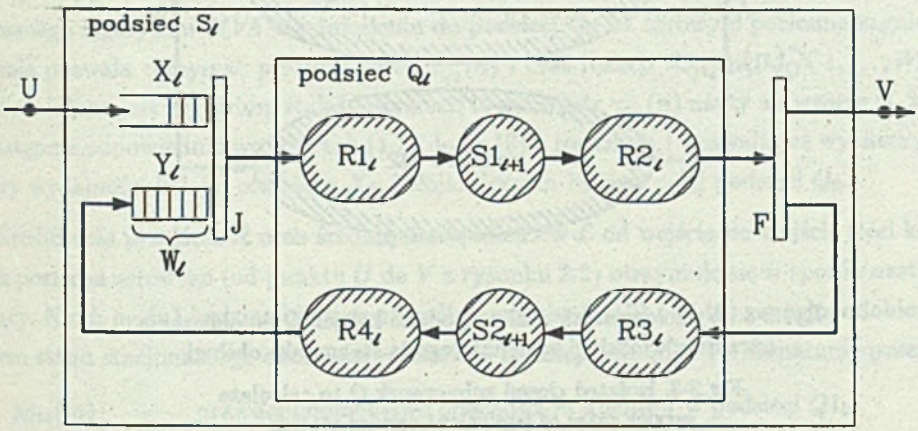
Fig.2-1. Queueing model of window flow control mechanism at layer l

W przedstawionym modelu klienci dodatkowi wracają do kolejki licznika Y_l przez oddzielną podsieć $Q2_l$. Rozbicie podsieci kolejek Q_l o limitowanej populacji klientów na dwie odseparowane od siebie podsieci $Q1_l$ i $Q2_l$ nie jest jednakże konieczne i zostało dokonane jedynie dla klarowności prezentowanego opisu i analizy. W przypadku ogólnym klienci dodatkowi wracają tą samą podsiecią, którą poruszają się klienci główni. Z powodu przeciwnego kierunku marszruty klientów dodatkowych obowiązują jednakże inne prawdopodobieństwa przejść pomiędzy stanowiskami podsieci, jak również mogą oni mieć inne czasy obsługi na tych stanowiskach.

Hierarchia poziomów mechanizmów sterowania przepływem za pomocą okna oznacza, że zaprezentowany na rysunku 2-1 model kolejkowy poziomu l jest elementem podsieci $Q1_{l-1}$ i $Q2_{l-1}$ modelu kolejkowego poziomu $l-1$. Jego własne podsieci $Q1_l$ i $Q2_l$ mogą z kolei zawierać analogiczne modele kolejkowe poziomu $l+1$ badanego systemu. Dla dalszych rozważań przyjmijmy, że istnieje L poziomów zagnieżdżenia modeli. Poziom zerowy przyjęto tu dla modelu, który nie jest zagnieżdżony w żadnym innym, tzn., nie jest fragmentem podsieci $Q1$ lub $Q2$ innego modelu, natomiast ostatni poziom zagnieżdżenia L – dla modelu, który nie posiada już zagnieżdżonych w sobie dalszych modeli. Dalsza część tego rozdziału prezentuje ideę analizy takich modeli.

2.2. ANALIZA MODELU

Przyjmijmy, że struktura modelu kolejkowego poziomu l, $l = 0, \ldots, L-1$, mechanizmu sterowania przepływem ma postać jak na rysunku 2-2. W odniesieniu do rysunku 2-1, podsieć kolejek $Q1_l$ złożona jest tu z podsieci $S1_{l+1}$ stanowiącej zagnieżdżony w $Q1_l$ model kolejkowy poziomu $l+1$ oraz dwóch klasycznych podsieci stanowisk obsługi $R1_l$ i $R2_l$ nie zawierających już dalszych zagnieżdżonych w sobie modeli. Analogicznie zbudowana jest podsieć kolejek $Q2_l$, która obejmuje podsieci $R3_l$, $R4_l$ oraz $S2_{l+1}$. Model kolejkowy o tej strukturze jest z kolei elementem struktury modelu z poziomu $l-1$, stanowiąc w nim podsieć $S1_l$ lub $S2_l$. W przypadku modelu kolejkowego ostatniego poziomu zagnieżdżenia L (podsieci S_L tj. $S1_L$ lub $S2_L$) nie ma już potrzeby rozbijania jego podsieci $Q1_L$ i $Q2_L$, gdyż nie zawierają one podsieci S.

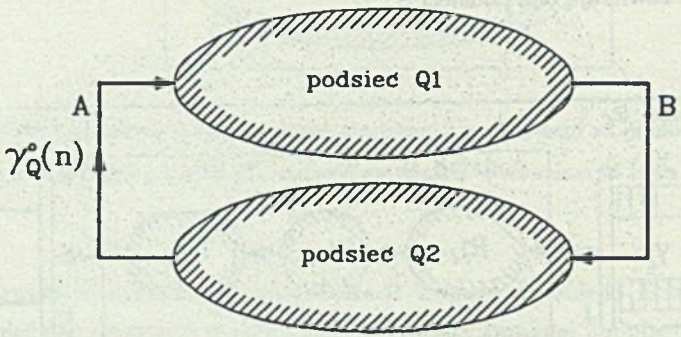


Rys.2-2. Podsieć kolejek S_l na poziomie zagnieżdżenia l, $l = 0, \ldots, L-1$, składająca się ze stacji semafora i kontrolowanej podsieci Q_l

Fig.2-2. Subnetwork S_l at nesting level l, $l = 0, \ldots, L-1$, composed of semaphore station and population constrained subnetwork Q_l

Stanowiska obsługi analizowanego modelu są typu *FCFS*, *LCFS-PR*, *PS* lub *IS*, z czasem obsługi o rozkładzie wykładniczym. Oznacza to, że jeśli nie uwzględniać stacji semaforów, tzn. limitowania populacji klientów w kontrolowanych podsieciach, to analizowany model kolejkowy odpowiada kryteriom modelu *BCMP* posiadającego rozwiązanie dokładne w postaci formy iloczynowej [BASK75].

Rozważmy podsieć kolejek S_L złożoną ze stacji semafora i kontrolowanych przez nią podsieci $Q1_L$ oraz $Q2_L$ na poziomie zagnieżdżenia L. Na tym poziomie zagnieżdżenia podsieci $Q1_L$ i $Q2_L$ nie zawierają innych podsieci o limitowanej liczbie klientów i spełniają założenia modelu *BCMP*. W myśl twierdzenia Nortona w odniesieniu do modeli kolejkowych (por. [CHAN75a]) oraz wniosków z analizy stacji semafora (rozdział 1), podsieć S_L jest zastępowana równoważnym stanowiskiem obsługi o przepustowości, tj. strumieniu obsługi $\mu^{\circ}_{S_L}(n)$ równym szybkości zwalniania semafora oraz czasie reakcji $\mathcal{R}^{\circ}_{S_L}(n)$ równym czasowi przejścia klienta od wejścia do wyjścia podsieci S_L. Wielkości $\mu^{\circ}_{S_L}(n)$ oraz $\mathcal{R}^{\circ}_{S_L}(n)$ otrzymuje się w wyniku analizy wyizolowanej i zamkniętej jak na rysunku 2-3 podsieci kolejek Q_L z n klientami, $n = 1, \ldots, W_L$, gdzie W_L jest rozmiarem okna–semafora poziomu L.



Rys.2-3. Wyizolowana zamknięta podsieć kolejek Q do obliczania parametrów obsługi zastępującego ją stanowiska obsługi

Fig.2-3. Isolated closed subnetwork Q to calculate service rates of equivalent service station

Posługując się znanym algorytmem *MVA* [REIS80] obliczamy dla każdej populacji n, $n = 1, \ldots, W_L$, klientów w zamkniętej sieci kolejek Q_L następujące parametry:

$\gamma^{\circ}_{Q_L}(n)$ — przepustowość podsieci Q_L (w punkcie A z rysunku 2-3),

$\mathcal{R}^{\circ}_{Q1_L}(n)$ — czas reakcji podsieci $Q1_L$ (od punktu A do punktu B z rysunku 2-3).

Ostatecznie podsieć kolejek S_L jest zastępowana stanowiskiem obsługi o strumieniu obsługi:

$$\mu^{\circ}_{S_L}(n) = \begin{cases} \gamma^{\circ}_{Q_L}(n) & \text{dla } n < W_L \ , \\ \gamma^{\circ}_{Q_L}(W_L) & \text{dla } n \geq W_L \ , \end{cases} \tag{2-1}$$

oraz czasie reakcji:

$$\mathcal{R}^{\circ}_{S_L}(n) = \begin{cases} \mathcal{R}^{\circ}_{Q1_L}(n) & \text{dla } n < W_L \,, \\ \mathcal{R}^{\circ}_{Q1_L}(W_L) + \frac{n-W_L}{\gamma^{\circ}_{Q_L}(W_L)} & \text{dla } n \geq W_L \,. \end{cases} \tag{2-2}$$

Dodatkowy składnik we wzorze (2-2) dla $n \geq W_L$ jest czasem oczekiwania klientów w kolejce semafora i pochodzi z reguły Little'a [LITT61].

Opisanej metody używa się do zastąpienia obu podsieci $S1_L$ i $S2_L$ na poziomie zagnieżdżenia L równoważnymi stanowiskami obsługi. Agregacje te doprowadzają do otrzymania sieci kolejek z $L-1$ poziomami zagnieżdżenia podsieci i, na tym samym poziomie $L-1$, stanowiskami obsługi o przepustowości i czasie obsługi zależnymi od populacji klientów na stanowisku. Stosując opracowany w tej pracy (rozdział 3) zmodyfikowany algorytm *MVA*, dostosowany do rozwiązywania zamkniętych sieci kolejek zawierających stanowiska o zależnym czasie obsługi, w podobny sposób można dokonać agregacji podsieci na kolejno zmniejszających się poziomach zagnieżdżenia $L-1, \ldots, 1$. Ostatnie zastosowanie zmodyfikowanego algorytmu *MVA* w odniesieniu do podsieci Q_0 na zerowym poziomie zagnieżdżenia pozwala otrzymać przepustowość $\gamma^{\circ}_{Q_0}(n)$ i czas reakcji $\mathcal{R}^{\circ}_{Q1_0}(n)$, $n = 1, \ldots, W_0$. Jeśli spełnione jest kryterium stabilności sieci, to parametr $\gamma^{\circ}_{Q_0}(n)$ użyty we wzorze (1-2), a następnie odpowiednio we wzorach (1-8) do (1-18) z rozdziału 1 pozwala już wyznaczyć miary wydajności kolejki semafora X_0, kolejki licznika Y_0 oraz całej podsieci Q_0.

Średni czas przejścia $\mathcal{R}$ oraz średnią liczbę klientów $\mathcal{L}$ od wejścia do wyjścia sieci kolejek poziomu zerowego (od punktu U do V z rysunku 2-2) otrzymuje się w sposób następujący. Niech $p_{Y_0}(n)$ będzie obliczonym według wzorów (1-10) i (1-11) prawdopodobieństwem stanu stacjonarnego obecności n klientów w kolejce licznika Y_0. Oznaczmy przez:

- $p_{Q1_0}(m)$ — prawdopodobieństwo obecności m klientów w podsieci $Q1_0$,
- $p^{\circ}_{Q1_0}(m|n)$ — prawdopodobieństwo obecności m klientów w podsieci $Q1_0$ liczone w wyizolowanej zamkniętej podsieci Q_0 z n klientami,
- $p_{Q_0}(n)$ — prawdopodobieństwo obecności n klientów w podsieci Q_0. Ponieważ łączna liczba klientów podsieci Q_0 i kolejki licznika Y_0 jest stała, równa rozmiarowi W_0 semafora, to $p_{Q_0}(n) = p_{Y_0}(W_0 - n)$.

Dla $m > n$ prawdopodobieństwo $p^{\circ}_{Q1_0}(m|n) = 0$. Uwzględniając, że [SAUE81]:

$$p_{Q1_0}(m) = \sum_{n=m}^{W_0} p^{\circ}_{Q1_0}(m|n) \cdot p_{Q_0}(n) \qquad m = 0, \ldots, W_0 \,, \tag{2-3}$$

otrzymujemy następujące wyrażenie na średnią liczbę klientów w podsieci $Q1_0$:

$$\begin{aligned}\mathcal{L}_{Q1_0} &= \sum_{m=1}^{W_0} m \cdot p_{Q1_0}(m) \\ &= \sum_{m=1}^{W_0} m \cdot \sum_{n=1}^{W_0} p^{\circ}_{Q1_0}(m|n) \cdot p_{Y_0}(W_0 - n) \\ &= \sum_{n=1}^{W_0} p_{Y_0}(W_0 - n) \cdot \sum_{m=1}^{W_0} m \cdot p^{\circ}_{Q1_0}(m|n) \; . \end{aligned} \tag{2-4}$$

Suma $\sum_{m=1}^{W_0} m \cdot p^{\circ}_{Q1_0}(m|n)$ jest średnią liczbą klientów w podsieci $Q1_0$ wyznaczoną podczas analizy podsieci Q_0 w izolacji z n klientami. Wartość średniej liczby klientów w podsieci $Q1_0$ jako elementu modelu otwartego ze stacją semafora wynosi zatem:

$$\mathcal{L}_{Q1_0} = \sum_{n=1}^{W_0} \mathcal{L}^{\circ}_{Q1_0}(n) \cdot p_{Y_0}(W_0 - n) = \sum_{n=1}^{W_0} \gamma^{\circ}_{Q_0}(n) \cdot \mathcal{R}^{\circ}_{Q1_0}(n) \cdot p_{Y_0}(W_0 - n) \; . \tag{2-5}$$

Otrzymane wyrażenie pozwala obliczyć poszukiwane globalne miary wydajności: średnią liczbę klientów $\mathcal{L}$ oraz średni czas przejścia klientów $\mathcal{R}$ od wejścia do wyjścia sieci kolejek poziomu zerowego ze wzorów:

$$\mathcal{L} = \mathcal{L}_{X_0} + \mathcal{L}_{Q1_0} \; , \tag{2-6}$$

$$\mathcal{R} = \frac{\mathcal{L}}{\lambda} \; , \tag{2-7}$$

z $\mathcal{L}_{X_0}$ otrzymanym ze wzoru (1-12).

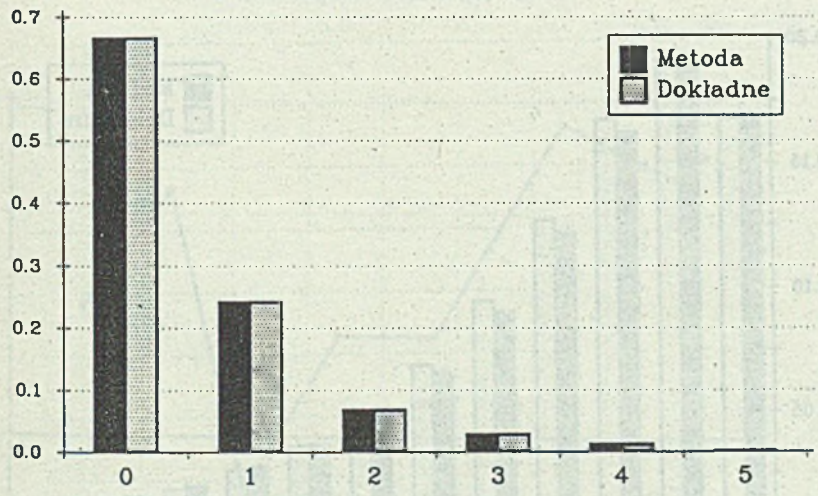
2.3. OCENA DOKŁADNOŚCI METODY ANALIZY

Opisana metoda analizy zagnieżdżonych podsieci kolejek z semaforami została zweryfikowana poprzez symulację oraz w jednym przypadku przez porównanie jej wyników ze znanymi wynikami dokładnymi zamieszczonymi w pracy [MAIL87]. W szczególności przedmiotem analizy był model jak na rysunku 2-2 z jednym, dwoma i trzema poziomami zagnieżdżenia. W dalszej części przedstawione są warunki oraz wyniki przeprowadzonych analiz. Wyniki porównawczej analizy symulacyjnej zostały wykreślone z przedziałami ufności na poziomie ufności 0.95.

A. Model jednopoziomowy ($L=0$)

W tym przypadku podsieci kolejek $Q1$ oraz $Q2$ kontrolowane przez stację semafora nie zawierają podsieci z semaforami $S1$ i odpowiednio $S2$. Dla analizy założono, że podsieci $R1$ oraz $R2$ będą reprezentowane przez pojedyncze stanowiska obsługi typu *FCFS*. Tym samym podsieć $Q1$ stanowi tandem dwóch stanowisk obsługi. Niech $\mu 1_0$ oraz $\mu 2_0$ będą strumieniami obsługi pierwszego i drugiego stanowiska. Podsieć $Q2$ została pominięta. Na rysunkach 2-4, 2-5 i 2-6 przedstawione są rozkłady łącznej liczby klientów w modelu tj. w kolejce semafora X i podsieci Q dla trzech różnych wartości strumienia λ przybywających klientów. Otrzymane prawdopodobieństwa porównane są ze znanymi wynikami dokładnymi otrzymanymi z numerycznego rozwiązania układu równań równowagi globalnej tego modelu i zamieszczonymi w pracy [MAIL87].

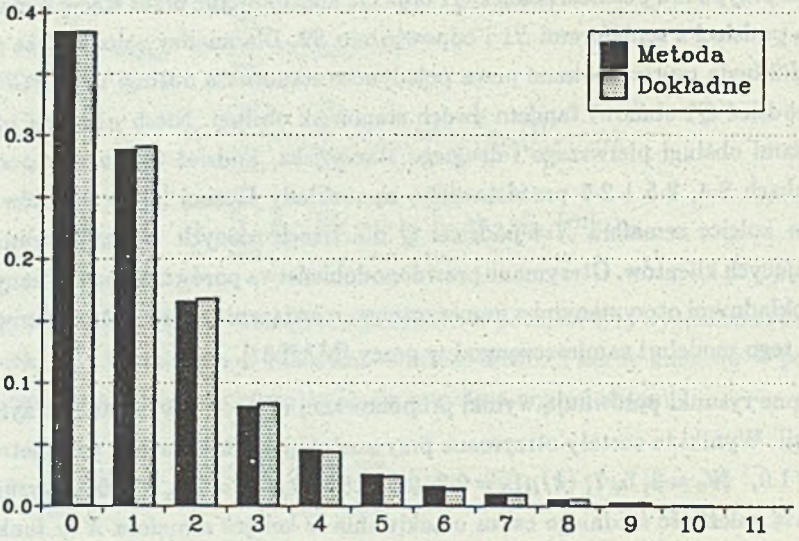
Następne rysunki porównują wyniki proponowanej metody z wynikami otrzymanymi z symulacji. Wyniki te zostały otrzymane przy następujących zmianach parametrów modelu: $\lambda = 1.0$, $W_0 = 3, 5, 7$, $1/\mu 1_0 = 0.2, 0.5, 0.8$ i $1/\mu 2_0 = 0.2, 0.5, 0.8$. Rysunek 2-7 przedstawia zależność średniego czasu oczekiwania w kolejce semafora X w funkcji wykorzystania U_X tej kolejki. Względny błąd pomiędzy wynikami metody i symulacyjnymi z rysunku 2-7 wykreślony jest na rysunku 2-8.



