POLITECHNIKA BIAŁOSTOCKA

SKRYPTY

Walenty Oniszczuk

METODY MODELOWANIA

WYDAWNICTWA POLITECHNIKI BIAŁOSTOCKIEJ

Recenzent:

prof. dr hab. Jacek Wojtowicz

Opracowanie redakcyjne Jadwiga Żukowska
Prace kreślarskie Tadeusz Malewicz

ISBN 83-86272-18-X

Drukowano w Zakładzie Poligraficznym Politechniki Białostockiej. Nakład 200 egz. Format B-5. Ark. wyd. 8,0 Ark. druk. 12,0. Oddano do druku w marcu 1995. Druk ukończono w marcu 1995. Zam. 16/95.

Spis treści

Rozdział 1

Elementy teorii strumienia wejściowego, procesy Markowa, proces narodzin i śmierci

1.1 Wiadomości wstępne

Modelowanie analityczne (matematyczne), tak często stosowane do ana lizy wydajności istniejących, często złożonych, systemów komputerowych, czy systemów do transmisji danych, oraz innych systemów, jak i do badania wydajności różnorodnych projektowanych systemów, szeroko wykorzystuje teorię kolejek (masowej obsługi). Teoria kolejek z kolei, wywodzi się z teorii procesów stochastycznych, będących jednym z działów teorii prawdopodobieństwa.

1.1.1 Wybrane pojęcia z teorii prawdopodobieństwa

W teorii kolejek, jak i w przedmiocie modelowania analitycznego kluczowym pojęciem jest pojęcie zmiennej losowej. Zmienne losowe dzielimy na : dyskretne (jak np. liczba zdarzeń (zgłoszeń) w kolejce, systemie itp.) i ciągłe (np. losowy czas obsługi zgłoszeń w kolejce, czas pobytu zgłoszenia w systemie itp.).

Rozkład $P = P_k$, dla $k = 0 , 1 , 2 , \cdots$ nieujemnej dyskretnej zmiennej losowej v opisany jest, przez prawdopodobieństwo przyjęcia przez nią wartości k, w sposób następujący:

$$P(v = k) = P_k \text{ , gdzie } k = 0 , 1 , 2 , \cdots , \text{ oraz } \sum_{k=0}^{\infty} P_k = 1 \quad (1.1)$$

Wartość oczekiwana (średnia) nieujemnej dyskretnej zmiennej losowej v jest definiowana jako:

$$E_v = \overline{v} = \sum_{k=0}^{\infty} k\ P_k \quad (1.2)$$

a wariancja:

$$V_v = \sigma^2 = E(v - E_v)^2 = \sum_{k=0}^{\infty}(k - E_k)^2 P_k \quad (1.3)$$

gdzie σ - odchylenie standardowe.

Jednym z najczęściej stosowanych rozkładów zmiennych losowych dyskretnych w teorii masowej obsługi jest rozkład Poissona. Mówimy, że zmienna losowa v ma rozkład Poissona z parametrem $\lambda > 0$, jeżeli:

$$P(v = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{dla } k = 0 , 1 , 2 , \cdots \quad (1.4)$$

Wartość oczekiwana i wariancja w rozkładzie Poissona wynoszą λ.

Rozkład dowolnej zmiennej losowej, oznaczmy ją przez ξ, charakteryzowany jest przez dystrybuantę F_ξ tej zmiennej losowej:

$$F_\xi(t) = P(\xi < t) \quad (1.5)$$

Zmienna losowa ξ nazywana jest ciągłą, jeżeli istnieje taka funkcja nieujemna $f(x)$, dla dowolnych a i b, że

$$P\{a \le \xi \le b\} = \int_a^b f(x)dx \quad (1.6)$$

zaś wartość oczekiwana i wariancja zmiennej losowej ξ równe są:

$$E_\xi = \int_{-\infty}^{\infty} x dF_\xi(x) = m_{F_\xi} \quad (1.7)$$

$$V_\xi = \int_{-\infty}^{\infty} (x - E_\xi)^2 dF_\xi(x) = \sigma^2_{F_\xi} \quad (1.8)$$