Rys.2-4. Rozkład liczby klientów w modelu

Fig.2-4. Distribution of the number of customers in the model

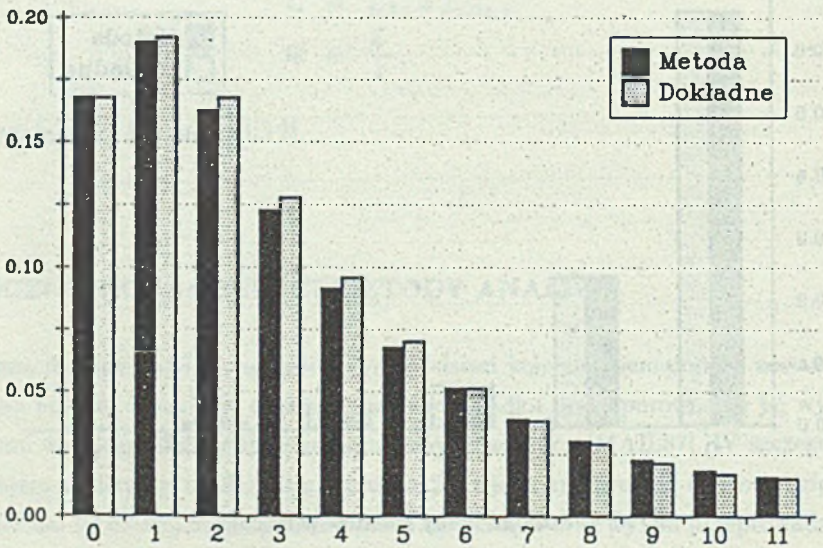
$L=0$, $\lambda=0.375$, $W_0=3$, $\mu 1_0=\mu 2_0=2$



Rys.2-5. Rozkład liczby klientów w modelu

Fig.2-5. Distribution of the number of customers in the model

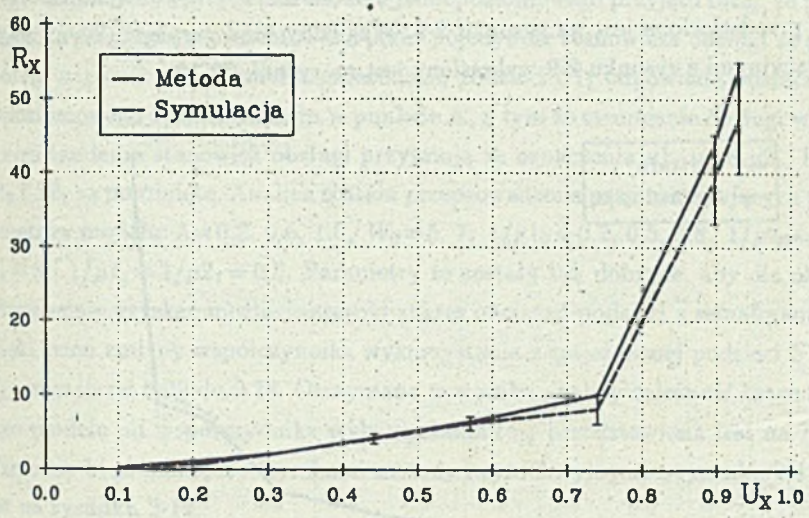
$L=0,\ \lambda=0.750,\ W_0=3,\ \mu 1_0=\mu 2_0=2$



Rys.2-6. Rozkład liczby klientów w modelu

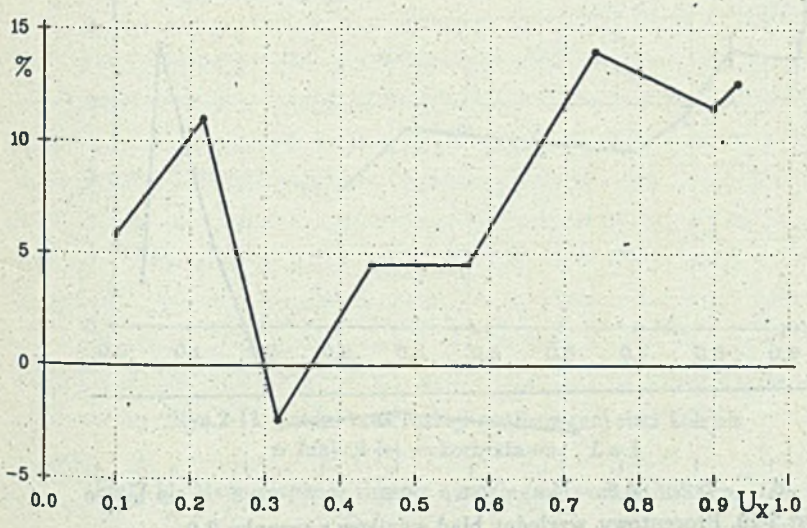
Fig.2-6. Distribution of the number of customers in the model

$L=0,\ \lambda=1.125,\ W_0=3,\ \mu 1_0=\mu 2_0=2$



Rys.2-7. Średni czas oczekiwania w kolejce semafora X w funkcji wykorzystania tej kolejki, $L=0$

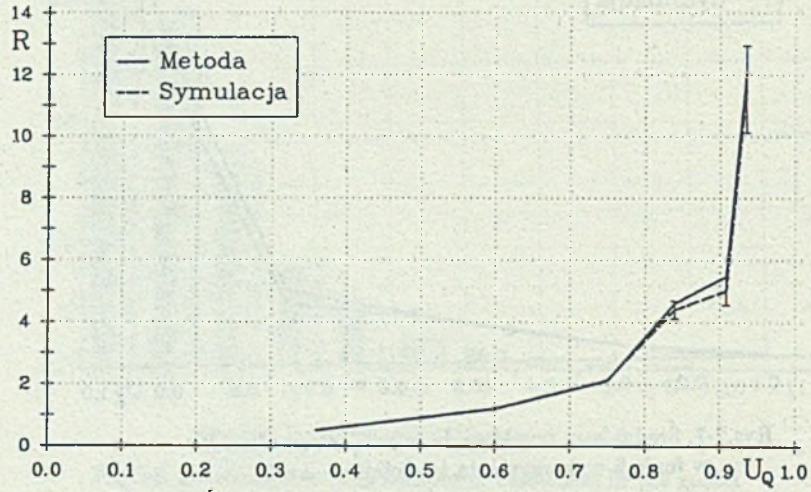
Fig.2-7. Mean waiting time in semaphore queue X vs semaphore queue utilisation, $L=0$



Rys.2-8. Procentowy, względny błąd wyników z rysunku 2-7

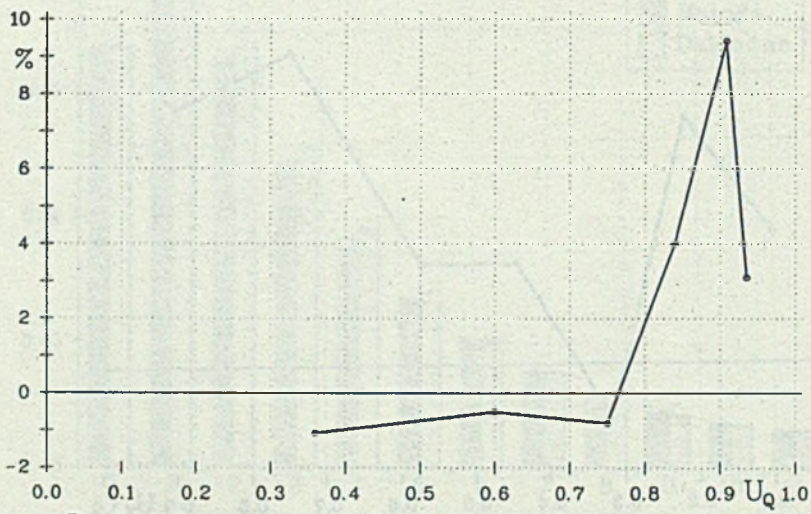
Fig.2-8. % relative error for the results given in figure 2-7

Dokładność proponowanej metody zależy od stopnia wykorzystania zespołu stacji semafora i kontrolowanej przez niego podsieci Q wyrażonym współczynnikiem U_Q (patrz wzór (1-18)). Zależność czasu reakcji analizowanej sieci kolejek od współczynnika wykorzystania U_Q przedstawiona jest na rysunku 2-9. Względny błąd pomiędzy wynikami metody i symulacyjnymi z rysunku 2-9 wykreślony jest na rysunku 2-10.



Rys.2-9. Średni czas reakcji analizowanej sieci kolejek w funkcji jej wykorzystania, $L=0$

Fig.2-9. Mean response time vs queueing network utilisation, $L=0$

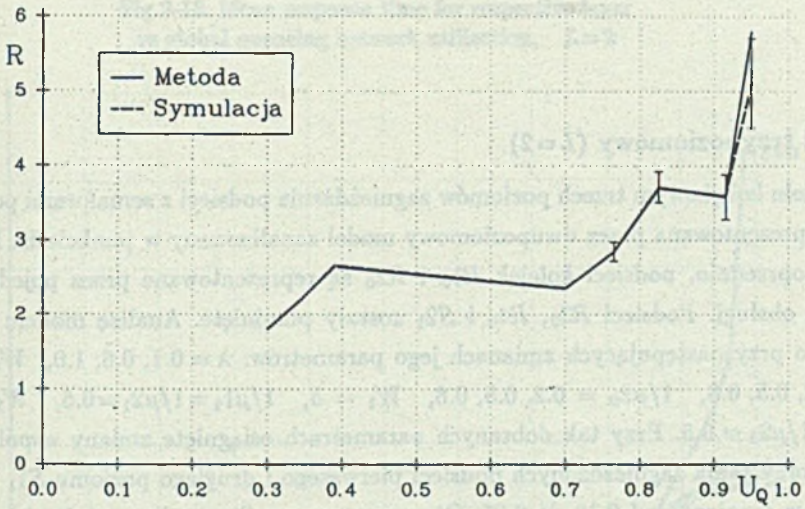


Rys.2-10. Procentowy, względny błąd wyników z rysunku 2-9

Fig.2-10. % relative error for the results given in figure 2-9

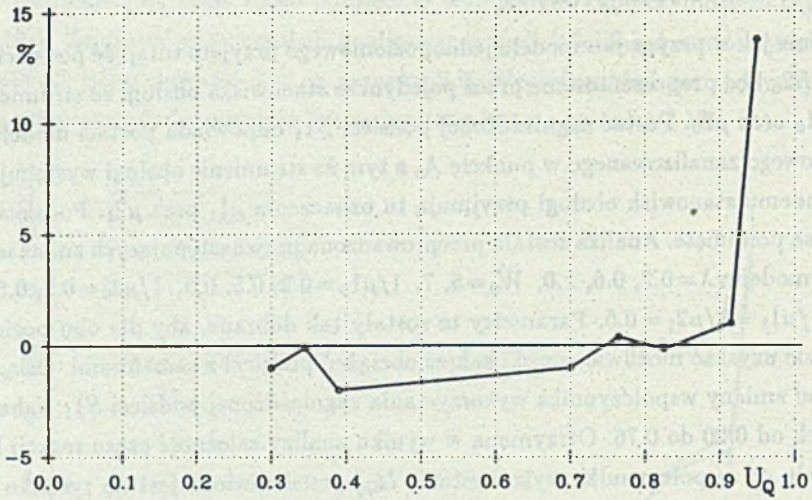
B. Model dwupoziomowy ($L=1$)

Podobnie jek w przypadku modelu jednopoziomowego przyjęto tutaj, że podsieci kolejek $R1_0$ i $R2_0$ będą reprezentowane przez pojedyncze stanowiska obsługi ze strumieniami obsługi $\mu 1_0$ oraz $\mu 2_0$. Postać zagnieżdżonej podsieci $S1_1$ odpowiada postaci modelu jednopoziomowego zanalizowanego w punkcie A, z tym że strumienie obsługi występującego w nim tandemu stanowisk obsługi przyjmują tu oznaczenia $\mu 1_1$ oraz $\mu 2_1$. Podsieci $R3_0$, $R4_0$ i $S2_1$ są pominięte. Analiza została przeprowadzona przy następujących zmianach parametrów modelu: $\lambda=0.2, 0.6, 1.0$, $W_0=5, 7$, $1/\mu 1_0=0.2, 0.5, 0.8$, $1/\mu 2_0=0.2, 0.5, 0.8$, $W_1=3$ i $1/\mu 1_1=1/\mu 2_1=0.5$. Parametry te zostały tak dobrane, aby dla obu poziomów jednocześnie uzyskać możliwie szeroki zakres obciążeń podsieci z semaforami. Osiągnięte dzięki temu zmiany współczynnika wykorzystania zagnieżdżonej podsieci $S1_1$ wahają się w granicach od 0.20 do 0.76. Otrzymana w wyniku analizy zależność czasu reakcji badanego modelu od współczynnika wykorzystania U_{Q_0} przedstawiona jest na rysunku 2-11. Względny błąd pomiędzy wynikami metody i symulacyjnymi z rysunku 2-11 wykreślony jest na rysunku 2-12.



Rys.2-11. Średni czas reakcji analizowanej sieci kolejek w funkcji jej wykorzystania, $L=1$

Fig.2-11. Mean response time vs queueing network utilisation, $L=1$

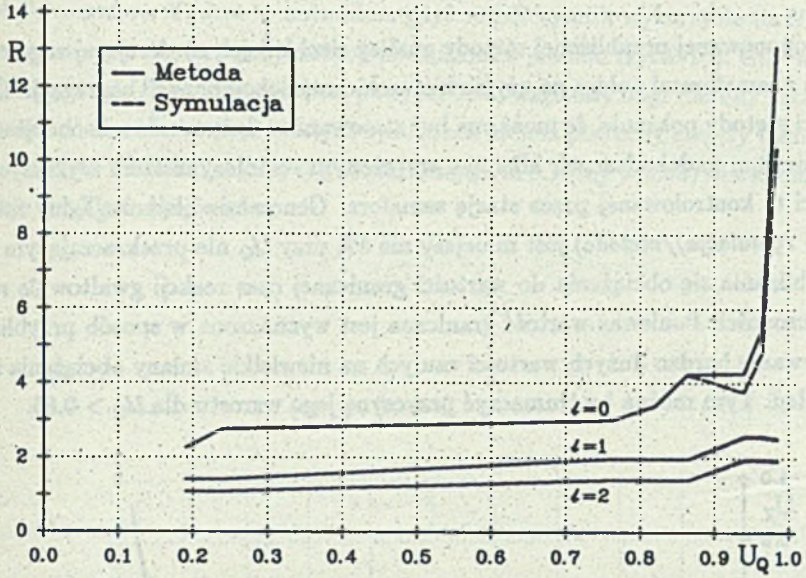


Rys.2-12. Procentowy, względny błąd wyników z rysunku 2-11

Fig.2-12. % relative error for the results given in figure 2-11

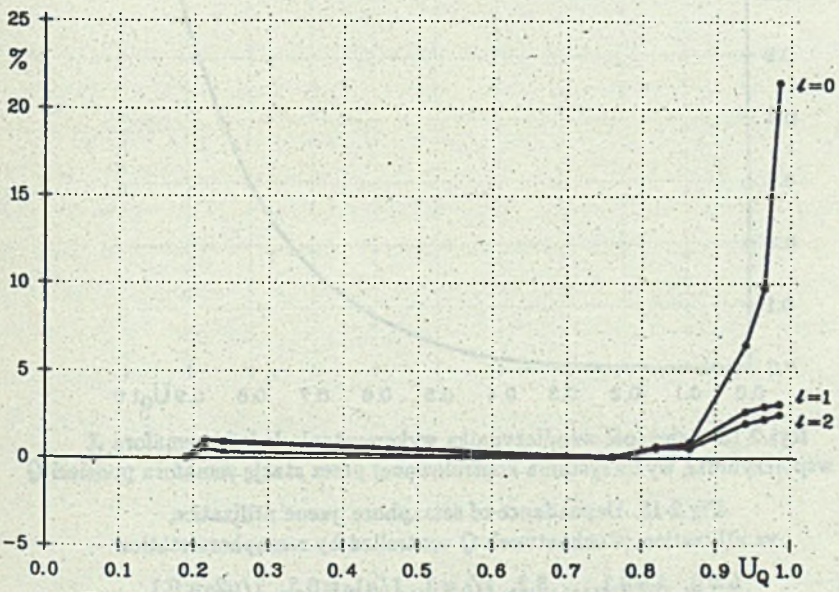
C. Model trzypoziomowy ($L=2$)

W modelu kolejkowym trzech poziomów zagnieżdżenia podsieci z semaforami podsieć $S1_1$ jest reprezentowana przez dwupoziomowy model zanalizowany w punkcie B. Podobnie jak poprzednio, podsieci kolejek $R1_0$ i $R2_0$ są reprezentowane przez pojedyncze stanowiska obsługi. Podsieci $R3_0$, $R4_0$ i $S2_1$ zostały pominięte. Analizę modelu przeprowadzono przy następujących zmianach jego parametrów: $\lambda = 0.1, 0.6, 1.0$, $W_0 = 7$, $1/\mu 1_0 = 0.2, 0.5, 0.8$, $1/\mu 2_0 = 0.2, 0.5, 0.8$, $W_1 = 5$, $1/\mu 1_1 = 1/\mu 2_1 = 0.5$, $W_2 = 3$ i $1/\mu 1_2 = 1/\mu 2_2 = 0.5$. Przy tak dobranych parametrach osiągnięte zmiany współczynników wykorzystania zagnieżdżonych podsieci pierwszego i drugiego poziomu $S1_1$ i $S1_2$ wahają się w granicach od 0.10 do 0.85. Otrzymane w wyniku analizy zależności czasu reakcji poszczególnych poziomów, tj. badanego modelu oraz zagnieżdżonych podsieci $S1_1$ i $S1_2$ od współczynnika wykorzystania U_{Q_0}, przedstawione są na rysunku 2-13. Względny błąd pomiędzy wynikami metody i symulacyjnymi z rysunku 2-13 wykreślony jest na rysunku 2-14.



Rys.2-13. Średni czas reakcji poszczególnych poziomów analizowanej sieci kolejek w funkcji jej globalnego wykorzystania, $L=2$

Fig.2-13. Mean response time for respective layer vs global queueing network utilisation, $L=2$

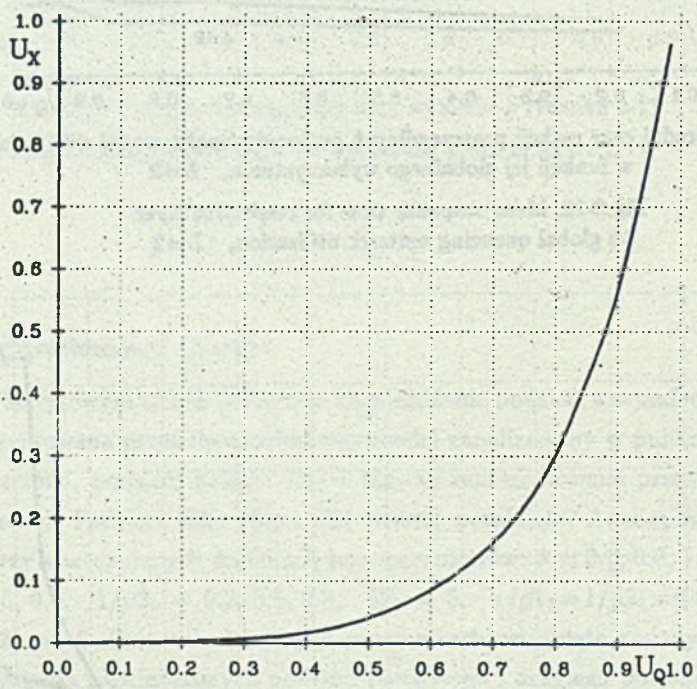


Rys.2-14. Procentowy, względny błąd wyników z rysunku 2-13

Fig.2-14. % relative error for the results given in figure 2-13

2.4. PODSUMOWANIE

Idea proponowanej przybliżonej metody analizy sieci kolejek zawierającej zagnieżdżone podsieci z semaforami polega na użyciu hierarchicznej dekompozycji i agregacji. Ocena dokładności metody pokazuje, że może ona być stosowana w dużym zakresie obciążeń modelu strumieniem nadchodzących klientów wyrażonym współczynnikiem wykorzystania U_Q podsieci Q kontrolowanej przez stację semafora. Generalnie błąd względny metody (*(metoda - symulacja)/metoda*) jest mniejszy niż 5% przy U_Q nie przekraczającym 0.85. W miarę zbliżania się obciążenia do wartości granicznej czas reakcji gwałtownie rośnie do nieskończoności. Ponieważ wartość graniczna jest wyznaczona w sposób przybliżony, to porównywanie bardzo dużych wartości czułych na niewielkie zmiany obciążenia może dać duży błąd. Tym można by tłumaczyć przyczynę jego wzrostu dla $U_Q > 0.85$.



Rys.2-15. Zależność współczynnika wykorzystania kolejki semafora X od współczynnika wykorzystania kontrolowanej przez stację semafora podsieci Q

Fig.2-15. Dependance of semaphore queue utilization vs utilization of subnetwork Q controlled by semaphore station

$L=0$, $\lambda=0.1,\ldots,3.2$, $W_0=3$, $1/\mu 1_0=0.3$, $1/\mu 2_0=0.1$

Dokładność metody jest bardziej wrażliwa na zmianę współczynnika wykorzystania U_X kolejki semafora X. Jest to naturalne, gdyż współczynnik wykorzystania U_X zaczyna gwałtownie rosnąć dopiero przy dużych obciążeniach modelu (rys.2-15). Gdy współczynnik wykorzystania U_Q osiąga wartości powyżej 0.90, względny błąd metody zaczyna przekraczać 5%. Jednakże w rzeczywistych systemach nie ma potrzeby analizy takich przypadków. Praktyka dowodzi, że przy U_X przekraczającym 0.7 ($U_Q > 0.93$) sieć komunikacyjna przestaje zwykle funkcjonować.

Rozdział 3

SIECI KOLEJEK Z SEMAFORAMI

3.1. WPROWADZENIE

Przybliżone rozwiązanie pozwalające otrzymać globalne miary wydajności otwartych sieci kolejek zawierających podsieci o limitowanej populacji klientów jest stosunkowo łatwe. Każdą podsieć modeluje się za pomocą zespołu stacji semafora i kontrolowanej przez nią podsieci, którego właściwości pozwalają zagregować go do równoważnego stanowiska obsługi o czasie reakcji i strumieniu obsługi zależnymi od liczby klientów na stanowisku. Po ostatnim kroku agregacji otrzymuje się sieć kolejek bez semaforów, dla której można już obliczyć globalne miary jej wydajności.

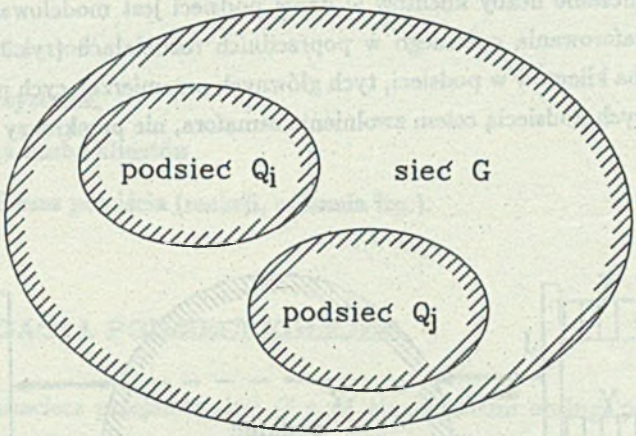
W niniejszym rozdziale zaprezentowana jest analiza, która rozszerza i uogólnia zakres powyższej metody aproksymacji. Podsieci kolejek o limitowanej populacji klientów mogą zawierać oprócz klasycznych stanowisk obsługi dalsze podsieci o podobnej strukturze. Model kolejkowy może być typu zamkniętego lub otwartego, ze stałym lub zależnym od liczby klientów w sieci strumieniem przybywających klientów. Efekt ograniczania liczby klientów w podsieciach kolejek jest tu modelowany również za pomocą mechanizmu semaforowania. Zaprezentowana metoda pozwala otrzymać zarówno globalne miary wydajności całej sieci kolejek, jak i każdej podsieci, każdego stanowiska oraz kolejek semaforów.

Rozdział jest zorganizowany w sposób następujący. W punkcie 2 wprowadzone są pojęcia odnoszące się do uogólnionej sieci kolejek z podsieciami o limitowanej liczbie klientów. Punkt 3 prezentuje technikę agregacji podsieci. Punkty 4 i 5 opisują metody rozwiązania sieci kolejek typu zamkniętego i otwartego otrzymanych w wyniku agregacji. Zaprezentowane algorytmy sugerują jednocześnie sposób ich programowej implementacji w narzędziu programowym analizy tego typu modeli. Podsumowanie i kilka ograniczeń stosowalności proponowanej metody zawarte są w punkcie 6.

3.2. MODEL

Rozważmy otwartą lub zamkniętą sieć kolejek G zwaną siecią główną lub podstawową, obejmującą dowolną liczbę podsieci Q. Niech Ω_i będzie zbiorem stanowisk obsługi w podsieci Q_i. Dopuszczalne są następujące relacje między każdymi dwoma podsieciami Q_i i Q_j:

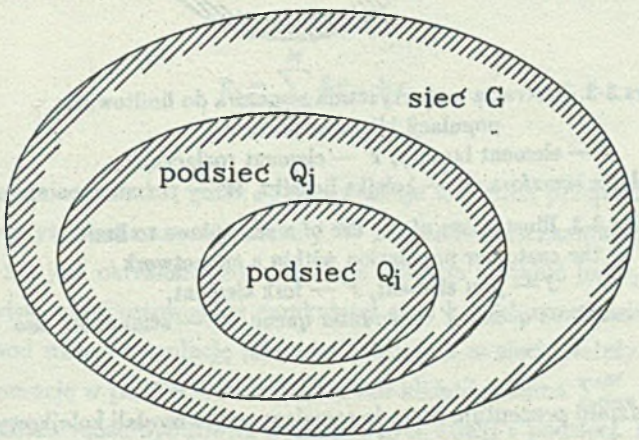
A. $\Omega_i \cap \Omega_j = \phi$ — podsieci Q_i i Q_j są niezależne (rys.3-1).



Rys.3-1. Przykład niezależnych podsieci

Fig.3-1. An example of independent subnetworks

B. $\Omega_i \cap \Omega_j = \Omega_i$ — podsieć Q_i jest zagnieżdżona w podsieci Q_j (rys.3-2).

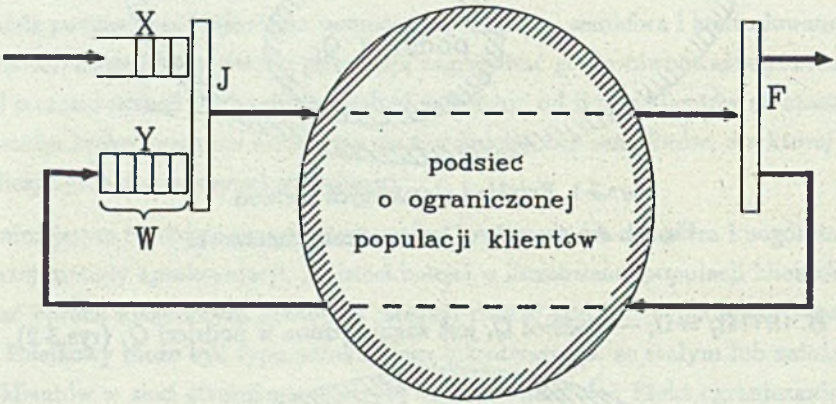


Rys.3-2. Przykład zagnieżdżonych podsieci

Fig.3-2. An example of nested subnetworks

Każdej podsieci odpowiada określony poziom jej zagnieżdżenia w sieci głównej. (Przyjmując zerowy poziom zagnieżdżenia dla sieci głównej, w przypadku B z rysunku 3-2 podsieć Q_j znajduje się na poziomie zagnieżdżenia 1, a podsieć Q_i na poziomie zagnieżdżenia 2). Załóżmy, że w analizowanym modelu kolejkowym podsieci sięgają do poziomu zagnieżdżenia L.

Stanowiska obsługi analizowanego modelu są typu *FCFS, LCFS-PR, PS* lub *IS*, z czasem obsługi o rozkładzie wykładniczym. Oznacza to, że jeśli nie uwzględniać limitowania liczby klientów w podsieciach, to sieć kolejek G odpowiada kryteriom modelu *BCMP* [BASK75]. Ograniczenie liczby klientów w danej podsieci jest modelowane za pomocą mechanizmu semaforowania opisanego w poprzednich rozdziałach (rys.3-3). Powoduje on, że łączna liczba klientów w podsieci, tych głównych przemierzających podsieć i dodatkowych wracających podsiecią celem zwolnienia semafora, nie przekroczy jego rozmiaru W.



Rys.3-3. Ilustracja wykorzystania semafora do limitowania populacji klientów podsieci
J — element łączący, F — element rozłączający,
X — kolejka semafora, Y — kolejka licznika, W — rozmiar semafora

Fig.3-3. Illustration of the use of a semaphore to limit the customer population within a subnetwork
J — join element, F — fork element,
X — semaphore queue, Y — counter queue, W — semaphore size

Dalsza część rozdziału prezentuje metodę rozwiązywania modeli kolejkowych zawierających semafory. Zgodnie z dotychczasowymi wnioskami wypływającymi z analizy stacji

semafora, idea metody opiera się na zastępowaniu podsieci z semaforami na kolejnych malejących poziomach zagnieżdżenia (od największego do najmniejszego) stanowiskami zastępczymi o czasie reakcji oraz strumieniu obsługi zależnymi od populacji klientów na stanowisku, a następnie rozwiązaniu otrzymanej na końcu sieci. Podczas każdej agregacji podsieci w zastępcze stanowisko obsługi liczone są warunkowe parametry wydajnościowe wszystkich stanowisk z dotychczas zagregowanych podsieci. Na końcu algorytmu oblicza się miary wydajności całej sieci kolejek, każdego stanowiska w sieci G, podsieci oraz kolejek semaforów. Są to:

$\mathcal{T}$ — przepustowość,

$\mathcal{U}$ — wykorzystanie,

$\mathcal{L}$ — średnia liczba klientów,

$\mathcal{R}$ — średni czas przejścia (reakcji, czekania itp.).

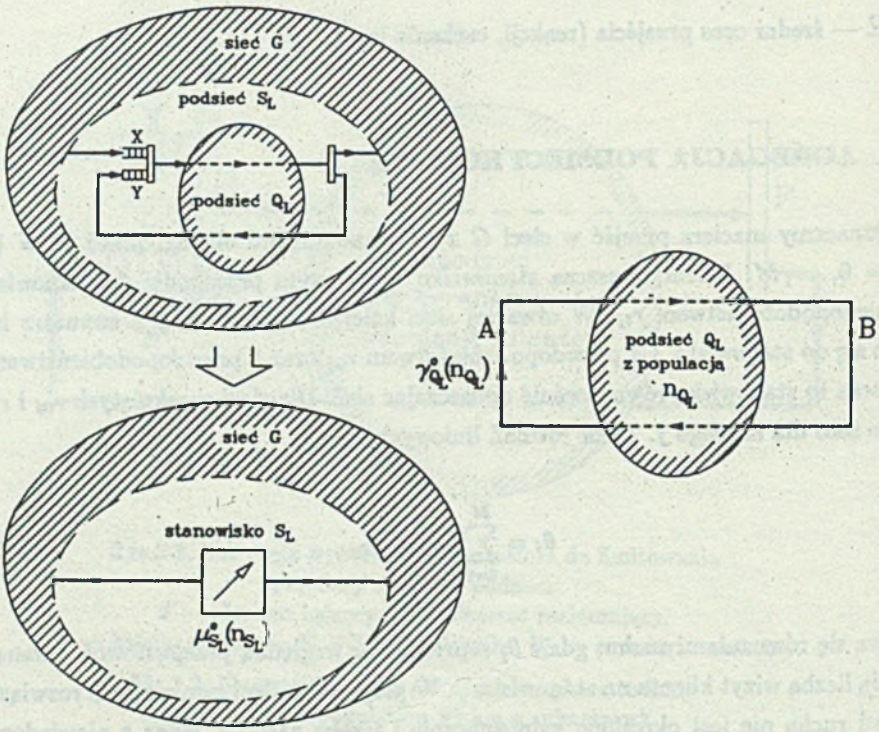
3.3. AGREGACJA PODSIECI KOLEJEK

Oznaczmy macierz przejść w sieci G z M stanowiskami obsługi przez $\mathbf{R} = \{r_{ij}\}$; $i,j = 0,\ldots,M$. Klient opuszcza stanowisko i, po czym przechodzi do stanowiska j z prawdopodobieństwem r_{ij}. W otwartej sieci kolejek przybywający z zewnątrz klient udaje się do stanowiska j z prawdopodobieństwem r_{0j} oraz z prawdopodobieństwem r_{j0} opuszcza to stanowisko równocześnie opuszczając sieć. Dla sieci zamkniętych r_{0j} i r_{j0} są równe zero dla każdego j. Zbiór równań liniowych:

$$\theta_j = \sum_{i=1}^{M} \theta_i r_{ij} + r_{0j} , \tag{3-1}$$

nazywa się równaniami ruchu, gdzie θ_j reprezentuje względną przepustowość zwaną też średnią liczbą wizyt klienta na stanowisku j. W przypadku sieci zamkniętych rozwiązanie równań ruchu nie jest określone jednoznacznie i trzeba zastąpić jedną z niewiadomych stałą. Rzeczywistą przepustowość w zamkniętej sieci kolejek otrzymuje się przez skalowanie biorąc pod uwagę populację klientów krążących w sieci. Należy zauważyć, że ze względu na generację w podsieciach dodatkowych klientów, suma $\sum_{j=0}^{M} r_{ij}$, która zwykle równa jest jedności, może być równa 2 dla pewnych stacji i podsieci. Poprzez wprowadzenie mechanizmu semaforowania stacjami tymi są elementy rozłączające F.

Rozważmy podsieć kolejek S_L złożoną z semafora i podsieci Q_L o limitowanej liczbie klientów na poziomie zagnieżdżenia L. Na tym poziomie zagnieżdżenia podsieć Q_L nie zawiera innych podsieci o limitowanej liczbie klientów i spełnia założenia modelu *BCMP*. W myśl twierdzenia Nortona w odniesieniu do modeli kolejkowych (porównaj [CHAN75a]) i wniosków wypływających z dotychczasowych rozważań podsieć S_L jest zastępowana równoważnym stanowiskiem obsługi o przepustowości–strumieniu obsługi (liczbą klientów, których można obsłużyć na jednostkę czasu) $\mu^{\circ}_{S_L}(n_{S_L})$, równym szybkości zwalniania semafora przez klientów oraz czasie reakcji $\mathcal{R}^{\circ}_{S_L}(n_{S_L})$, równym czasowi przejścia klienta od wejścia do wyjścia podsieci S_L. Wielkości $\mu^{\circ}_{S_L}(n_{S_L})$ oraz $\mathcal{R}^{\circ}_{S_L}(n_{S_L})$ otrzymuje się licząc przepustowość $\gamma^{\circ}_{Q_L}(n_{Q_L})$ oraz czas przejścia $\mathcal{R}^{\circ}_{Q_L}(n_{Q_L})$ w zamkniętej wyizolowanej od sieci głównej podsieci Q_L z n_{Q_L} klientami; $n_{Q_L} = 1, \ldots, W_{S_L}$, gdzie W_{S_L} jest rozmiarem semafora.



Rys.3-4. Idea agregacji podsieci S_L. Podsieć S_L jest zastępowana stanowiskiem ze strumieniami obsługi $\mu^{\circ}_{S_L}(n_{S_L})$ oraz czasami obsługi $\mathcal{R}^{\circ}_{S_L}(n_{S_L})$ otrzymanymi z analizy wyizolowanej podsieci Q_L

Fig.3-4. The idea of subnetwork S_L aggregation. Subnetwork S_L is replaced by a queue-dependent centre whose service rate $\mu^{\circ}_{S_L}(n_{S_L})$ and response time $\mathcal{R}^{\circ}_{S_L}(n_{S_L})$ have been obtained from the analysis of subnetwork Q_L in isolation

W zamkniętej jak na rysunku 3-4 podsieci kolejek Q_L otrzymujemy najpierw względną przepustowość $\theta_i^{S_L}$, dla każdego stanowiska i w tej podsieci przyjmując względną przepustowość w punkcie A równą 1. Można pokazać, że względna przepustowość θ_i stanowiska i w sieci głównej, otrzymana z rozwiązania zbioru równań (3-1), spełnia następującą relację:

$$\theta_i = \theta_i^{S_L} \cdot \theta_{S_L} , \tag{3-2}$$

gdzie θ_{S_L} jest względną przepustowością klientów w równoważnym stanowisku obsługi otrzymanym z agregacji podsieci S_L. Niech Ω_{S_L} będzie zbiorem stanowisk obsługi w podsieci Q_L, a M liczbą wszystkich stanowisk w sieci głównej G. Dla $i, j = 1, \ldots, M$ otrzymujemy wtedy:

$$\theta_{S_L} = \sum_{j:j\in\Omega_{S_L}} \Big(\sum_{i:i\notin\Omega_{S_L}} \theta_i r_{ij} + r_{0j} \Big) = \sum_{i:i\notin\Omega_{S_L}} \sum_{j:j\in\Omega_{S_L}} (\theta_j r_{ji} + r_{j0}) . \tag{3-3}$$

Posługując się algorytmem *MVA* [REIS80] obliczamy teraz dla każdej populacji n_{Q_L}: $n_{Q_L} = 0, \ldots, W_{S_L}$, klientów w zamkniętej sieci kolejek Q_L na poziomie zagnieżdżenia L następujące parametry:

$\gamma^{\circ}_{Q_L}(n_{Q_L})$ — przepustowość podsieci (w punkcie A z rysunku 3-4),
$\mathcal{R}^{\circ}_{Q_L}(n_{Q_L})$ — czas przejścia klienta przez podsieć Q_L (od punktu A do punktu B z rysunku 3-4),
$\mathcal{L}^{\circ}_i(n_{Q_L})$ — średnia liczba klientów na stanowisku $i : i \in \Omega_{S_L}$.

Oznaczmy przez n_{S_L} populację klientów w podsieci S_L obejmującą klientów z podsieci Q_L oraz kolejki X semafora. Wtedy:

$$n_{Q_L} = min(n_{S_L}, W_{S_L}) .$$

Ostatecznie podsieć kolejek S_L jest zastępowana stanowiskiem obsługi o względnej przepustowości θ_{S_L} i strumieniu obsługi:

$$\begin{aligned} \mu^{\circ}_{S_L}(n_{S_L}) &= \gamma^{\circ}_{Q_L}(n_{Q_L}) \\ &= \begin{cases} \gamma^{\circ}_{Q_L}(n_{S_L}) & \text{dla } n_{S_L} < W_{S_L} , \\ \gamma^{\circ}_{Q_L}(W_{S_L}) & \text{dla } n_{S_L} \geq W_{S_L} , \end{cases} \end{aligned} \tag{3-4}$$

oraz czasie reakcji:

$$\mathcal{R}^{\circ}_{S_L}(n_{S_L}) = \begin{cases} \mathcal{R}^{\circ}_{Q_L}(n_{S_L}) & \text{dla } n_{S_L} \leq W_{S_L} , \\ \mathcal{R}^{\circ}_{Q_L}(W_{S_L}) + \frac{n_{S_L} - W_{S_L}}{\gamma^{\circ}_{Q_L}(W_{S_L})} & \text{dla } n_{S_L} > W_{S_L} . \end{cases} \tag{3-5}$$

Dodatkowy składnik we wzorze (3-5) dla $n_{S_L} > W_{S_L}$ jest czasem oczekiwania klientów w kolejce semafora i pochodzi z reguły Little'a [LITT61].

Do zastąpienia pozostałych podsieci kolejek o limitowanej liczbie klientów na poziomie zagnieżdżenia L należy użyć tej samej metody. Agregacje te doprowadzają do otrzymania sieci kolejek z $L-1$ poziomami zagnieżdżenia podsieci i, na tym samym poziomie $L-1$, stanowiskami obsługi reprezentującymi zagregowane podsieci o strumieniu obsługi i czasie reakcji zależnymi od populacji klientów na stanowisku. Stosując zmodyfikowany algorytm *MVA* (patrz punkt 4 tego rozdziału), dostosowany do rozwiązywania zamkniętych sieci kolejek zawierających stanowiska o zależnym czasie obsługi, w podobny sposób można dokonać agregacji podsieci na kolejno zmniejszających się poziomach zagnieżdżenia $L-1,\ldots,1$. Na końcu otrzymuje się równoważną sieć kolejek G^* już bez semaforów, której miary wydajności mogą być obliczone np. metodami proponowanymi w punktach 4 i 5 rozdziału.

3.4. ZAMKNIĘTA SIEĆ KOLEJEK

Rozważmy zamkniętą sieć kolejek G nie zawierającą semaforów i spełniającą założenia modelu *BCMP*. Z M stanowisk obsługi wchodzących w skład sieci wyodrębniamy podsieć O składającą się z U stanowisk obsługi ponumerowanych od 1 do U. Niech N będzie populacją krążących w sieci G klientów, natomiast n_O stanem podsieci, tj. liczbą klientów przebywających w podsieci O, a n_m stanem stanowiska m: $m = 1,\ldots,M$, w sieci G. Iloczynowa forma rozwiązania zamkniętej sieci kolejek określa prawdopodobieństwo stanu stacjonarnego, w którym populacja klientów n_O znajduje się w podsieci O przy N klientach krążących w sieci G [CHAN75a]:

$$p_O(n_O|N) = \frac{\mathrm{V}_{1\ldots U}(n_O)\cdot \mathrm{V}_{U+1\ldots M}(N-n_O)}{\mathrm{V}_{1\ldots M}(N)}\,, \tag{3-6}$$

gdzie:

$$\mathrm{V}_{i\ldots j}(n) = \sum_{(n_i\ldots n_j)} \prod_{m=i}^{j} V_m(n_m)\,,$$

$(n_i \ldots n_j)$ – wszystkie możliwe stany takie, że $\sum_{m=i}^{j} n_m = n$,

$$V_m(n_m) = \prod_{k=1}^{n_m} \frac{\theta_m}{\mu_m(k)}\,,$$

z $\mu_m(k)$ określającym strumień obsługi klientów na stanowisku m gdy przebywa w nim k klientów ($\mu_m(k) = \mu_m$ dla stanowisk typu *FCFS*, *LCFS-PR* i *PS*; $\mu_m(k) = k\mu_m$ dla stanowiska typu *IS*) oraz θ_m otrzymanym z równania (3-1). Jeśli uwzględnić [REIS81a], że przepustowość podsieci O w sieci G z N klientami jest równa:

$$\mathcal{T}_O(N) = \frac{V_{1...M}(N-1)}{V_{1...M}(N)} \cdot \theta_O \,, \tag{3-7}$$

oraz przepustowość wyizolowanej od sieci G zamkniętej podsieci O z n_O klientami jest równa:

$$\gamma_O^o(n_O) = \frac{V_{1...U}(n_O-1)}{V_{1...U}(n_O)} \cdot \theta_O \,, \tag{3-8}$$

z θ_O obliczonymi według (3-3), to po kilku prostych przekształceniach otrzymujemy rekurencyjny wzór na wartość prawdopodobieństwa stanu stacjonarnego podsieci O w sieci G:

$$p_O(n_O|N) = \frac{\mathcal{T}_O(N)}{\gamma_O^o(n_O)} \cdot p_O(n_O-1|N-1) \,, \tag{3-9}$$

oraz wzór określający czas reakcji podsieci O (czas przejścia klienta przez podsieć O) przy N klientach krążących w sieci G:

$$\mathcal{R}_O(N) = \sum_{n_O=1}^{N} \frac{n_O}{\gamma_O^o(n_O)} \cdot p_O(n_O-1|N-1) \tag{3-10}$$

$$\mathcal{R}_O(N) = \sum_{n_O=1}^{N} \mathcal{R}_O^o(n_O) \cdot p_O(n_O-1|N-1) \,, \tag{3-11}$$

czasem reakcji $\mathcal{R}_O^o(n_O)$ otrzymanym z analizy podsieci O w izolacji.

Na powyższej rekurencji opiera się znany iteracyjny algorytm *MVA* [REIS80], wyznaczający parametry wydajnościowe poszczególnych stanowisk obsługi w funkcji stopniowo rosnącej populacji klientów w sieci. Jeśli bowiem podsieć O reprezentuje pojedyncze stanowisko obsługi m, to wzór (3-10) upraszcza się do postaci:

$$\mathcal{R}_m(N) = \begin{cases} \frac{1}{\mu_m}(1 + \mathcal{L}_m(N-1)) & \text{dla stanowisk } \textit{FCFS,LCFS-PR} \text{ i } \textit{PS}, \\ \frac{1}{\mu_m} & \text{dla stanowiska } \textit{IS}, \end{cases} \tag{3-12}$$

z μ_m określającymi strumień obsługi klientów na stanowisku m.