Dla nieujemnych zmiennych losowych (szeroko stosowanych w modelowaniu analitycznym) wartość oczekiwaną oblicza się według następującego wzoru:

$$E_\xi = \int_0^{\infty} [1 - F_\xi(x)]\, dx \quad (1.9)$$

Dla ciągłych zmiennych losowych posiadających jednostajnie ciągłą dystrybuantę F_ξ, dystrybuantę tę można przedstawić w postaci:

$$F_\xi(t) = \int_{-\infty}^{t} f_\xi(x)dx \quad (1.10)$$

gdzie gęstość $f_\xi(x)$ jest nieujemna oraz:

$$\int_{-\infty}^{\infty} f_\xi(x)dx = 1 \quad (1.11)$$

Wzory na wartość oczekiwaną i wariancję przyjmą w tym przypadku następującą postać:

$$E_\xi = \int_{-\infty}^{\infty} x f_\xi(x)dx = m_{F_\xi} \quad (1.12)$$

$$V_\xi = \int_{-\infty}^{\infty} (x - E_\xi)^2 f_\xi(x)dx = \sigma^2_{F_\xi} \quad (1.13)$$

i tam, gdzie nie będzie to powodowało dwuznaczności, F_ξ i f_ξ oznaczmy przez F i f.

Rozkład wykładniczy: ciągła zmienna losowa ξ ma rozkład wykładniczy z parametrem $\lambda > 0$, gdy jej dystrybuanta F ma postać:

$$P(\xi < t) = F(t) = \begin{cases} 0 & \text{dla } t \le 0 \\ 1 - e^{-\lambda t} & \text{dla } t > 0 \end{cases} \quad (1.14)$$

a odpowiadająca jej gęstość dana jest wzorem:

$$f(t) = \lambda\, e^{-\lambda\, t} \qquad \text{dla } t \geq 0 \tag{1.15}$$

więc wartość oczekiwana i wariancja zmiennej losowej ξ odpowiednio wynoszą:

$$E_\xi = \lambda^{-1} \qquad \text{oraz } V_\xi = \lambda^{-2} \tag{1.16}$$

Rozkład ten charakteryzuje się tzw. brakiem pamięci. Pewnym rozkładem, blisko powiązanym z rozkładem wykładniczym, jest rozkład Erlanga. Zmienną losową ξ o rozkładzie Erlanga stopnia k, z parametrem λ, można przedstawić jako sumę wzajemnie niezależnych zmiennych losowych ξ_i, dla $i = 1, 2, \cdots, k$, o jednakowych rozkładach wykładniczych, z paramertem λ:

$$\xi = \xi_1 + \xi_2 + \cdots + \xi_k \tag{1.17}$$

1.2 Procesy stochastyczne

Proces stochastyczny traktowany jest jako abstrakcja matematyczna pewnych empirycznych układów, zmieniających się losowo w czasie. Z punktu widzenia teorii prawdopodobieństwa, proces stochastyczny definiowany jest jako przeliczalny lub nieprzeliczalny zbiór zmiennych losowych X_t, gdzie t jest parametrem rzeczywistym indeksującym te zmienne.

Gdy t przbiega zbiór liczb naturalnych, to proces oznaczamy przez $\{X_n, n \in N_1\}$, gdzie N_1 - zbiór liczb naturalnych, i wtedy proces stochastyczny nazywany jest ciągiem losowym. Jeżeli t jest liczbą rzeczywistą ze zbioru $T = \{t \geq 0\}$ i gdy parametr t interpretowany jest jako czas, to zbiór zmiennych losowych X_t tworzy właściwy proces stochastyczny, oznaczany $\{X_t, t \in T\}$.

Zbiór S możliwych wartości zmiennych losowych X_t, tworzących proces stochastyczny $\{X_t, t \in T\}$, gdzie T oznacza dowolny zbiór liczb rzeczywistych, może być skończony, przeliczalny lub nieprzeliczalny. Zbiór S nazywany jest przestrzenią fazową procesu stochastycznego.