Z powyższych zależności wynika, że podsieć O może być zastąpiona pojedynczym stanowiskiem obsługi z czasem reakcji $\mathcal{R}_O^\circ(n_O)$ oraz przepustowością $\gamma_O^\circ(n_O)$ zależnymi od populacji n_O klientów na stanowisku i otrzymanymi z analizy podsieci O w izolacji. Na tym stwierdzeniu bazuje idea analizy sieci kolejek z semaforami, polegająca na użyciu parametrów obsługi podsieci S z semaforami w miejsce parametrów obsługi podsieci O. W tym celu we wzorze (3-11) jako czas reakcji $\mathcal{R}_O^\circ(n_O)$ zostaje użyty czas reakcji $\mathcal{R}_S^\circ(n_S)$ podsieci S otrzymany ze wzoru (3-5). Czas reakcji $\mathcal{R}_S^\circ(n_S)$ został wyznaczony w funkcji populacji n_S wszystkich klientów krążących w izolowanej podsieci S. Populacja ta obejmuje zarówno klientów głównych, jak i wygenerowanych klientów dodatkowych, zwalniających semafory, gdy tymczasem ci ostatni nie są ujęci w populacji N klientów sieci podstawowej G. Zatem wymagana we wzorze (3-11) wartość prawdopodobieństwa $p_S(n_S|N)$ z poprzedniego kroku iteracji jest w przybliżony sposób wyznaczana ze wzoru (3-9) przy zastąpieniu przepustowości $\gamma_O^\circ(n_O)$ przepustowością $\mu_S^\circ(n_S)$ podsieci S z semaforami otrzymaną ze wzoru (3-4). Ostatecznie otrzymujemy:

$$\mathcal{R}_S(N) = \sum_{n_S=1}^{N} \mathcal{R}_S^\circ(n_S) \cdot p_S(n_S - 1|N-1) , \tag{3-13}$$

$$p_S(n_S|N) = \frac{\mathcal{T}_S(N)}{\mu_S^\circ(n_S)} \cdot p_S(n_S - 1|N-1) . \tag{3-14}$$

Klienci dodatkowi, wracający podsiecią Q (rys.3-3) celem zwolnienia semafora, rywalizują w czasie z klientami głównymi, którzy po przejściu przez resztę zamkniętej sieci G z powrotem przybywają do kolejki X semafora podsieci S. Zaproponowane przybliżenie ma sens, jeśli dla każdej populacji n_S przepustowość $\mathcal{T}_S(N)$ podsieci S w sieci G jest mniejsza od przepustowości $\mu_S^\circ(n_S)$. Oznacza to, że musi istnieć małe prawdopodobieństwo wystąpienia większej liczby klientów w podsieci S niż w sieci G (tj. $n_S > N$). Warunek ten jest spełniony w przypadku analizy modeli kolejkowych mechanizmu sterowania przepływem za pomocą okna stosowanego w sieciach komunikacyjnych, co jest przedmiotem rozważań tej pracy. W przypadkach ogólnych nie jest on prawdziwy i może prowadzić do dużych błędów. Łatwo pokazać, że w pewnych przypadkach zastosowanie wzoru (3-14) może dać wartość prawdopodobieństwa większą od jedności.

Uwzględniając powyższe założenia, zmodyfikowany iteracyjny algorytm *MVA* analizy zamkniętej sieci kolejek G^* zawierającej stanowiska odpowiadające zagregowanym podsieciom z semaforami przedstawia się następująco. Niech Ω_M będzie zbiorem wszystkich zwykłych stanowisk obsługi w sieci G^*, a Ω_S zbiorem wszystkich stanowisk reprezentujących zagregowane podsieci z semaforami. Zbiór $\Omega_M \cup \Omega_S$ obejmuje wszystkie stanowiska występujące w sieci G^*.

- Warunki początkowe:

  $N := 0;$
  $\mathcal{L}_i(0) := 0; \quad i \in \Omega_M \cup \Omega_S;$
  $p_s(0|0) := 1; \quad s \in \Omega_S.$

- Krok 1:

  $N := N + 1;$
  Obliczenie czasu reakcji stanowisk:
  $\mathcal{R}_m(N)$, $m \in \Omega_M$, ze wzoru (3-12), $\mathcal{R}_s(N)$, $s \in \Omega_S$, ze wzoru (3-13).

- Krok 2:

  Obliczenie przepustowości stanowisk przy użyciu reguły Little'a [LITT61] w odniesieniu do całej sieci G^*:
  $\mathcal{T}_i(N) := \frac{N}{\sum_{j \in \Omega_M \cup \Omega_S} \mathcal{R}_j(N) \cdot \theta_j} \cdot \theta_i; \quad i \in \Omega_M \cup \Omega_S.$

- Krok 3:

  Obliczenie średniej liczby klientów na każdym ze stanowisk przy użyciu reguły Little'a w odniesieniu do pojedynczego stanowiska (w przypadku zagregowanych podsieci z semaforami liczba ta nie obejmuje wygenerowanych klientów dodatkowych):
  $\mathcal{L}_i(N) := \mathcal{R}_i(N) \cdot \mathcal{T}_i(N); \quad i \in \Omega_M \cup \Omega_S.$

- Krok 4:

  Obliczenie prawdopodobieństw stanu stanowisk reprezentujących zagregowane podsieci z semaforami (stan n_s obejmuje tutaj klientów głównych i wygenerowanych klientów dodatkowych):
  $p_s(n_s|N); \quad s \in \Omega_S,$ ze wzoru (3-14) dla każdej osiągalnej populacji n_s,
  $p_s(0|N) := 1 - \sum_{n_s=1}^{N} p_s(n_s|N)$.

Kroki 1 do 4 są powtarzane do osiągnięcia zadanej końcowej populacji N. Brakujący parametr, jakim jest wykorzystanie stanowiska, oblicza się na końcu iteracji za pomocą wzoru:

$$\mathcal{U}_i(N) = \begin{cases} \frac{\mathcal{T}_i(N)}{\mu_i} & \text{dla stanowisk } \textit{FCFS, LCFS-PR} \text{ i } \textit{PS}, \\ 1 - p_i(0|N) & \text{dla innego typu stanowisk.} \end{cases} \tag{3-15}$$

Na końcu algorytmu otrzymujemy poszukiwane miary wydajności sieci kolejek G^* przy populacji N krążących w niej klientów. Dla stanowisk o zależnych parametrach obsługi reprezentujących zagregowane podsieci z semaforami otrzymujemy również prawdopodobieństwa $p_s(n_s|N)$, $s \in \Omega_S$, używane do wyznaczania miar wydajności stanowisk zagregowanych podsieci oraz kolejek semaforów według wzorów:

— średnia liczba klientów w kolejce X semafora podsieci S:

$$\mathcal{L}_X(N) = \begin{cases} \sum_{n_S=W_S}^{N} (n_S - W_S) \cdot p_S(n_S|N) & \text{dla } N > W_S \,, \\ 0 & \text{dla } N \leq W_S \,; \end{cases} \tag{3-16}$$

— przepustowość $\mathcal{T}_X(N)$ równa jest przepustowości $\mathcal{T}_S(N)$ stanowiska–podsieci S;

— średni czas oczekiwania w kolejce X:

$$\mathcal{R}_X(N) = \frac{\mathcal{L}_X(N)}{\mathcal{T}_X(N)} \,; \tag{3-17}$$

— wykorzystanie kolejki semafora X:

$$\mathcal{U}_X(N) = \begin{cases} 1 - \sum_{n_S=0}^{W_S} p_S(n_S|N) & \text{dla } N > W_S \,, \\ 0 & \text{dla } N \leq W_S \,; \end{cases} \tag{3-18}$$

— prawdopodobieństwa stanu podsieci Q zawartej w podsieci S:

$$p_Q(n_Q|N) = \begin{cases} p_S(n_Q|N) & \text{dla } n_Q < W_S \,, \\ \sum_{k=W_S}^{N} p_S(k|N) & \text{dla } n_Q = W_S \text{ i } N \geq W_S \,, \\ 0 & \text{w innym przypadku;} \end{cases} \tag{3-19}$$

— średnia liczba klientów w podsieci Q (liczba ta obejmuje zarówno klientów głównych jak i wygenerowanych klientów dodatkowych):

$$\mathcal{L}_Q(N) = \sum_{n_Q=1}^{W_S} n_Q \cdot p_Q(n_Q|N) \,; \tag{3-20}$$

— przepustowość $\mathcal{T}_Q(N)$ równa jest przepustowości $\mathcal{T}_S(N)$;

— średni czas reakcji podsieci Q:

$$\mathcal{R}_Q(N) = \mathcal{R}_S(N) - \mathcal{R}_X(N) \,. \tag{3-21}$$

Prawdopodobieństwa stanu podsieci Q pozwalają wyznaczyć miary wydajności dla stacji w niej zawartych. Niech c będzie wybranym stanowiskiem obsługi, kolejką semafora lub zagnieżdżoną podsiecią o limitowanej populacji klientów zawartą w podsieci Q. Wtedy:

— przepustowość stacji c:

$$\mathcal{T}_c(N) = \mathcal{T}_S(N) \cdot \frac{\theta_c}{\theta_S}, \tag{3-22}$$

z θ_c i θ_S otrzymanymi odpowiednio ze wzorów (3-1) oraz (3-3);

— średnia liczba klientów w stacji c:

$$\mathcal{L}_c(N) = \sum_{n_Q=1}^{W_S} \mathcal{L}_c^o(n_Q) \cdot p_Q(n_Q|N), \tag{3-23}$$

gdzie: $\mathcal{L}_c^o(n_Q)$ jest średnią liczbą klientów w stacji c otrzymaną podczas analizy sieci Q w izolacji;

— średni czas reakcji stacji c:

$$\mathcal{R}_c(N) = \frac{\mathcal{L}_c(N)}{\mathcal{T}_c(N)}. \tag{3-24}$$

Tego samego zmodyfikowanego algorytmu *MVA* używa się również do agregacji podsieci na kolejno zmniejszających się poziomach $l : l = L-1, \ldots, 1$, ich zagnieżdżenia w sieci głównej G. Występujący w algorytmie parametr N odpowiada wówczas populacji klientów w wyizolowanej agregowanej podsieci S_l; Ω_M jest zbiorem pojedynczych stanowisk tej podsieci, a Ω_S – zbiorem stanowisk odpowiadających już zagregowanym podsieciom z poziomu zagnieżdżenia $l+1$. Wyliczone parametry wydajnościowe, po dokonaniu agregacji podsieci S_l, są następnie wykorzystywane przy analizie modelu na płytszym poziomie zagnieżdżenia $l-1$.

3.5. OTWARTA SIEĆ KOLEJEK

Niech $\lambda(N)$ będzie Poissonowskim strumieniem klientów przybywających do sieci kolejek G^* z M^* stanowiskami obsługi, zależnym od aktualnej liczby N, $N = 0, \ldots, \infty$, klientów w sieci, a $\mathcal{S}(N)$ zbiorem stanów $(n_i, \ldots, n_{M^*})$ sieci, takich, że $\sum_{i=1}^{M^*} n_i = N$ i $n_i \geq 0$.

Iloczynowa forma rozwiązania otwartej sieci kolejek spełniającej założenia modelu *BCMP* jest znanym wzorem określającym prawdopodobieństwo stanu stacjonarnego, w którym n klientów znajduje się na stanowisku i [BASK75]:

$$p_i(n) = \frac{\sum_{N=n}^{\infty} \Lambda(n) \cdot \sum_{\mathcal{S}(N):n_i=n} \mathbf{V}(N)}{\mathbf{H}}, \tag{3-25}$$

gdzie:

$$\mathbf{H} = \sum_{N=0}^{\infty} \Lambda(N) \cdot \sum_{\mathcal{S}(N)} \mathbf{V}(N),$$

$$\Lambda(N) = \prod_{j=0}^{N-1} \lambda(j),$$

$$\mathbf{V}(N) = V_1(n_1) \cdot V_2(n_2) \cdot \ldots \cdot V_{M^*}(n_{M^*}),$$

$$V_i(n_i) = \prod_{j=1}^{n_i} \frac{\theta_i}{\mu_i(j)},$$

z $\mu_i(j)$ określającym strumień obsługi na stanowisku i, gdy przebywa w nim j klientów ($\mu_i(j) = \mu_i$ dla stanowisk typu *FCFS*, *LCFS-PR* i *PS*; $\mu_i(j) = j\mu_i$ dla stanowiska typu *IS*) oraz θ_i otrzymanym z równań (3-1) lub (3-3). Powyższe równania i następne, prezentowane poniżej, są prawdziwe przy założeniu, że istnieje stan stacjonarny sieci, tzn. że sieć jest w stanie przepuścić wszystkich przybywających ze strumieniem $\lambda(N)$ klientów, innymi słowy, że ich liczba w sieci nie rośnie wraz z czasem do nieskończoności.

Ze względów programowej implementacji algorytmu należy ograniczyć dowolność zależności strumienia przybywających klientów $\lambda(N)$ od liczby N klientów w sieci w ten sposób, aby dla pewnej skończonej wartości a, $a \geq 0$, zachodziło $\lambda(N) = \lambda(a)$ dla $N > a$. Przy tym założeniu wygodnie jest rozpatrzyć trzy niezależne przypadki:

A) niezależny stały strumień klientów, tj. $\lambda(N) = \lambda$, $N = 0, \ldots, \infty$;

B) sieć o ograniczonej populacji klientów, tj. $\lambda(a) = 0$, $a > 0$;

C) sieć o nieograniczonej populacji klientów, tj. $\lambda(a) > 0$, $a > 0$.

A. Niezależny stały strumień klientów

Jeżeli strumień λ klientów przybywających do sieci kolejek spełniającej założenia modelu *BCMP* jest stały, wtedy, co łatwo wykazać, każde stanowisko i sieci może być analizowane niezależnie od pozostałych, z indywidualnym strumieniem λ_i przybywających klientów, równym $\lambda\theta_i$.

Przyjmijmy upraszczająco, że w sieci kolejek G^* zawierającej stanowiska reprezentują-

ce zagregowane podsieci z semaforami strumień λ; zachowuje swój charakter poissonowski. Założenie to pozwala uprościć analizę sieci i również analizować każde z jej stanowisk z osobna, niezależnie od pozostałych. Sieć taka posiada stan stacjonarny, jeśli kryteria jego istnienia są spełnione indywidualnie dla każdego ze stanowisk. Globalne miary wydajności sieci, jak czas reakcji lub średnia liczba klientów, są sumą odpowiednich parametrów wydajnościowych poszczególnych stanowisk.

Rozważmy stanowisko S z czasem obsługi zależnym od liczby klientów na stanowisku, reprezentujące zagregowaną podsieć S z poziomu zagnieżdżenia 1 w sieci G. Analogicznie do równania (3-4) strumień obsługi na stanowisku S wynosi:

$$\mu_S(n_S) = \begin{cases} \gamma^o_{Q_1}(n_S) & \text{dla } n_S < W_S\,, \\ \gamma^o_{Q_1}(W_S) & \text{dla } n_S \geq W_S\,, \end{cases}$$

gdzie W_S jest rozmiarem semafora z podsieci S. Z wyrażenia (3-25) otrzymujemy następujący wzór na wartość prawdopodobieństwa $p_S(n_S)$ stanu stacjonarnego znalezienia n klientów na stanowisku S:

$$p_S(n_S) = \frac{\prod_{j=1}^{n_S} \frac{\lambda\theta_S}{\mu_S(j)}}{\sum_{N=0}^{\infty} \prod_{j=1}^{N} \frac{\lambda\theta_S}{\mu_S(j)}} = \frac{\prod_{j=1}^{n_S} \frac{\lambda\theta_S}{\mu_S(j)}}{\prod_{j=1}^{W_S} \frac{\lambda\theta_S}{\mu_S(j)}} \cdot \frac{\prod_{j=1}^{W_S} \frac{\lambda\theta_S}{\mu_S(j)}}{\sum_{N=0}^{\infty} \prod_{j=1}^{N} \frac{\lambda\theta_S}{\mu_S(j)}}\,,$$

$$p_S(n_S) = \begin{cases} \prod_{j=n+1}^{W_S} \frac{\mu_S(j)}{\lambda\theta_S} \cdot p_S(W_S) & \text{dla } n_S < W_S\,, \\ \left(\frac{\lambda\theta_S}{\mu_S(W_S)}\right)^{n_S - W_S} \cdot p_S(W_S) & \text{dla } n_S \geq W_S\,, \end{cases} \tag{3-26}$$

wraz z

$$p_S(W_S) = \frac{\prod_{j=1}^{W_S} \frac{\lambda\theta_S}{\mu_S(j)}}{\sum_{N=0}^{W_S-1} \prod_{j=1}^{N} \frac{\lambda\theta_S}{\mu_S(j)} + \sum_{N=W_S}^{\infty} \prod_{j=1}^{N} \frac{\lambda\theta_S}{\mu_S(j)}}\,,$$

$$[p_S(W_S)]^{-1} = \sum_{N=1}^{W_S} \prod_{j=N}^{W_S} \frac{\mu_S(j)}{\lambda\theta_S} + \frac{1}{1 - \frac{\lambda\theta_S}{\mu_S(W_S)}} \tag{3-27}$$

i warunkiem istnienia stanu stacjonarnego $\frac{\lambda\theta_S}{\mu_S(W_S)} < 1$. Otrzymane prawdopodobieństwa $p_S(n_S)$ z sumą:

$$\sum_{n_S=W_S}^{\infty} p_S(n_S) = \frac{1}{1 - \frac{\lambda\theta_S}{\mu_S(W_S)}} \cdot p_S(W_S)\,, \tag{3-28}$$

używa się następnie do wyznaczenia miar wydajności podsieci S oraz stanowisk w niej zawartych. Dla podsieci S obejmującej stację semafora z kolejką semafora X oraz podsieć Q o limitowanej populacji klientów otrzymujemy:

— średnią liczbę klientów w kolejce semafora X:

$$\begin{aligned} \mathcal{L}_X &= \sum_{n_S=W_S}^{\infty} (n_S - W_S) \cdot p_S(n_S) \\ &= \frac{\frac{\lambda\theta_S}{\mu_S(W_S)}}{\left(1 - \frac{\lambda\theta_S}{\mu_S(W_S)}\right)^2} \cdot p_S(W_S) \, ; \end{aligned} \tag{3-29}$$

— przepustowość $\mathcal{T}_X$ równa jest strumieniowi $\lambda\theta_S$ klientów przybywających do podsieci–stanowiska S;

— średni czas oczekiwania w kolejce X:

$$\mathcal{R}_X = \frac{\mathcal{L}_X}{\mathcal{T}_X} \, ; \tag{3-30}$$

— wykorzystanie kolejki semafora X:

$$\begin{aligned} \mathcal{U}_X &= 1 - \sum_{n_S=0}^{W_S} p_S(n_S) \\ &= \frac{\frac{\lambda\theta_S}{\mu_S(W_S)}}{1 - \frac{\lambda\theta_S}{\mu_S(W_S)}} \cdot p_S(W_S) \, ; \end{aligned} \tag{3-31}$$

— prawdopodobieństwa stanu podsieci Q zawartej w podsieci S:

$$p_Q(n_Q) = \begin{cases} p_S(n_Q) & \text{dla } n_Q < W_S \, , \\ \frac{1}{1 - \frac{\lambda\theta_S}{\mu_S(W_S)}} \cdot p_S(W_S) & \text{dla } n_Q = W_S \, ; \end{cases} \tag{3-32}$$

— średnią liczbę klientów w podsieci Q (liczba ta obejmuje zarówno klientów głównych jak i wygenerowanych klientów dodatkowych):

$$\mathcal{L}_Q = \sum_{n_Q=1}^{W_S} n_Q \cdot p_Q(n_Q) \, ; \tag{3-33}$$

— średnią liczbę klientów głównych w podsieci Q (liczba ta nie obejmuje generowanych w podsieci Q klientów dodatkowych):

$$\mathcal{L}_{Q_m} = \sum_{n_Q=1}^{W_S} \mathcal{L}^{\circ}_{Q_m}(n_Q) \cdot p_Q(n_Q)$$

$$= \sum_{n_Q=1}^{W_S} \mathcal{R}^{\circ}_{Q}(n_Q) \cdot \gamma^{\circ}_{Q}(n_Q) \cdot p_Q(n_Q) \; ; \qquad (3\text{-}34)$$

— przepustowość $\mathcal{T}_Q$ równa jest strumieniowi $\lambda\theta_S$ klientów przybywających do podsieci-stanowiska S;

— średni czas reakcji podsieci Q:

$$\mathcal{R}_Q = \frac{\mathcal{L}_{Q_m}}{\mathcal{T}_Q} \; ; \qquad (3\text{-}35)$$

— średnią liczbę klientów w stanowisku–podsieci S:

$$\mathcal{L}_S = \mathcal{L}_X + \mathcal{L}_Q \; ; \qquad (3\text{-}36)$$

— średni czas reakcji stanowiska–podsieci S:

$$\mathcal{R}_S = \mathcal{R}_X + \mathcal{R}_Q \; . \qquad (3\text{-}37)$$

Prawdopodobieństwa stanu podsieci Q pozwalają również wyznaczyć miary wydajności dla stacji w niej zawartych w sposób analogiczny do przypadku analizy zamkniętych sieci kolejek (wzory (3-22), (3-23) i (3-24)).

B. Sieć o ograniczonej populacji klientów

Jeśli strumień przybywających klientów jest taki, że $\lambda(N) > 0$ dla $N < a$ i $\lambda(N) = 0$ dla $N \geq a$, to sieć kolejek G^* można zastąpić równoważną zamkniętą siecią G°, obejmującą sieć G^* oraz dodatkowe stanowisko O zastępujące źródło, o czasie obsługi zależnym od liczby klientów na stanowisku. Klienci opuszczający stanowisko O udają się do sieci G^*, po czym, po przejściu przez tę sieć, powracają do stanowiska O. Jeśli dla stanowiska O

przyjmiemy $\theta_O = 1$ i $\mu_O(n_O) = \lambda(a-n_O)$, a a będzie liczbą klientów krążących w równoważnej zamkniętej sieci G^o, to miary wydajności każdego stanowiska z sieci G^* oraz G^o będą identyczne.

Na podstawie równania (3-25) dla stanowiska i z sieci kolejek G^* otrzymujemy:

$$p_{i,G^*}(n) = \frac{\sum_{N=n}^{a} \prod_{j=0}^{N-1} \lambda(j) \cdot \sum_{\mathcal{S}(N):n_i=n} \mathbf{V}(N)}{\sum_{N=0}^{a} \prod_{j=0}^{N-1} \lambda(j) \cdot \sum_{\mathcal{S}(N)} \mathbf{V}(N)}$$

$$= \frac{\prod_{k=0}^{a-1} \lambda(k) \cdot \sum_{N=n}^{a} \prod_{j=N}^{a-1} \frac{1}{\lambda(j)} \cdot \sum_{\mathcal{S}(N):n_i=n} \mathbf{V}(N)}{\prod_{k=0}^{a-1} \lambda(k) \cdot \sum_{N=0}^{a} \prod_{j=N}^{a-1} \frac{1}{\lambda(j)} \cdot \sum_{\mathcal{S}(N)} \mathbf{V}(N)}$$

$$= \frac{\sum_{N=n}^{a} \prod_{j=1}^{a-N} \frac{1}{\lambda(a-j)} \cdot \sum_{\mathcal{S}(N):n_i=n} \mathbf{V}(N)}{\sum_{N=0}^{a} \prod_{j=1}^{a-N} \frac{1}{\lambda(a-j)} \cdot \sum_{\mathcal{S}(N)} \mathbf{V}(N)} . \qquad (3\text{-}38)$$

Dla dodatkowego stanowiska O:

$$V_O(n_O) = \prod_{j=1}^{n_O} \frac{1}{\mu_O(j)} = \prod_{j=1}^{n_O} \frac{1}{\lambda(a-j)} . \qquad (3\text{-}39)$$

Ostatecznie otrzymujemy więc:

$$p_{i,G^*}(n) = \frac{\sum_{N=n}^{a} V_O(a-N) \cdot \sum_{\mathcal{S}(N):n_i=n} \mathbf{V}(N)}{\sum_{N=0}^{a} V_O(a-N) \cdot \sum_{\mathcal{S}(N)} \mathbf{V}(N)} = \frac{\sum_{\mathcal{S}^o(a):n_i=n} \mathbf{V}^o(a)}{\sum_{\mathcal{S}^o(a)} \mathbf{V}^o(a)} = p_{i,G^o}(n) , \qquad (3\text{-}40)$$

gdzie: $\mathcal{S}^o(a)$ oznacza zbiór stanów $(n_0, n_1, \ldots, n_{M^*})$ sieci G^o takich, że $\sum_{i=0}^{M^*} n_i =$ i $n_i \geq 0$, natomiast $\mathbf{V}^o(a) = V_O(n_O) \cdot V_1(n_1) \cdot, \ldots, \cdot V_{M^*}(n_{M^*})$.

Przez użycie zmodyfikowanego algorytmu *MVA* uwzględniającego stanowiska o zale nym czasie obsługi (opisanego w poprzednim punkcie tego rozdziału) otrzymuje się ost tecznie poszukiwane miary wydajności globalnej oraz poszczególnych stanowisk sieci k lejek G. Przepustowość $\mathcal{T}_O(a)$ obliczona dla dodatkowego stanowiska O jest jednocześn przepustowością otwartej sieci G.

C. Sieć o nieograniczonej populacji klientów

W przypadku gdy strumień klientów przybywających do sieci kolejek jest taki, $\lambda(N) > 0$ dla $N > a$ i $\lambda(N) = \lambda(a)$, $\lambda(a) > 0$, dla $N \geq a$, stała normalizacyjna

z równania (3-25) przybiera postać:

$$H = \sum_{N=0}^{\infty} \prod_{j=0}^{N-1} \lambda(j) \cdot \sum_{S(N)} V(N)$$

$$= \sum_{N=0}^{a-1} \prod_{j=0}^{N-1} \lambda(j) \cdot \sum_{S(N)} V(N) + \frac{\prod_{j=0}^{a-1} \lambda(j)}{(\lambda(a))^a} \cdot \sum_{N=a}^{\infty} (\lambda(a))^N \cdot \sum_{S(N)} V(N)$$

$$= \sum_{N=0}^{a-1} \prod_{j=0}^{N-1} \lambda(j) \cdot \sum_{S(N)} V(N) +$$

$$\frac{\prod_{j=0}^{a-1} \lambda(j)}{(\lambda(a))^a} \left\{ \sum_{N=0}^{\infty} (\lambda(a))^N \cdot \sum_{S(N)} V(N) - \sum_{N=0}^{a-1} (\lambda(a))^N \cdot \sum_{S(N)} V(N) \right\}$$

$$= \frac{\sum_{N=0}^{a-1} \prod_{j=1}^{a-1-N} \frac{1}{\lambda(a-1-j)} \cdot \sum_{S(N)} V(N)}{\prod_{j=1}^{a-1} \frac{1}{\lambda(a-1-j)}} +$$

$$\frac{\prod_{j=0}^{a-1} \lambda(j)}{(\lambda(a))^a} \left\{ \sum_{N=0}^{\infty} (\lambda(a))^N \cdot \sum_{S(N)} V(N) - \frac{\sum_{N=0}^{a-1} \frac{1}{(\lambda(a))^{a-1-N}} \cdot \sum_{S(N)} V(N)}{\frac{1}{(\lambda(a))^{a-1}}} \right\}$$

$$= \frac{1}{p_{G^o}(0)} + \frac{\prod_{j=0}^{a-1} \lambda(j)}{(\lambda(a))^a} \left\{ \frac{1}{p(0|\lambda = \lambda(a))} - \frac{1}{p_{G^o}(0|\mu_O = \lambda(a))} \right\}, \qquad (3\text{-}41)$$

gdzie:

$p_{G^o}(0)$ — prawdopodobieństwo, że sieć kolejek G^* zamknięta przez dodatkowe stanowisko O ze strumieniem obsługi $\mu_O(n_O) = \lambda(a-1-n_O)$ jest pusta. Niech G^o będzie zamkniętą siecią kolejek zawierającą sieć G^* i stanowisko O, a $a-1$ populacją klientów w sieci G^o; wtedy $p_{G^o}(0) = p_O(a-1)$;

$p(0|\lambda = \lambda(a))$ — prawdopodobieństwo, że sieć kolejek G^* ze stałym strumieniem przybywających klientów równym $\lambda(a)$ jest pusta;

$p_{G^o}(0|\mu_O = \lambda(a))$ — prawdopodobieństwo analogiczne do $p_{G^o}(0)$, ale ze stałym $\mu_O = \lambda(a)$.

Podobnie oznaczmy przez:

$p_{i,G^o}(n)$ — prawdopodobieństwo obecności n klientów na stanowisku i sieci kolejek G^* zamkniętej przez dodatkowe stanowisko O; $\mu_O(n_O) = \lambda(a-1-n_O)$, a $a-1$ jest populacją klientów w sieci G^o;

$p_i(n|\lambda=\lambda(a))$ — prawdopodobieństwo obecności n klientów na stanowisku i w sieci kolejek G^* ze stałym strumieniem przybywających klientów równym $\lambda(a)$;

$p_{i,G^o}(n|\mu_O=\lambda(a))$ — prawdopodobieństwo analogiczne do $p_{i,G^o}(n)$, ale ze stałym $\mu_O = \lambda(a)$.

Analogicznie do przypadku stałej normalizacyjnej H równanie (3-25) przybiera postać:

$$p_i(n) = \frac{\frac{p_{i,G^o}(n)}{p_{G^o}(0)} + \frac{\prod_{j=0}^{a-1}\lambda(j)}{(\lambda(a))^a}\left\{\frac{p_i(n|\lambda=\lambda(a))}{p(0|\lambda=\lambda(a))} - \frac{p_{i,G^o}(n|\mu_O=\lambda(a))}{p_{G^o}(0|\mu_O=\lambda(a))}\right\}}{\frac{1}{p_{G^o}(0)} + \frac{\prod_{j=0}^{a-1}\lambda(j)}{(\lambda(a))^a}\left\{\frac{1}{p(0|\lambda=\lambda(a))} - \frac{1}{p_{G^o}(0|\mu_O=\lambda(a))}\right\}}, \tag{3-42}$$

z $p_{i,G^o}(n)$ oraz $p_{i,G^o}(n|\mu_O=\lambda(a))$ równymi zero dla $n \geq a$.

Z wyrażenia (3-42) wynika sposób obliczania parametrów wydajnościowych dla sieci G^*, i ostatecznie dla sieci G, polegający na wstępnym obliczeniu tych parametrów w zamkniętych sieciach kolejek Q^o i $Q^o(\mu_O = \lambda(a))$ oraz w sieci G^* ze stałym strumieniem przybywających klientów równym $\lambda(a)$. Kryteria istnienia stanu stacjonarnego sieci G są również takie same jak dla sieci G^* z $\lambda = \lambda(a)$.

Niech α będzie jedną z miar wydajności, oznaczającą: przepustowość, średnią liczbę klientów lub wykorzystanie: sieci G, albo dowolnego stanowiska, kolejki semafora, lub podsieci w G. Ze wzoru (3-42) otrzymujemy:

$$\alpha = \frac{\frac{\alpha_{G^o}}{p_{G^o}(0)} + \frac{\prod_{j=0}^{a-1}\lambda(j)}{(\lambda(a))^a}\left\{\frac{\alpha(\lambda=\lambda(a))}{p(0|\lambda=\lambda(a))} - \frac{\alpha_{G^o}(\mu_0=\lambda(a))}{p_{G^o}(0|\mu_0=\lambda(a))}\right\}}{\frac{1}{p_{G^o}(0)} + \frac{\prod_{j=0}^{a-1}\lambda(j)}{(\lambda(a))^a}\left\{\frac{1}{p(0|\lambda=\lambda(a))} - \frac{1}{p_{G^o}(0|\mu_0=\lambda(a))}\right\}} \tag{3-43}$$

Średni czas reakcji otrzymamy na końcu algorytmu wykorzystując regułę Little'a.

3.6. PODSUMOWANIE

W rozdziale tym została przedstawiona metodyka rozwiązywania modeli kolejkowych obejmujących liczne podsieci o limitowanej populacji przebywających w nich klientów. Efekt limitowania populacji jest modelowany za pomocą dodatkowych elementów modelu kolejkowego – semaforów. Specyfiką analizowanych modeli kolejkowych jest uwzględnienie niezerowego czasu zwolnienia semafora odpowiadającego czasowi powrotu przez podsieć dodatkowo wygenerowanego klienta. Cecha ta sprawia, że modele te szczególnie nadają się do analizy wydajności mechanizmu sterowania przepływem za pomocą okna stosowanego w pakietowych sieciach komunikacyjnych. Poczynione założenia nie pozwalają jednak wykorzystywać opracowanej metody w sposób ogólny. Stosowanie jej przy nie spełnionych założeniach w pewnych teoretycznych przypadkach modeli kolejkowych nie mających związku z modelowaną rzeczywistością może zawieść i dać nieprawidłowe wyniki. Na marginesie należy zauważyć, że jeśli czas zwolnienia semafora jest zerowy, tzn. klienci po opuszczeniu podsieci kolejek o limitowanej populacji klientów natychmiast zwalniają semafor, to prezentowane w rozdziale algorytmy są algorytmami dokładnymi i jedynym przybliżeniem metody jest samo zastąpienie podsieci z semaforami równoważnym stanowiskiem obsługi.

Zaprezentowane algorytmy są łatwe w programowej implementacji i z uwzględnieniem podanych ograniczeń odnoszą się do różnych typów sieci kolejek, zamkniętych i otwartych, o stałym lub zależnym od stanu sieci strumieniu przybywających klientów. Metoda została zweryfikowana poprzez symulację dając zadowalające rezultaty, co zostało przedstawione w poprzednim rozdziale pracy.

Rozdział 4

MODELOWANIE PROTOKOŁU X.25

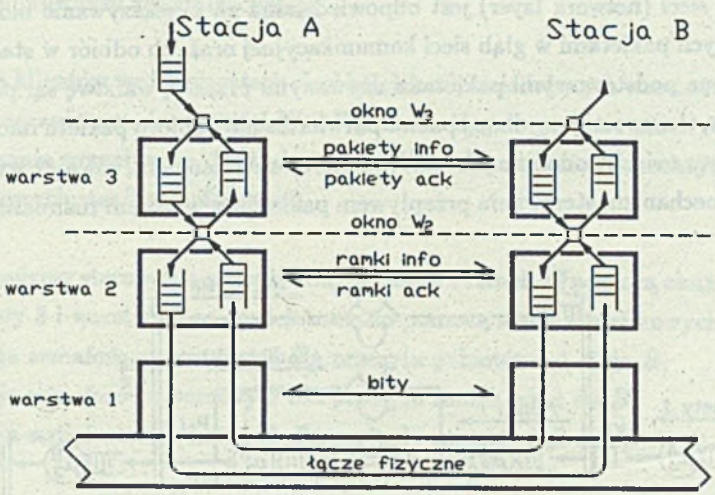
4.1. WPROWADZENIE

Ostatnie lata przyniosły znaczące prace w dziedzinie normalizacji standardów sieci komunikacyjnych. Prace te są najbardziej zaawansowane dla tzw. otwartych, rozległych sieci komputerowych; trwają nadal intensywnie nad sieciami lokalnymi. Opracowany w 1982 roku przez ISO model referencyjny systemu otwartego [ISO82] dekomponuje interfejs komunikacyjny na 7 poziomów–warstw protokołów. Dla każdej z warstw zdefiniowany jest zakres usług komunikacyjnych udostępnianych warstwie wyższej oraz realizujące te usługi protokoły. Realizacja usług każdej z warstw przebiega w sposób asynchroniczny, co powoduje tworzenie sie kolejek transmitowanych danych pomiędzy warstwami (rys.4-1). Kolejki te wynikają również z ograniczonej rozmiarem tzw. okna liczby pakietów i ramek, które można wysłać bez uzyskania potwierdzenia ich odbioru.

Złożoność architektury komunikacyjnej sprawia, że jej analiza wydajnościowa jak i dobór parametrów systemowych protokołów i stacji komunikacyjnych (rozmiary okien, buforów, bloków danych, priorytety, ...) nie są trywialne i tym samym wymagają wykorzystania odpowiednich metod badawczych. Dalsza część rozdziału prezentuje próbę obliczenia wydajności mechanizmu sterowania przeplywem za pomocą okna stosowanego w pakietowych sieciach komunikacyjnych pracujących z wykorzystaniem pocztowego protokołu komunikacyjnego rozległych sieci komputerowych – X.25 [CCITT84].

4.2. PREZENTACJA SYSTEMU

Rozważmy mechanizm wymiany informacji na poziomie trzech pierwszych warstw interfejsu komunikacyjnego kontrolowanych przez protokół X.25 (rys.4-1).



Rys.4-1. Trzy pierwsze warstwy interfejsu komunikacyjnego

Fig.4-1. First three communication layers

Znaczenie kolejnych warstw tego interfejsu jest następujące:

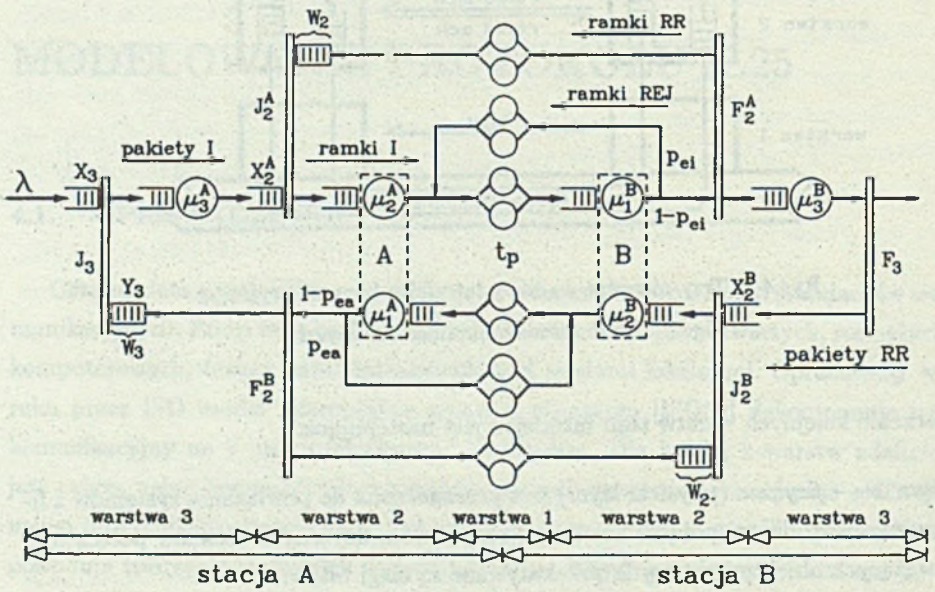
Warstwa 1 — *fizyczna* (physical layer) jest przeznaczona do powiązania systemów z fizycznymi środkami połączenia. Za pośrednictwem fizycznych środków połączenia, wiążących obiekty warstwy 2, przekazywane są ciągi bitów.

Warstwa 2 — *sterowania kanałem* (data link layer) udostępnia warstwie 3 usługi polegające na przekazywaniu bez błędów bloków danych. Bloki danych, którymi operuje ta warstwa, są nazywane ramkami. Trzema głównymi ramkami wykorzystywanymi przez protokół X.25 są:

- Ramka *I* – ramka informacyja, przekazująca blok danych;
- Ramka *RR* – ramka potwierdzająca poprawny odbiór ramki informacyjnej;
- Ramka *REJ* – ramka odmowy, żądająca ponownego nadania błędnie otrzymanej ramki informacyjnej.

Warstwa sterowania kanałem, celem ochrony przed błędami transmisji, wyposażona jest w mechanizm sterowania przepływem ramek, który zatrzymuje transmisję, gdy liczba wysłanych i nie potwierdzonych ramek informacyjnych przekroczy pewną z góry zadaną liczbę zwaną rozmiarem okna.

Warstwa 3 — *sieci* (network layer) jest odpowiedzialna za przekazywanie bloków danych zwanych pakietami w głąb sieci komunikacyjnej oraz ich odbiór w stacji adresata. Dwoma podstawowymi pakietami używanymi przez tę warstwę są: pakiet informacyjny, transportujący dane i pakiet potwierdzenia odbioru pakietu informacyjnego. Warstwa sieci, podobnie jak warstwa sterowania kanałem, również wyposażona jest w mechanizm sterowania przepływem pakietów z własnym rozmiarem okna.



Rys.4-2. Model kolejkowy mechanizmu sterowania przepływem protokołu X.25

Rys.4-2. The queueing model of X.25 flow control mechanism

Model kolejkowy odpowiadający powyższemu systemowi transmisji danych jest przedstawiony na rysunku 4-2. Charakteryzujące ten model elementy oraz przyjęte dla niego założenia są następujące:

- Strumień przybywających do sieci kolejek klientów reprezentujących przesyłane bloki danych ma charakter poissonowski z parametrem λ. Dla celów analizy w modelu został przyjęty jednokierunkowy przepływ danych (od stacji A do stacji B z rysunku 4-1).

- Klienci przemierzając sieć kolejek przechodzą przez kolejne stanowiska obsługi. Czasy pobytu w tych stanowiskach, wyznaczane według rozkładu wykładniczego, odpo-

wiadają czasom związanym z obsługą i transmisją bloków danych w kolejnych warstwach interfejsu komunikacyjnego.

- Liczba klientów we fragmentach sieci kolejek odpowiadających warstwie 3 i warstwie 2 jest ograniczona do wartości W_3 i W_2 stanowiących rozmiary okien mechanizmów sterowania przepływem dla tych warstw. Klienci nadmiarowi oczekują w kolejkach wejściowych stacji semaforowych.

- Mechanizmy sterowania przepływem pakietów i ramek za pomocą okna na poziomie warstwy 3 i warstwy 2 są modelowane za pomocą stacji semaforowych:
  – stacja semaforowa warstwy 3 dla przesyłu pakietów od A do B,
  – stacja semaforowa warstwy 2 dla przesyłu ramek od A do B,
  – stacja semaforowa warstwy 2 dla przesyłu ramek od B do A.

Klienci przybywający do sieci kolejek kierowani są do kolejki semafora X_3. Jeśli przepuszczenie przez stację semafora warstwy 3 jest możliwe, tzn. kolejka Y_3 jest niepusta, przybywający klient opuszcza kolejkę X_3 i udaje się do stanowiska ze strumieniem obsługi μ_3^A. Czas pobytu na tym stanowisku reprezentuje obsługę transmisji przez warstwę 3 po stronie nadawcy (stacja A). Po zakończeniu obsługi klient trafia do kolejki semafora X_2^A warstwy 2. Jeśli są spełnione warunki przepuszczenia, wędrujący klient jest obsługiwany na stanowisku ze strumieniem obsługi μ_2^A. W obsługę tę wkalkulowany jest czas transmisji ramki, który jest funkcją jej długości i przepustowości łącza. Następnie klient przechodzi przez stanowisko typu IS z czasem obsługi t_p równym czasowi propagacji sygnału przez łącze, po czym zjawia się w stanowisku warstwy 2 po stronie odbiorcy (stacja B) ze strumieniem obsługi μ_1^B. Po zakończeniu obsługi klient może zostać zawrócony z prawdopodobieństwem p_{ei} do stacji nadawcy. Wartość p_{ei} jest prawdopodobieństwem otrzymania błędnej ramki i konieczności jej retransmisji. W tym przypadku wysyłana jest ramka *REJ*. Błędna ramka jest retransmitowana przez warstwę 1 na zasadzie zwykłego powtórzenia fizycznej transmisji. Z prawdopodobieństwem $1-p_{ei}$ klient reprezentujący ramkę bez błędów jest kierowany do stanowiska warstwy 3 ze strumieniem obsługi μ_3^B i ostatecznie, opuszczając to stanowisko, opuszcza sieć, co odpowiada dotarciu bloku danych do użytkownika stacji B. W tym samym czasie klient wygenerowany w elemencie rozłączającym F i reprezentujący pakiet potwierdzenia jest kierowany w drogę powrotną przez sieć kolejek. Jego droga, podlegająca analogicznym prawidłom, co droga klienta reprezentującego pakiet informacyjny, rozpoczyna się wizytą w kolejce semafora X_2^B i kończy w kolejce licznika Y_3.

W realnym systemie komunikacyjnym nadawaniem i odbieraniem ramek warstwy 2 - sterowania kanałem - zajmuje się jeden procesor, przy czym odbieranie ma bezwzględny priorytet nad nadawaniem. W modelu oznacza to, że prace stanowisk obsługi ze strumieniami obsługi μ_1^A i μ_2^A (analogicznie μ_1^B i μ_2^B) wykluczają się wzajemnie. Klient przybywający do stanowiska μ_1^A wstrzymuje pracę stanowiska μ_2^A (analogicznie μ_1^B - μ_2^B). Na rysunku 4-2 skojarzone stanowiska modelu otoczone są linią przerywaną.

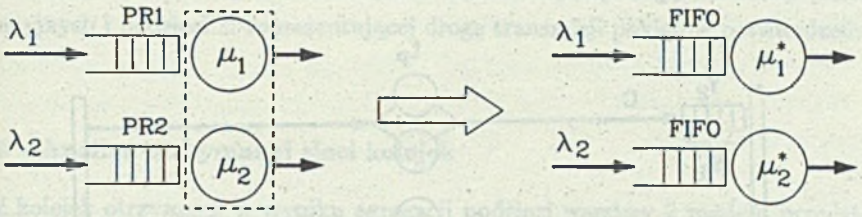
4.3. ANALIZA MODELU

Model przedstawiony na rysunku 4-2 jest zbyt złożony, aby można było bezpośrednio do niego zastosować metodę analitycznej analizy. Trzeba zatem uciec się do techniki dekompozycji pozwalającej studiować sukcesywnie kolejną z warstw modelu. W rezultacie stosowania tej warstwowej techniki rozwiązania detaliczna analiza każdej z warstw używa tylko wizji globalnej warstw niższego poziomu.

W pierwszym kroku (etap 1 i 2) zostanie zanalizowana podsieć kolejek reprezentująca warstwę 2 - sterowania kanałem. Po rozwiązaniu problemu priorytetu odbioru ramek nad ich nadawaniem w stanowiskach μ_1^A i μ_2^A oraz μ_1^B i μ_2^B (etap 1), w etapie 2 dokonana zostanie agregacja charakterystyk tej warstwy (głównie chodzi o czas reakcji) z wykorzystaniem metody zaprezentowanej w poprzednich rozdziałach pracy. W ten sposób etap 3, dotyczący analizy warstwy 3 modelu, operuje siecią kolejkową, w której warstwy 1 i 2 są reprezentowane globalnie przez zastępcze stanowisko obsługi. Analiza sieci kolejek z etapu 3 wykorzystuje również metodę opisaną w poprzednich rozdziałach tej pracy.

Etap 1 - Dekompozycja stanowisk warstwy 2

Rozróżnia się dwie fazy obsługi na tych stanowiskach: przetwarzanie ramki nadawanej (niższy priorytet) i przetwarzanie ramki odbieranej (wyższy priorytet). Zarówno priorytet względny, jak i bezwzględny mogą być tu brane pod uwagą. Zachowując wartość średnią strumienia wejściowego klientów, model obsługi priorytetowej może być zdekomponowany w sposób dokładny dla obu typów priorytetu. W analizowanym modelu kolejkowym protokołu X.25 nasze rozważania ograniczają się do przypadku priorytetu bezwzględnego. Do dekompozycji stanowisk z obsługą o tym priorytecie stosuje się metodę redukcji zajętości przedstawioną np. w [SCHM84] (rys.4-3).



Rys.4-3. Dekompozycja stanowisk z obsługą priorytetową

Fig.4-3. Decomposition of the priority service centers

W wyniku otrzymujemy wyrażenia na wartość równoważnych strumieni obsługi przy założeniu niezależności stanowisk:

$$\mu_1^* = \mu_1 , \tag{4-1}$$

$$\mu_2^* = \mu_2 \cdot (1 - \alpha) , \tag{4-2}$$

gdzie: α jest prawdopodobieństwem, że klient przybywający do stanowiska 2 zastanie zajęte stanowisko 1. Jeśli strumienie λ_1 i λ_2 przybywających klientów są niezależne, to parametr α jest równy współczynnikowi zajętości (wykorzystania) stanowiska 1 wyrażonym przez λ_1/μ_1. W modelu przedstawionym na rysunku 4-2 liczba klientów w sieci jest limitowana do wartości W_3 i tym samym warunek o niezależności strumieni nie jest spełniony. Respektując to ograniczenie, przybliżona wartość parametru α powinna być równa udziałowi czasu obsługi $W_3 - 1$ klientów w stanowisku 1 (przy jednym kliencie przebywającym na stanowisku 2) w czasie wyznaczonym nadejściem W_3 klientów:

$$\alpha = \frac{(W_3 - 1) \cdot \lambda_1}{W_3 \cdot \mu_1} . \tag{4-3}$$

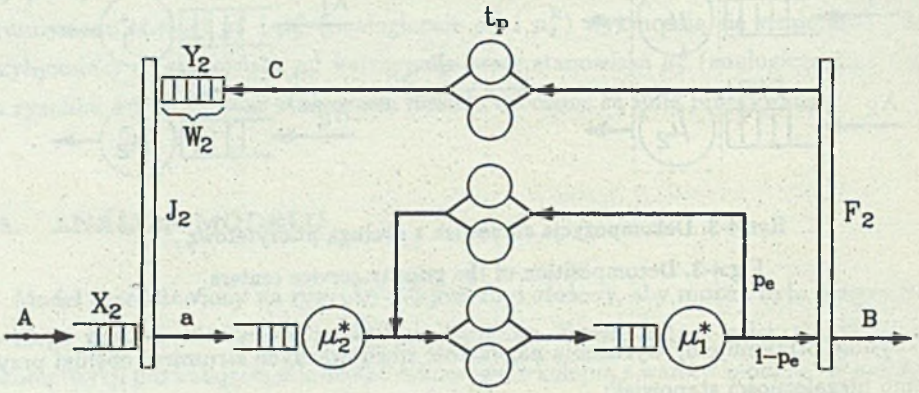
Jeśli rozmiar okna W_3 równy jest jedności, priorytet z braku konkurujących o obsługę klientów nigdy nie "zadziała"; wartość parametru α dla tego przypadku zgodnie z oczekiwaniami równa jest zeru. Przy wzroście wartości W_3 efekt limitowania populacji ma coraz mniejszy wpływ na uzależnienie strumieni λ_1 i λ_2, zatem i wartość parametru α zbliża się do wartości λ_1/μ_1.

Brakujący parametr, jakim jest strumień λ_1, otrzymuje się po uwzględnieniu strumienia retransmisji błędnie przesłanych ramek (rys.4-2):

$$\lambda_1 = \begin{cases} \lambda \cdot \frac{1}{1-p_{eA}} & \text{dla stacji } A, \\ \lambda \cdot \frac{1}{1-p_{eB}} & \text{dla stacji } B. \end{cases} \tag{4-4}$$

Etap 2 - Agregacja podsieci kolejek warstwy 2

Podsieć kolejek reprezentująca jeden z dwu komponentów warstwy 2 analizowanego modelu prezentuje rysunek 4-4.



Rys.4-4. Podsieć kolejek warstwy 2

Fig.4-4. Queueing subnetwork of layer 2

Zgodnie z techniką agregacji opisaną w rozdziale 3 tej pracy, analizowana podsieć kolejek jest zastępowana równoważnym stanowiskiem obsługi o parametrach obsługi zależnych od stanu stanowiska–podsieci. Dwa parametry są tu istotne: czas reakcji mierzony między punktami A i B z rysunku 4-4 oraz przepustowość wyrażona strumieniem klientów zwalniających semafor. Strumień ten, w funkcji liczby klientów przebywających w podsieci, wyrażony jest zależnością:

$$\gamma_2(k) = \begin{cases} \frac{k}{\mathcal{R}_{2o}(k)} & \text{dla } k \leq W_2 \ , \\ \frac{W_2}{\mathcal{R}_{2o}(W_2)} & \text{dla } k > W_2 \ , \end{cases} \tag{4-5}$$

gdzie: $\mathcal{R}_{2o}(k)$ jest czasem reakcji zamkniętej podsieci kolejek z usuniętymi elementami J i F, przy k klientach krążących w podsieci, mierzonym między punktami a i C. Czas reakcji pomiędzy punktami A i C wynosi zatem:

$$\mathcal{R}_{2s}(k) = k \cdot \frac{1}{\gamma_2(k)} \ . \tag{4-6}$$

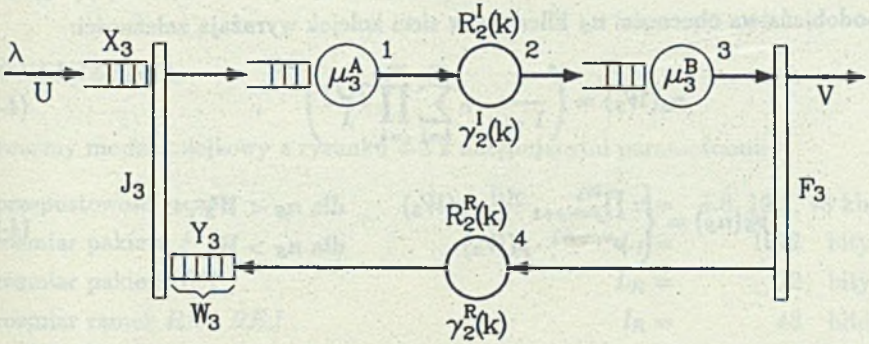
Ostatecznie, czas reakcji podsieci pomiędzy punktami A i B, niezbędny dla dalszej analizy modelu, otrzymuje się ze wzoru:

$$\mathcal{R}_2(k) = \mathcal{R}_{2s}(k) - t_p \ . \tag{4-7}$$

Przedstawiona technika agregacji ma zastosowanie dla obu komponentów warstwy 2 analizowanego modelu – podsieci kolejek I reprezentującej drogę transmisji pakietów informacyjnych i podsieci R reprezentującej drogę transmisji pakietów potwierdzeń.

Etap 3 - Analiza otrzymanej sieci kolejek

Sieć kolejek otrzymaną w wyniku agregacji podsieci warstwy 2 modelu przedstawia rysunek 4-5. Jest to sieć kolejkowa typu otwartego, zawierająca jedną stację semafora oraz kontrolowaną przez nią podsieć złożoną z czterech stanowisk obsługi.



Rys.4-5. Sieć kolejek otrzymana po agregacji podsieci warstwy 2 modelu

Fig.4-5. Queueing network obtained after aggregation of subnetwork of layer 2

Proponowana w pracy metoda analizy takiej sieci kolejek zakłada w pierwszym kroku obliczenie strumienia $\gamma(n_Q)$ klientów w zamkniętej podsieci, kontrolowanej przez stację semafora. W tym celu należy zastosować zmodyfikowany algorytm *MVA* dopuszczający stanowiska o parametrach obsługi zależnych od stanu. Każda iteracja tego algorytmu dla sukcesywnie rosnącej populacji n_Q klientów w podsieci daje:

- czasy reakcji poszczególnych stanowisk podsieci:
  $\mathcal{R}_1(n_Q)$, $\mathcal{R}_2(n_Q)$, $\mathcal{R}_3(n_Q)$ i $\mathcal{R}_4(n_Q)$;
- średni czas obiegu klientów w podsieci równy sumie czasów reakcji stanowisk:
  $\mathcal{R}^\circ(n_Q) = \mathcal{R}_1(n_Q) + \mathcal{R}_2(n_Q) + \mathcal{R}_3(n_Q) + \mathcal{R}_4(n_Q)$;
- strumień klientów w podsieci (jest to strumień klientów zwalniających semafor):
  $\gamma(n_Q) = n_Q/\mathcal{R}^\circ(n_Q)$;

oraz dla stanowisk reprezentujących zagregowane podsieci (stanowisko 2 i 4):

- prawdopodobieństwa $p_2(k|n_Q)$ i $p_4(k|n_Q)$, $k = 0,\ldots,n_Q$, obecności k klientów na stanowisku–podsieci przy n_Q klientach krążących w sieci. Należy pamiętać, że liczba k obejmuje zarówno klientów głównych, krążących w sieci, jak i wygenerowanych w zagregowanej podsieci kolejek warstwy 2 klientów dodatkowych zwalniających semafor tej warstwy.

W następnym kroku, otrzymane wielkości zgodnie ze wzorami wyprowadzonymi w poprzednim rozdziale (wzory (3-26) do (3-37)) pozwalają obliczyć poszukiwane miary wydajności analizowanego modelu. Kryterium istnienia stanu stacjonarnego, tj. istnienia rozwiązania dla tego modelu, jest spełnione, jeśli $\lambda < \gamma(W_3)$. Niech $\rho = \lambda/\gamma(W_3)$ oraz prawdopodobieństwa obecności n_S klientów w sieci kolejek wyrażają zależności:

$$p_S(W_3) = \left(\frac{1}{1-\rho} + \sum_{k=1}^{W_3} \prod_{j=k}^{W_3} \frac{\gamma(j)}{\lambda} \right)^{-1} , \tag{4-8}$$

$$p_S(n_S) = \begin{cases} \prod_{j=n_S+1}^{W_3} \frac{\gamma(j)}{\lambda} \cdot p(W_3) & \text{dla } n_S < W_3 , \\ \rho^{n_S - W_3} \cdot p(W_3) & \text{dla } n_S > W_3 . \end{cases} \tag{4-9}$$

Wtedy:

— średnia liczba klientów w kolejce semafora X_3:

$$\mathcal{L}_X = \frac{\rho}{(1-\rho)^2} \cdot p(W_3) ; \tag{4-10}$$

— średni czas oczekiwania w kolejce X_3:

$$\mathcal{R}_X = \frac{\mathcal{L}_X}{\lambda} ; \tag{4-11}$$

— prawdopodobieństwa obecności n_Q klientów w podsieci kolejek Q kontrolowanej przez stację semafora:

$$p_Q(n_Q) = \begin{cases} p_S(n_Q) & \text{dla } n_Q < W_3 , \\ \frac{1}{1-\rho} \cdot p_S(W_3) & \text{dla } n_Q = W_S ; \end{cases} \tag{4-12}$$

— średnia liczba klientów w podsieci Q:

$$\mathcal{L}_Q = \sum_{n_Q=1}^{W_3} n_Q \cdot p_Q(n_Q) ; \tag{4-13}$$

— średni czas przejścia przez podsieć Q:

$$\mathcal{R}_Q = \frac{\sum_{n_Q=1}^{W_3} n_Q \cdot \frac{\mathcal{R}_1(n_Q)+\mathcal{R}_2(n_Q)+\mathcal{R}_3(n_Q)}{\mathcal{R}^0(n_Q)} \cdot p_Q(n_Q)}{\lambda} ; \qquad (4\text{-}14)$$

— czas reakcji całej sieci kolejek (między punktami U i V z rysunku 4-5):

$$\mathcal{R} = \mathcal{R}_X + \mathcal{R}_Q . \qquad (4\text{-}15)$$

Miary wydajności stanowisk obsługi oraz zagregowanych podsieci oblicza się na podstawie prawdopodobieństw $p_Q(n_Q)$ oraz $p_2(k|n_Q)$ i $p_4(k|n_Q)$, stosując odpowiednio wzory (3-23) i (3-24).

4.4. OBLICZENIA

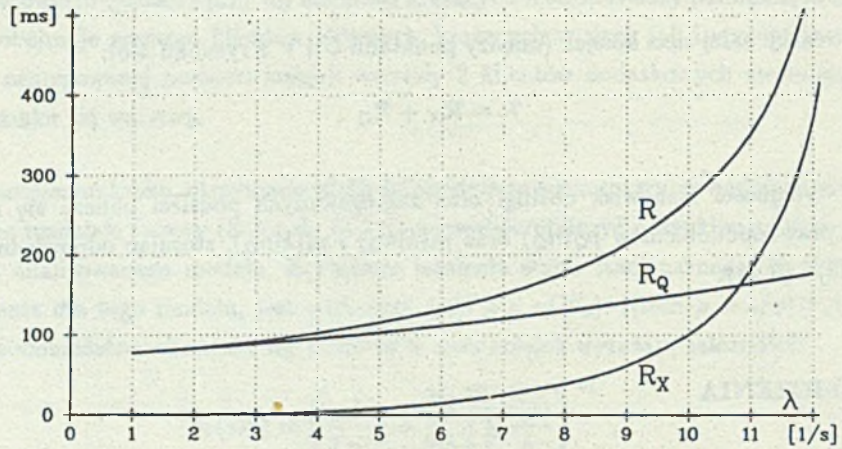
Rozważmy model kolejkowy z rysunku 4-2 z natępującymi parametrami:

- przepustowość łącza $v =$ 4.8, 19.2, 48 kb/s
- rozmiar pakietu I $L_I =$ 1072 bity
- rozmiar pakietu RR $L_R =$ 72 bity
- rozmiar ramek RR i REJ $l_R =$ 48 bitów
- prawdopodobieństwo błędu transmisji jednego bitu $p_{bit} =$ 10^{-7}

  $p_{ei} = 1-(1-p_{bit})^{1072}$, $p_{ea} = 1-(1-p_{bit})^{72}$,
- czasy obsługi na stanowiskach
  $1/\mu_3^A =$ 1 ms
  $1/\mu_3^B =$ 1.5 ms
  $1/\mu_2^A = 1 + L_I/v$ ms
  $1/\mu_2^B = 1 + L_R/v$ ms
  $1/\mu_1^A = 1/\mu_1^B = 2 + l_R/v$ ms
- czasy propagacji przez łącze
  - dla przepustowości 4.8 kb/s $t_p =$ 11 ms
  - dla przepustowości 19.2 i 48 kb/s $t_p =$ 4 ms

Poniższe rysunki od 4-6 do 4-10 prezentują niektóre z rezultatów otrzymanych w wyniku analizy modelu w funkcji zmian strumienia λ przybywających klientów oraz rozmiarów semaforów W_3 i W_2. Zasadniczym kryterium wydajności, które wzięto pod uwagę, jest czas reakcji R badanej sieci kolejek odpowiadający czasowi transmisji pakietu. Czas ten gwałtownie maleje wraz ze wzrostem rozmiarów okien W_3 i W_2, aby się ustabilizować dla

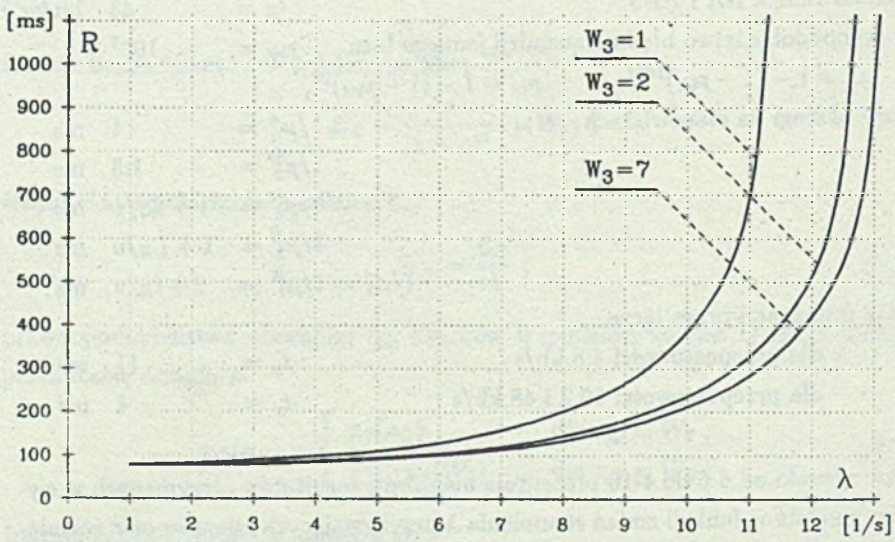
W_3 i $W_2 > 5$. Można zauważyć, że w rozważanym przykładzie optymalnymi rozmiarami okien są: $W_3 = 3$ i $W_2 = 2$.



Rys.4-6. Komponenty czasu reakcji systemu $\mathcal{R} = \mathcal{R}_X + \mathcal{R}_Q$

Fig.4-6. Components of system response time $\mathcal{R} = \mathcal{R}_X + \mathcal{R}_Q$

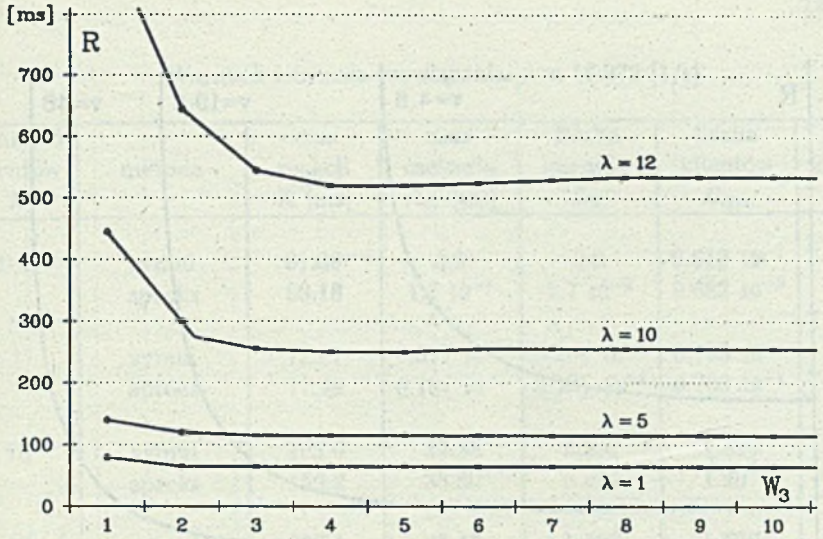
$v = 19.2$ kb/s, $W_3 = 3$, $W_2 = 1$



Rys.4-7. Czas reakcji systemu w funkcji strumienia przybywających pakietów

Fig.4-7. System response time vs packet arrival stream

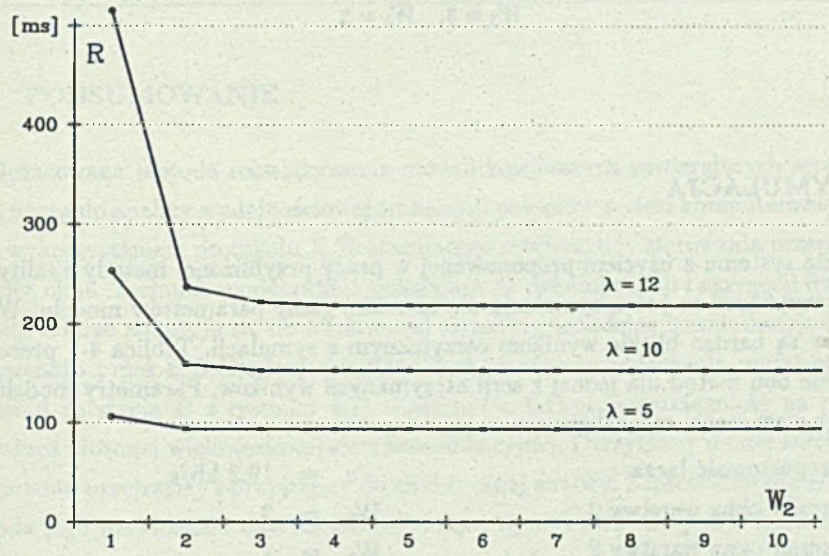
$v = 19.2$ kb/s, $W_2 = 1$



Rys.4-8. Czas reakcji systemu w funkcji rozmiaru okna W_3

Fig.4-8. System response time vs window size W_3

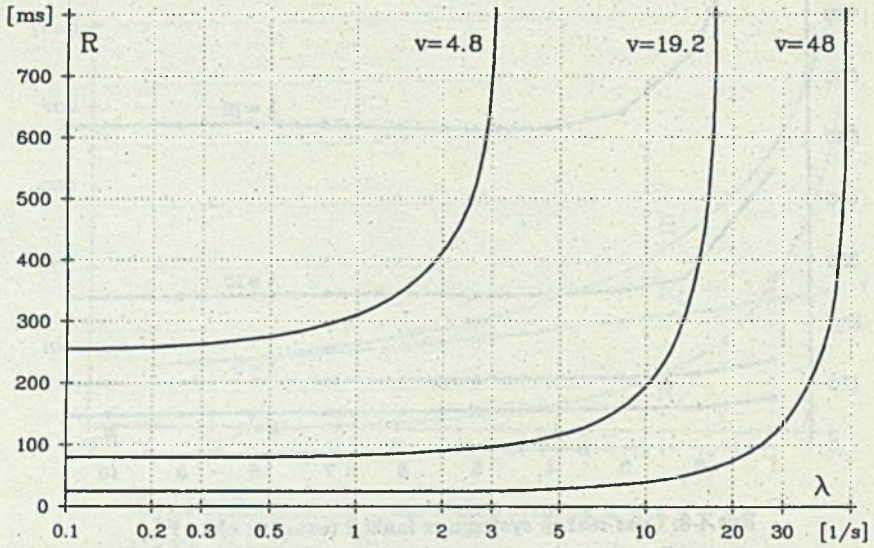
$v = 19.2$ kb/s, $W_2 = 1$



Rys.4-9. Czas reakcji systemu w funkcji rozmiaru okna W_2

Fig.4-9. System response time vs window size W_2

$v = 19.2$ kb/s, $W_3 = 7$



Rys.4-10. Porównanie czasów reakcji systemu w funkcji przepustowości łącza

Fig.4-10. Comparison of system response time vs line capacity

$W_3 = 3, \quad W_2 = 2$

4.5. SYMULACJA

Analiza systemu z użyciem proponowanej w pracy przybliżonej metody analitycznej została zweryfikowana przez symulację dla szerokiej gamy parametrów modelu. Wyniki tej analizy są bardzo bliskie wynikom otrzymanym z symulacji. Tablica 4-1 prezentuje porównanie obu metod dla jednej z serii otrzymanych wyników. Parametry modelu, dla których je otrzymano, są następujące:

- przepustowość łącza $v = 19.2$ kb/s
- rozmiar okna warstwy 3 $W_3 = 3$
- rozmiar okna warstwy 2 $W_2 = 2$

Tablica 4-1

Warunek istnienia rozwiązania: $\lambda < 16.272$ [1/s]

strumień klientów λ	metoda	czas reakcji $\mathcal{R}$ [ms]	czas czekania $\mathcal{R}_X$ [ms]	liczba klientów $\mathcal{L}_X$	liczba klientów $\mathcal{L}_{Qm}$	liczba klientów $\mathcal{L}_Y$
0.1	symul	61.39	0.0	0.0	$0.612\ 10^{-2}$	2.993
	aproks	68.18	$1.7\ 10^{-7}$	$1.7\ 10^{-9}$	$0.682\ 10^{-2}$	2.992
1	symul	72.91	$0.377\ 10^{-1}$	$0.364\ 10^{-4}$	$0.703\ 10^{-1}$	2.916
	aproks	71.49	$0.181\ 10^{-1}$	$0.181\ 10^{-4}$	$0.703\ 10^{-1}$	2.914
10	symul	163.6	39.36	0.387	1.220	1.625
	aproks	159.2	38.50	0.385	1.207	1.642
12	symul	237.1	97.45	1.166	1.669	1.142
	aproks	229.9	94.94	1.139	1.619	1.197
15	symul	913.0	750.8	11.17	2.413	0.344
	aproks	755.4	597.3	8.959	2.373	0.393

4.6. PODSUMOWANIE

Opracowana metoda rozwiązywania modeli kolejkowych zawierających semafory została użyta do analizy wydajnościowej transmisji pakietów w sieci komputerowej pracującej z wykorzystaniem protokołu X.25 stosującego mechanizmy sterowania przepływem za pomocą okna. Technika modelowania polegająca na dekompozycji i agregacji fragmentów modelu dobrze adaptuje się do warstwowej struktury interfejsu komunikacyjnego zaproponowanego przez ISO. Badany przykład bezpośredniego połączenia punkt-punkt sieci (stacji A ze stacją B z rysunku 4-1) może być z łatwością rozciągnięty na połączenie na wskroś złożonej wielowęzłowej sieci komunikacyjnej. Otrzymany model kolejkowy jest stosunkowo przejrzysty i przystępny do analitycznej analizy. Zaproponowana przybliżona metoda jego rozwiązania daje satysfakcjonujące wyniki zarówno pod względem formy – prostota formuł matematycznych, jak i pod względem wiarygodności – duża zgodność wyników z wynikami otrzymanymi z symulacji. W rezultacie pozwalają one na trafny dobór parametrów eksploatacyjnych stosowanego protokołu komunikacyjnego oraz stacji transmisyjnych (rozmiary okien, priorytety, przepustowości łączy itp.).

Zakończenie

W pracy podjęto próbę opracowania metody analizy wydajnościowej systemów informatycznych, uwzględniającej ich istotne elementy synchronizacji wymagane przy realizacji dostępu do pewnych zasobów systemu. Metodyka analizy została oparta na wykorzystaniu i rozszerzeniu modeli kolejkowych od lat znajdujących uznanie w tej dziedzinie modelowania. Analiza modelu kolejkowego dostarcza konkretnych wyników ilościowych, które wprost wskazują na sposób strojenia parametrów eksploatacyjnych badanego systemu. Jednakże klasa modeli kolejkowych posiadających rozwiązanie dokładne jest mocno ograniczona. Warunek o niezależności parametrów modelu nie dopuszcza żadnych mechanizmów synchronizacji, stąd modele takie są zwykle dużym uproszczeniem rzeczywistości. W miarę wprowadzania do modelu kolejkowego elementów przybliżających jego własności do modelowanej rzeczywistości napotyka się na coraz to większe trudności w znalezieniu dla niego metody analizy. Siłą rzeczy, każda z zaproponowanych metod staje się jedną z wielu metod przybliżonych, której jakość weryfikuje się poprzez symulację i porównanie z innymi metodami.

Problemem przewodnim pracy była próba obliczenia wydajności mechanizmu sterowania przepływem bloków danych za pomocą okna w pakietowych sieciach komunikacyjnych pracujących np. z wykorzystaniem protokołu X.25. Zagadnienie to zostało rozwiązane w sposób następujący:

1. Na podstawie analizy protokołu X.25, jako model mechanizmu sterowania przepływem za pomocą okna w sieci komputerowej została zaproponowana sieć kolejek z zagnieżdżonymi na wielu poziomach podsieciami o limitowanej populacji klientów. Limitowanie populacji wynika z ograniczonej tzw. rozmiarem okna liczby bloków danych (pakietów, ramek), które można wysłać bez otrzymania potwierdzenia ich odbioru.

2. Do modelowania ograniczania populacji klientów zostały użyte dodatkowe elementy modelu kolejkowego – semafory. Wprowadzenie mechanizmu semaforowa-

nia pozwala również uwzględnić w modelu niezerowy czas przesłania potwierdzenia odbioru bloków danych transmitowanych w sieci komputerowej.

3. Przedstawiona została analiza stacji semafora jako elementu modelu kolejkowego. Analiza ta uzasadnia metodę rozwiązywania sieci kolejek z semaforami polegającą na hierarchicznej dekompozycji i agregacji. Dekompozycja pozwala określić strumień klientów zwalniających semafor, który jest następnie wykorzystywany do rozwiązania stacji semafora. W wyniku agregacji stacja semafora z kontrolowaną przez semafor podsiecią są zastępowane pojedynczym, zastępczym stanowiskiem obsługi.

4. Opracowane zostały algorytmy obliczania parametrów wydajnościowych sieci kolejek ze stanowiskami obsługi reprezentującymi zagregowane podsieci z semaforami. Algorytmy te odnoszą się do sieci zamkniętych lub otwartych ze stałym lub zależnym od stanu strumieniem przybywających klientów. Pierwszy z algorytmów jest wykorzystywany również w fazie dekompozycji do obliczania parametrów stanowiska zastępczego.

5. Przydatność opracowanej przybliżonej metody analizy została zweryfikowana przez symulację. Do weryfikacji metody został użyty model odniesienia z kilkoma poziomami zagnieżdżeń podsieci kolejek z semaforami.

6. Został opracowany model kolejkowy mechanizmu sterowania przepływem za pomocą okna dla trzech pierwszych warstw interfejsu komunikacyjnego według ISO objętych działaniem protokołu X.25. Model ten obejmuje stacje semaforów ograniczających liczbę klientów w podsieciach do wartości określonej rozmiarem okna danej warstwy.

7. Sporządzony model kolejkowy został zanalizowany za pomocą opracowanej metody. Uzyskane wyniki zostały następnie porównane z wynikami otrzymanymi z symulacji tego samego modelu.

Uzyskane wyniki prowadzą do sformułowania wielu istotnych spostrzeżeń i wniosków.

1. Użycie semafora powoduje ograniczenie co do maksymalnej liczby klientów, jaka może przebywać w kontrolowanym przez semafor fragmencie (podsieci) sieci kolejek. Liczbę tę wyznacza tzw. rozmiar semafora. Każdy klient przybywający do podsieci o limitowanej populacji jest zatrzymywany w stacji semafora, jeśli liczba klientów aktualnie przebywających w podsieci osiągnie wspomniane ograniczenie. Przetrzymany klient może wejść do podsieci, gdy inny w niej przebywający opuści

ją. Fakt ten jest sygnalizowany w momencie opuszczania podsieci generowaniem dodatkowego klienta, który podąża do semafora celem zwolnienia go. Ten dodatkowy klient może być przekazany do semafora natychmiast bądź odbyć drogę przez pewien fragment sieci kolejek. W tym drugim przypadku czas zwolnienia semafora jest niezerowy i może być dowolnie kształtowany topologią i parametrami sieci. Tak zaproponowany model kolejkowy szczególnie dobrze ujmuje rzeczywisty mechanizm sterowania przepływem bloków danych stosowany w wielu protokołach komunikacyjnych, w tym między innymi w jednym z najbardziej popularnych pocztowych protokołów X.25. Nie mniej jednak, co należy mocno podkreślić, jego zastosowanie może być z powodzeniem rozciągnięte na inne systemy z mechanizmami synchronizacji i to nie tylko systemy komputerowe.

2. Metoda analizy modeli kolejkowych zawierających semafory oparta jest na hierarchicznej dekompozycji i agregacji. Każdą podsieć kolejek z kontrolującym ją semaforem izoluje się od reszty sieci i oblicza dla niej globalne parametry wydajnościowe, głównie czas reakcji $\mathcal{R}_S(n)$ oraz przepustowość $\gamma_S(n)$. Następnie dokonuje się agregacji zastępując w sieci kolejek wyizolowane podsieci z semaforami zastępczymi stanowiskami o wyliczonych globalnych parametrach obsługi. W ten sposób w otrzymanym modelu kolejkowym zachowuje się tylko "wizję globalną" podsieci z semaforem, reprezentowaną przez jej stanowisko zastępcze. Przez resztę sieci kolejek jest ono jednak widziane jako stanowisko, w którym niejako nie zachodzi prawo Little'a. Oznacza to, że iloczyn $\mathcal{R}_S(n) \cdot \gamma_S(n) \neq n$ jest równy liczbie klientów głównych, przemierzających podsieć, gdy tymczasem ich faktyczna liczba n w podsieci jest sumą klientów głównych jak i klientów dodatkowych, generowanych celem zwolnienia semafora.

3. W literaturze nie istnieją metody rozwiązywania takich sieci kolejek. Niniejsza praca zawiera propozycję takiej metody. Jest ona oparta na modyfikacji znanych metod dokładnych nie dopuszczających stanowisk o podanych wyżej właściwościach. W modelu kolejkowym typu zamkniętego posłużono się znanym iteracyjnym algorytmem *MVA*. Stanowiska reprezentujące podsieci z semaforami zostały potraktowane jako stanowiska o szczególnej obsłudze zależnej od stanu. Dla takiego stanowiska niezależnie od innych stanowisk modelu w trakcie każdego kroku iteracji wyznacza się rozkład całkowitej liczby klientów przebywających w reprezentowanej przez niego podsieci. Wyznaczony rozkład pozwala określić zachowanie się stanowiska jako elementu całej sieci. Ograniczeniem metody iteracyjnej dla modeli kolejkowych typu zamkniętego jest to, że wyliczona przepustowość całej sieci nie może

być większa od przepustowości $\gamma(n)$ któregokolwiek ze stanowisk zastępczych. Na marginesie należy zauważyć, że jeśli nie uwzględniać czasu zwolnienia semafora, to powyższe ograniczenie nigdy nie będzie miało miejsca, a opracowana metoda w ramach algorytmu *MVA* w sposób naturalny staje się metodą dokładną. Dla modelu kolejkowego typu otwartego zostały wyprowadzone nowe wzory wynikające z otrzymanych prawdopodobieństw stanu stacjonarnego systemu. Pokazano, że dla modeli ze strumieniem przybywających klientów zależnym od stanu sieci kolejek miary wydajności otrzymuje się z rozwiązania modeli cząstkowych typu zamkniętego i otwartego.

4. Zaproponowana metoda analizy sieci kolejek z semaforami została zweryfikowana na drodze symulacji poprzez porównanie wyników dla wielu wariantów parametrów modelu odniesienia reprezentującego warstwową architekturę interfejsu komunikacyjnego. Otrzymane rezultaty są satysfakcjonujące. Dla obciążeń modelu nie przekraczających 0.85 względny błąd czasu reakcji nie przekracza 5%. Praktyka dowodzi, że przy obciążeniu powyżej 0.7 (dane z literatury) sieć komunikacyjna nie nadaje się do pracy. Oznacza to, że w interesującym nas zakresie metoda daje realne wyniki.

5. Przykładem, w którym wykorzystano opracowaną metodę, jest analiza mechanizmu sterowania przepływem pakietów i ramek dla trzech pierwszych warstw interfejsu komunikacyjnego według ISO, objętych działaniem protokołu X.25. Złożoność struktury jego modelu kolejkowego wymagała także rozwiązania problemu priorytetowej obsługi odbioru danych nad ich nadawaniem. Mimo to otrzymano rezultaty zadowalające zarówno pod względem formy (prostota formuł matematycznych), jak i wiarygodności (duża zgodność wyników z wynikami otrzymanymi z symulacji). Przeprowadzona dla tego przykładu analiza pozwala między innymi na optymalny dobór rozmiarów okien systemu, który jest kompromisem pomiędzy liczbą niezbędnych buforów stacji nadawczo-odbiorczych a szybkością transmisji wyrażoną średnim czasem niezbędnym na przesłanie jednego pakietu. Świadczy to o przydatności metody w procesie strojenia parametrów eksploatacyjnych badanego systemu komunikacyjnego.

Przedstawione w pracy modele kolejkowe badanych rzeczywistych systemów cechuje wiele upraszczających założeń w stosunku do rzeczywistości. Przyjmuje się np. poissonowskie strumienie przybywających klientów oraz wykładnicze czasy obsługi na stanowiskach. Dzięki temu jednak model daje się zanalizować choćby przybliżoną, ale szybką metodą

analitycznej analizy. Podejście takie ma na celu szybkie zawężenie zbioru poszukiwanych, możliwych do przyjęcia dla badanego systemu rozwiązań będących funkcją jego parametrów ustawianych, jak np. wymiarów okien mechanizmów sterowania przepływem. Rozwiązania te można następnie zanalizować dokładnie, stosując kosztowne i czasochłonne pomiary lub metody symulacyjne. Z drugiej strony praktyka dowodzi, że rozkłady wykładnicze lub rozkłady będące ich złożeniami częstokroć dość dobrze aproksymują rzeczywistość i tym samym otrzymane wyniki analityczne są w pełni satysfakcjonujące.

Dalsze prace nad analizą modeli kolejkowych zawierających semafory powinny iść w kierunku uogólnienia metody na wiele klas klientów oraz rozszerzenia jej stosowalności poza zakres istniejących ograniczeń. Pozwoli to przymierzyć się do analizy złożonych wielowęzłowych sieci komunikacyjnych, w których każda z klas klientów reprezentowałaby jeden z badanych kierunków przesłania. Pierwsze kroki w tej dziedzinie zostały już podjęte. W [WILK90a] jest zaprezentowana idea rozwiązania wieloklasowych modeli z semaforami. Praca [WILK90b] zawiera formalny opis metody analizy dla sieci kolejkowych typu otwartego i zamkniętego. Nadal obowiązują ją takie same ograniczenia, jak metody prezentowanej w niniejszej pracy. W przypadku nierozróżniania klas klientów wyprowadzone dla niej zależności upraszczają się do znanych postaci zaprezentowanych w tej pracy.

Literatura

[AGRA83] S.C.Agrawal, J.P.Buzen, "The Aggregate Server Method for Analyzing Serialization Delays in Computer Systems", *ACM Trans.Comput.Syst.*, vol.1, no.2, 1983, pp.116-143.

[AGRA84] S.C.Agrawal, J.P.Buzen, A.W.Shum, "Response Time Preservation: A General Technique for Developing Approximate Algorithms for Queueing Networks", *Performance Eval.Rev.*, vol.12, no.3, 1984, pp.63-77.

[AGRE85] J.R.Agre, S.K.Tripathi, "Approximate Solution to Multichain Queueing Networks with State Dependent Service Rates", *Performance Eval.*, vol.5, no.1, 1985, pp.45-65.

[AJMO87] M.Ajmone Marsan, G.Balbo, G.Chiola, G.Conte, "Modeling the Software Architecture of a prototype Parallel Machine", *Proc. ACM/Sigmetrics Conf. on Measurement and Modeling of Computer Systems*, Banff, Alberta, Canada, May 1987.

[AKYI88] I.F.Akyildiz, A.Sieber, "Approximate Analysis of Load Dependent General Queueing Networks", *IEEE Trans.Software Eng.*, vol.14, no.11, 1988, pp.1537-1545.

[BALB79] G.Balbo, "Approximate Solutions of Queueing Network Models of Computer Systems", *Ph.D.Thesis*, Purdue University, 1979.

[BALB85] G.Balbo, S.C.Bruel, S.Ghanta, "Modeling Priority Schemes", *Proc. ACM/Sigmetrics Conf. on Measurement and Modeling of Computer Systems*, Austin, August 1985.

[BALB86] G.Balbo, S.C.Bruel, S.Ghanta, "Combining Queueing Network and Generalized Stochastic Petri Net Model for the Analysis of Some Software Blocking Phenomena", *IEEE Trans.Software Eng.*, vol.12, no.4, 1986, pp.561-576.

[BALB88] G.Balbo, S.C.Bruel, S.Ghanta, "Combining Queueing Networks and Generalized Stochastic Petri Nets for the Solution of Complex Models of System Behaviour", *IEEE Trans.Comp.*, vol.37, no.10, 1988, pp.1251-1268.

[BALS82] S.Balsamo, G.Iazeolla, "An extension of Norton's Theorem for queueing networks", *IEEE Trans.Software Eng.*, vol.8, no.4, 1982, pp.298-305.

[BASK75] F.Basket, K.M.Chandy, R.Muntz, F.Palacios, "Open, Closed and Mixed Networks of Queues with Different Classes of Customers", *J. Assoc.Comput.Mach.*, vol.22, no.2, 1975, pp.248-260.

[BENK86] A.Benkiran, H.N.Thuc, "Utilisation du protocole X.25 paquet dans l'environnement de réseaux locaux et d'interconnection", *Res.Rep. 166*, Lab. MASI, Univ. Paris 6, 1986.

[BORO76] A.A.Borovkov, *Stochastic Processes in Queueing Theory*, Springer-Verlag, New York, 1976.

[BOSC88] M.Bosch, O.Gihr, W.Kiesel, "Modelling and Performance Comparison of two Application Layer Protocols for Manufacturing Automation", *International Teletraffic Congress*, Torino, 1988.

[BRAM83] G.W.Brams, *Réseaux de Petri: Théorie et Analyse*, Masson, Paris, 1983.

[BRAN79] A.Brandwajn, "An Iterative Solution of Two-Dimensional Brith and Death Processes", *Oper.Res.*, vol.27, no.3, 1979, pp.595-605.

[BRAN84] A.Brandwajn, W.M.McCormac, "Efficent Approximation for Models of Multiprogramming with Shared Domains", *Performance Eval.Rev.*, vol.12, no.3, 1984, pp.186-194.

[BUI89] B.N.Bui, "Réseaux de files d'attente ouverts multiclasses à contrainte partagée", *Proc. Deuzième Conference sur l'Ingenierie de Performance et la Technologie de l'Information*, Nice, France, Mai 1989.

[CCITT84] "Réseau de communications de données; interfaces; Recomandations X.20 à X.32", *CCITT*, Livre Rouge, Tome VIII - Fascicule VIII.3, 1984.

[CHAN74] A.Chang, S.S.Lavenberg, "Work-rates in Closed Queueing Systems", *Oper. Res.*, vol.22, no.4, 1984.

[CHAN75a] K.M.Chandy, U.Herzog, L.S.Woo, "Parametric Analysis of Queueing Networks", *IBM J.Res.& Dev.*, vol.19, no.1, 1975, pp.36-42.

[CHAN75b] K.M.Chandy, U.Herzog, L.S.Woo, "Approximate Analysis of General Queueing Networks", *IBM J.Res.& Dev.*, vol.19, no.1, 1975, pp.43-49.

[CHAN78] K.M.Chandy, C.H.Sauer, "Approximate Methods for Analyzing Queueing Network Models of Computing Systems", *Computer Survey*, vol.10, no.3, 1978, pp.281-317.

[CHAN80] K.M.Chandy, C.H.Sauer, "Computational Algorithms for Product Form Queueing Networks", *Commun.ACM*, vol.23, no.10, 1980, pp.573-583.

[CHEN88] W-Y.Cheng, J.W.S.Liu, "Performance of ARQ Schemes in Token Ring Networks", *IEEE Trans.Comp.*, vol.37, no.7, 1988, pp.826-834.

[CHRE84] P.Chrétienne, "Exécutions controlées des réseaux de Petri temporisés", *Technique et Science Informatique*, vol.3, no.1, 1984, pp.23-31.

[COUR75] P.J.Courtois, "Decomposability, Instabilities and Saturation in Multiprogramming Systems", *Commun.ACM*, vol.18, 1975, pp.371-377.

[COUR77] P.J.Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, 1977.

[CZAC82] T.Czachórski, M.Kowalówka, A.Wilk, "A Software for Computer System Performance Analysis - one more effort", *7-th Winterschool on Operating Systems*, Visegrád, Hungary, 1982.

[CZAC87] T.Czachórski, M.Kowalówka, Z.Szczerbiński, A.Wilk, W.Wołowiec, "Modelowanie systemów komputerowych za pomocą AMOK-u", *Podstawy Sterowania*, vol.17, no.3-4, 1987, pp.103-127.

[DALL88] Y.Dallery, "Approximate Analysis of General Open Queueing Networks with Restricted Capacity", *Draft paper*, Lab. MASI, Univ. Paris 6, 1988.

[DAO83] M.Dao, "Etude du contrôle de fleux par fenêtre dans un réseau à cirquits virtuels", *Thèse de l'Université*, Centre d'Orsay, Univ. Paris-Sud, 1983.

[DIJK65] E.W.Dijkstra, *Cooperating sequential process*, Technological University, Eindhoven, The Netherlands, 1965.

[DRIX82] P.Drix, "Algorithme de réseaux de files d'attente fermés ECMP et heuristiques associées", *Thèse de docteur ingenieur*, Univ. Paris 6, 1982.

[DUDA87] A.Duda, T.Czachórski, "Performance Evaluation of Fork and Join Synchronization Primitives", *Acta Informatica*, vol.24, 1987, pp.525-535.

[FDID84a] S.Fdida, "Étude de systèmes à partage de ressources par réseaux de files d'attente. Application à làrchitecture multiprocesseur SM90.", *Thèse de 3-ème cycle*, Lab. MASI, Univ. Paris 6, 1984.

[FDID84b] S.Fdida, D.Mailles, G.Pujolle, "Queueing Systems with Resource Sharing.", *Int. Workshop on Modeling and Performance Evaluation of Parallel Systems*, Grenoble, December 1984.

[FDID84c] S.Fdida, G.Pujolle, "Etude de réseaux de files d'attente avec des ressources partagées", *Res.Rep. 48*, Lab. MASI, Univ. Paris 6, 1984.

[FDID86] S.Fdida, G.Pujolle, D.Mailles, "Réseaux de files d'attente avec sémaphores", *Res.Rep. 96*, Lab. MASI, Univ. Paris 6, 1985, also: *RAIRO Tech.Sciences Inform.*, vol.5, no.3, 1986, pp.187-196.

[FDID87] S.Fdida, A.Wilk, H.Perros, "Semaphore Queues: Modelling Window Flow Control Mechanism", *Res.Rep. 181*, Lab. MASI, Univ. Paris 6, 1987.

[FDID90] S.Fdida, H.Perros, A.R.Wilk, "Semaphore Queues: Modeling Multi-Layered Window Flow Control Mechanism", *IEEE Trans.Commun.*, vol.38, no.3, 1990, pp.309-317.

[FDID88] S.Fdida, U.Korner, H.G.Perros, G.Sharpio, "End to End Delays in a Catenet Environment", *Res.Rep. 182*, Lab. MASI, Univ. Paris 6, 1987, also: *Performance Eval.Rev.*, vol.15, no.3-4, 1988, pp.20-28.

[FREU83] D.J.Freund, J.N.Bexfield, "A New Aggregation Approximation Procedure for Solving Closed Qucueing Networks with Simultaneous Resource Possesion", *Proc. ACM/Sigmetrics Conf. on Measurement and Modeling of Computer Systems*, Minneapolis, August 1983, pp.214-224.

[GANZ88] A.Ganz, I.Chlamtac, "Queueing Analysis of Finite Buffer Token Networks", *Proc. ACM/Sigmetrics Conf. on Measurement and Modeling of Computer Systems*, Santa Fe, Mew Mexico, May 1988.

[GERL80] M.Gerla, L.Kleinrock, "Flow Control: A Comparative Survey", *IEEE Trans. Commun.*, vol.28, no.4, 1980, pp.553-574.

[GIHR85] O.Gihr, P.J.Kuehn, "Comparison of Communication Services with Connection-Oriented and Conectionnless Data Transmission", *Proc. International Seminar on Computer Networking and Performance Evaluation*, Tokyo, September 1985.

[GOTO83] K.Goto, Y.Takahashi, A.Hasegawa, "An Approximate Analysis of Controlled Tandem Queues", *Proc. Int. Sem. on Modelling and Performance Evaluation Methodology*, vol.1, Paris, January 1983, pp.43-52.

[HAYE84] J.F.Hayes, *Modeling and Analysis of Computer Communications Networks*, Plenum Press, New York, 1984.

[HEID84] P.Heidelberger, S.S.Lavenberg, "Computer Performance Evaluation Methodology", *IEEE Trans.Comp.*, vol.33, no.12, 1984, pp.1195-1220.

[HERZ75] U.Herzog, L.Woo, K.M.Chandy, "Solution of Queueing Problems by a Recursive Technique", *IBM J.Res.& Develop.*, vol.19, no.5, 1975, pp.295-300.

[HOLL85] M.A.Hollyday, M.K.Vernon, "A Generalized Timed Petri Net Model for Performance Analysis", *Proc. Int. Workshop Timed Petri Nets*, Torino, Italy, July 1985, IEEE Computer Society Press.

[ISO82] "Data Processing - Open System Interconnection - Basic Reference Model", Draft Proposal ISO/DP 7498, International Organization for Standardization, 1982.

[JACO82] P.A.Jacobson, E.D.Lazowska, "Analyzing Queueing Networks with Simultaneous Resource Possesion", *Commun.ACM*, vol.25, no.2, 1982, pp.142-151.

[JACO83] P.A.Jacobson, E.D.Lazowska, "A Reduction Technique for Evaluating Queueing Networks with Serialization Delays", *Proc. Performance 83*, University of Maryland, College Park, North-Holland, 1983, pp.45-49.

[KLEI75] L.Klienrock, *Queueing Systems. I: Theory. II: Computer Applications*, John Wiley Interscience, New York, 1975.

[KOBA77] H.Kobayashi, A.G.Konheim, "Queueing Models for Computer Communications System Analysis", *IEEE Trans.Commun.*, vol.25, no.1, 1977, pp.2-29.

[KOER87] U.Koerner, S.Fdida, H.G.Perros, G.Shapiro, "End to End Delays in a Catenet Environment", *Proc. Third Int. Conference on Data Communnication Systems and their Performance*, Rio de Janeiro, June 1987, pp.453-464.

[KRIT86] P.S.Kritzinger, "A Performance Model of the OSI Communication Architecture", *IEEE Trans.Commun.*, vol.34, no.6, 1986, pp.554-563.

[KRZE85] A.E.Krzesinski, P.Teunissen, "Multiclass Queueing Networks with Population Constrained Subnetworks", *Performance Eval.Rev.*, vol.13, no.2, 1985, pp.128-139.

[KUEH69] P.J.Kuehn, "Approximate Analysis of General Queueing Networks by Decomposition", *IEEE Trans.Commun.*, vol.27, no.1, 1979, pp.113-126.

[LABE80] J.Labetoulle, G.Pujolle, "Isolation Method in a Network of Queues", *IEEE Trans.Soft.Eng.*, vol.6, no.4, 1980, pp.373-381.

[LABE81] J.Labetoulle, G.Pujolle, "A Study of Flows Through Virtual Circuits Computer Networks", *Computer Networks*, no.5, 1981, pp.119-126.

[LAM77] S.S.Lam, "Queueing Networks with Population Size Constraints", *IBM J.Res. Develop.*, vol.21, no.4, 1977, pp.370-378.

[LAM81] S.S.Lam, Y.I.Lien, "Modeling and Analysis of Flow-Controlled Packet Switching Networks", *Proc. 7-th Data Communications Symp.*, Mexico City, October 1981, pp.98-107.

[LAM82] S.S.Lam, J.W.Wong, "Queueing Network Models of Packet Switching Networks. Part 2: Networks with Population Size Constraints", *Performance Eval.*, vol.2, no.2, 1982, pp.161-180.

[LAM87] S.S.Lam, C.-T.Hsieh, "Modeling, Analysis and Optimal Routing of Flow-Controlled Communication Networks", *Tech.Rep. TR-87-24*, Dept. of Computer Sciences, Univ. of Texas at Austin, 1987.

[LAVE78] S.S.Lavenberg, "Stability and Maximum Departure Rate of Certain Open Queueing Networks Having Finite Capacity Constraints", *RAIRO Informatique/Computer Science*, vol.12, no.4, 1978, pp.353-370.

[LAZO82] E.D.Lazowska, J.Zahorjan, "Multiple Class Memeory Constrained Queueing Networks", *ACM/Sigmetrics Symp.on Measurement, Modeling and Evaluation of Comp. Systems*, Seattle, 1982.

[LAZO84] E.D.Lazowska, J.Zahorian, G.S.Graham, K.C.Sevcik, *Quantitative System Performance - Computer System Analysis Using Queueing Network Models*, Prentice-Hall, Englewood Cliffs, 1984.

[LIM87] J-T.Lim, S.M.Meerkov, "Theory of Markovian Acces to Collision Channels", *IEEE Trans.Commun.*, vol.35, no.12, 1987, pp.1278-1288.

[LITT61] J.D.C.Little, "A Proof of the Queueing Formula $L = \lambda W$", *Oper.Res.*, vol.9, no.3, 1961, pp.383-387.

[MAGO89] J.Magott, "Sieci Petriego w ocenie wydajności systemów komputerowych", Praca Naukowe Instytutu Cybernetyki Technicznej Politechniki Wrocławskiej, Seria: monografie, no.80, Wrocław 1989.

[MAIL87] D.Mailles, "Files d'attente descriptive pour la modélisation de la synchronisation dans les systèmes informatiques", *Thèse d'Etat*, Univ. Paris 6, 1987.

[MOLL82] M.K.Molly, "Performance Analysis Using Stochastic Petri Nets", *IEEE Trans.Comp.*, vol.31, no.9, 1982, pp.913-917.

[NATK85] S.Natkin, "Réseaux de Pétri Stochastiques. Théorie et applications", *Thèse d'Etat*, Univ. Paris 6, 1985.

[PUJO78] G.Pujolle, "Modélisation et évaluation de performances des réseaux à commutation de paquets", *Thèse d'Etat*, Univ. Paris-Sud, 1978.

[PUJO84] G.Pujolle, S.Fdida, "General Resource Sharing Systems", *Res.Rep 48*, Lab. MASI, Univ. Paris 6, 1984.

[REIS75] M.Reiser, H.Kobayashi, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms", *IBM J.Res.Develop.*, vol.19, no.3, 1975, pp.283-294.

[REIS79] M.Reiser, "A Queueing Network Analysis of Computer Communication Networks with Window Flow Control", *IEEE Trans.Commun.*, vol.27, no.8, 1979, pp.1199-1209.

[REIS80] M.Reiser, S.S.Lavenberg, "Mean-Value Analysis of Closed Multichain Queueing Networks", *J.ACM*, vol.27, no.2, 1980, pp.313-322.

[REIS81a] M.Reiser, "Mean Value Analysis and Convolutional Method for Queue-Dependent Servers in Closed Queueing Networks", *Performance Eval.*, vol.1, no.1, 1981, pp.7-18.

[REIS81b] M.Reiser, "Admission Delays on Virtual Routes with Window Flow Control", *Proc. Performance of Data Communication Systems and their Applications*, North Holland, 1981, pp.67-76.

[REIS82] M.Reiser, "Performance Evaluation of Data Communication Systems", *Proc. IEEE*, vol.70, 1982, pp.171-196.

[REIS85] M.Reiser, "Communication System Models Embredded in the OSI-Reference Model. A Survey", *Proc. International Seminar on Computer Networking and Performance Evaluation*, Tokyo, September 1985.

[SAUE78] C.H.Sauer, E.A.MacNair, "Simultaneous Resource Possesion in Queueing Models of Computers", *Performance Eval.Rev.*, vol.7, no.1-2, 1978, pp.41-52.

[SAUE79] C.H.Sauer, K.M.Chandy, "Approximate Solution of Queueing Models of Computer Systems", *IBM Res.Rep. RC-7785 (#33680)*, IBM Res. Center, Yorktown Heights, 1979.

[SAUE81a] C.H.Sauer, K.M.Chandy, *Computer Systems Performance Modeling*, Prentice-Hall, Englewood Cliffs, 1981.

[SAUE81b] C.H.Sauer, "Approximate Solution of Queueing Networks with Simultaneous Resource Possesion", *IBM J.Res.Develop.*, vol.25, no.6, 1981, pp.894-903.

[SAUE83] C.H.Sauer, "Computational Algorithms for State-Dependent Queueing Networks", *ACM Trans.Comput.Syst.*, vol.1, no.1, 1983, pp.67-92.

[SCHM84] W.Schmitt, "On a Decomposition of Markovian Priority Queues and Their Application to the Analysis of Closed Queueing Networks", *Proc. Performance 84*, North Holland, 1984, pp.393-407.

[SCHW82] M.Schwartz, "Performance Analysis of the SNA Virtual Route Packing Control", *IEEE Trans.Commun.*, vol.30, 1982, pp.172-184.

[SERE85] D.Seret, "Evaluation de Performances des Protocoles des Communication – application aux reseaux locaux et aux transmissions par satelite", *Thèse d'Etat*, Univ. Paris 6, 1985.

[SMIT80] C.U.Smith, J.C.Browne, "Aspects of Software Design Analysis: Concurency and Blocking", *Perf.Eval.Rev.*, vol.9, no.2, 1980, pp.245-253.

[THOM83] A.Thomasian, P.Bay, "Analysis Techniques for Queueing Network Models of Multicomputer Systems with Shared Resources", *Comput.Performance*, vol.4, no.3, 1983, pp.151-166.

[THOM84] A.Thomasian, P.Bay, "Analysis of Queueing Network Models with Population Size Constraints and Delayed Blocked Customers", *Performance Eval.Rev.*, vol.12, no.3, 1984, pp.202-215.

[TUCC82] S.Tucci, E.A.MacNair, "Implementation of Mean-Value Analysis for Open, Closed and Mixed Queueing Networks", *Computer Performance Theory*, vol.3, no.4, 1982, pp.233-239.

[VARG83] G.Varghese, W.Chou, A.A.Nilson, "Queueing Delays on Virtual Cirquits Using a Sliding Window Flow Control Scheme", *Proc. ACM/Sigmetrics Conf. on Measurement and Modeling of Comp. Systems*, Minneapolis, August 1983, pp.275-281.

[WILK86] A.R.Wilk, S.Fdida, "Gestion des ressources par sémaphores: modélisation des mechanismes de contrôle de flux du protocole X.25", *Res.Rep. 124*, Lab. MASI, Univ. Paris 6, 1986.

[WILK87] A.R.Wilk, "Jednoklasowe sieci kolejek z podsieciami o limitowanej populacji klientów", *Podstawy Sterowania*, vol.17, no.3-4, 1987, pp.129-147.

[WILK88] A.R.Wilk, "Single Class Open and Closed Queueing Networks with Population Constrained Subnetworks", *Proc. 2nd International Seminar on Modelling and Performance Evaluation of Distributed Computer Systems*, Wendisch Rietz, GDR, November 1988, pp.15-34.

[WILK90a] A.R.Wilk, "Multiclass Queueing Networks with Population Size Constraints", *Proc. Second Symp. on Queueing Theory and Related Topics*, Karpacz, January 1990.

[WILK90b] A.R.Wilk, "Multiple Class Open and Closed Queueing Networks with Population Constrained Subnetworks", w druku w *Archiwum Informatyki Teoretycznej i Stosowanej*, 1990.

[WONG82] J.W.Wong, S.S.Lam, "Queueing Network Models of Packet Switching Networks. Part 1: Open Networks", *Performance Eval.*, no.2, 1982, pp.9-21.

[WONG89] F.Wong, J.R.B. de Marca, "Fairness in Window Flow Controlled Computer Networks", *IEEE Trans.Commun.*, vol.37, no.5, 1989, pp.475-480.

[ZAHO84] J.Zahorjan, "Incorporating Load Dependent Servers in Approximate Mean Value Analysis", *Performance Eval. Rev.*, vol. 12, no.3, 1984, pp.63-77.

MODELE KOLEJKOWE Z SEMAFORAMI. MODELOWANIE PROTOKOŁU X.25

Streszczenie

Przedmiotem niniejszej pracy jest próba obliczenia wydajności mechanizmu sterowania przepływem pakietów i ramek za pomocą okna, stosowanego w pakietowych sieciach komunikacyjnych pracujących w oparciu o protokół X.25. Do rozwiązania problemu zostały użyte modele kolejkowe w postaci sieci stanowisk obsługi, do których wprowadzono dodatkowe elementy – semafory – synchronizujące ruch przemieszczających się w sieci klientów.

Praca przedstawia funkcjonowanie i analizę semaforów w modelu kolejkowym. Opracowana została przybliżona metoda obliczeniowa parametrów wydajnościowych dla sieci kolejek z semaforami typu zamkniętego i otwartego ze stałym lub zmiennym strumieniem przybywających klientów. Dokładność metody została zweryfikowana przez porównanie jej wyników z wynikami otrzymanymi z symulacji.

Opracowana metoda została wykorzystana do obliczenia wydajności mechanizmów sterowania przepływem pakietów i ramek pomiędzy dwoma przykładowymi węzłami sieci X.25. Otrzymane wyniki pozwalają dobrać właściwy rozmiar okien tych mechanizmów.

QUEUEING NETWORK MODELS WITH SEMAPHORES. X.25 PROTOCOL MODELLING

Abstrct

This study is inspired by an attempt to compute performance measures of the window flow control mechanism used in packet communication networks operating with the X.25 protocol. To solve this problem the queueing network models are used with additional elements – semaphores – synchronizing the customer circulation in the network.

The study demonstrates the conception and analysis of the semaphores in the queueing network models. The developed approximate method permits to obtain performance measures for the open or closed queueing networks with state-dependent arrival rates. The approximation is validated against simulation data.

The proposed technique is applied to compute performance measures of the flow control mechanisms used for packet and frame flow control between two exemplary nodes of X.25 network. The obtained results permit to choose a correct window size of these mechanisms.

МОДЕЛИ СИСТЕМ МАССОВОГО ОБСЛУЖИБАНИЯ С СЕМАФОРАМИ. МОДЕЛИРОВАНИЕ ПРОТОКОЛА X.25

Резюме

Предметом работы являетсья попытка расчёта производительности механизма управления потоком пакетов и рамок при помощи окна, применяемого в сетьях коммутации пакетов работающих с протоколом X.25. Дла решения проблемы использовано модели систем массового обслуживания в виде сети постов обслуживания в которые введено добавочные элементы – семафоры – которые синхронизируют движение перемещающихся в сети заявок.

В работе представлено функционирование и анализ семафоров являющихся элементом системы массового обслуживаня. Разработано приближённый метод расчёта параметров характеризующих производительность сетей очередей с семафорами, закрытого и открытого типа, с постоянным или зависимым потоком приходящих заявок. Точность метода проверено посредством сравненя полученных результатов с результатами симуляции

Разработанный метод использовано для расчёта производительности механизмов управления потоком пакетов и рамок между двумья избранными узлами сети типа X.25. Полученные результаты позволяют подобрать соответствующий резмер окон этих механизмов.

BIBLIOTEKA GŁÓWNA
Politechniki Śląskiej
P. 4193/91/17