Dla dowolnego procesu stochastycznego zdarzenie polegające na tym, że zmienna losowa X_t przyjęła wartość k, oznaczane jest jako $X_t = k$, znaczy to, że proces stochastyczny w momencie $t \in T$ jest w stanie

$k \in S$. Jeżeli zmienne losowe X_t procesu $\{X_t, t \in T\}$ przyjmują wartości 1, 2, $\cdots$, r, to przestrzenią fazową tego procesu jest $S = \{1, 2, \cdots, r\}$, przy czym liczby 1, 2, $\cdots$, r nazywane są dyskretnymi stanami fazowymi procesu (pojęcie to szeroko stosowane jest w modelowaniu analitycznym).

1.3 Teoria strumienia wejściowego

Teoria strumienia wejściowego, obejmuje między innymi, pojęcie strumienia zdarzeń, sposoby jego określania, klasyfikację strumieni oraz pojęcie równoważności strumieni.

Strumieniem zdarzeń nazywamy taki proces stochastyczny, w którym:

a) jeżeli $v(t)$ jest liczbą zdarzeń do momemtu t i w momencie $t = 0$ nie nastąpiło żadne zdarzenie, to:

$$v(0) = 0 \tag{1.18}$$

b) $v(t)$ dla każdego $t \geq 0$ przyjmuje tylko wartości całkowite nieujemne,

c) trajektorie procesu $v(t)$ nie maleją.

Niech $t_1, t_2, \cdots$ będą kolejnymi momentami pojawienia się zdarzeń, oraz $t_{k-1} \leq t_k$, dla $k \geq 1$ i dla $t_0 = 0$, to:

$$z_k = t_k - t_{k-1}, \qquad \text{dla } k \geq 1 \tag{1.19}$$

Przyjmuje się, że strumień zdarzeń jednego typu jest określony, jeżeli dla każdej liczby całkowitej $n \geq 1$ jest określony rozkład wektora losowego

$$(z_1, \cdot z_2, \cdots, z_n)$$

jeżeli $z_1, z_2, \cdots$ są niezależnymi zmiennymi losowymi, to strumień zdarzeń jest strumieniem o ograniczonych następstwach. Dla określenia takiego strumienia wystarczy podać dystrybuanty

$$F_k(t) = P\{z_k < t\} \qquad \text{dla } k \geq 1 \tag{1.20}$$

Strumień zdarzeń o ograniczonych następstwach, dla którego $F_2(t) = F_3(t) = \cdots = F(t)$ nazywamy opóźnionym strumieniem rekurencyjnym (ogólny strumień odnowy), określonym przez dystrybuanty $F_1(t)$ i $F(t)$.

Jeżeli $F_k(t) = F(t)$, dla $k \geq 1$, to mówimy o strumieniu rekurencyjnym (odnowy) określonym przez dystrybuantę $F(t)$.

Strumień rekurencyjny, dla którego $F(t) = 1 - e^{-\lambda t}$ (dystrybuanta rozkładu wykładniczego), dla $t \geq 0$, gdy λ jest stałą dodatnią, nazywamy strumieniem Poissona, z kolei λ nazywamy parametrem (intensywnością) strumienia Poissona.

Strumień $v(t)$ nazywamy strumieniem bez następstw, jeżeli proces stochastyczny $v(t)$ jest procesem o przyrostach niezależnych, tzn. zmienne losowe, przyrosty procesu, są niezależne.

Strumień $v(t)$ nazywamy stacjonarnym, jeżeli dla każdej liczby całkowitej $n \geq 1$ i dowolnego ciągu liczb nieujemnych $\tau_1, \tau_2, \cdots, \tau_n$ rozkład wektora losowego

$$\{v(c+\tau_k) - v(c) \qquad k = 1, 2, \cdots, n\} \tag{1.21}$$

nie zależy od wyboru liczby $c \geq 0$. Idea stacjonarności, związana jest blisko z koncepcją ergodyczności, odnoszącą się do problemu "pomiaru" procesu stochastycznego $X(t)$ w pojedynczej realizacji. O procesie stochastycznym $X(t)$ można powiedzieć, że jest ergodyczny, gdy jego parametry mogą być aproksymowane przez wartości otrzymane z pojedynczej realizacji $X_0(t)$ tego procesu. Proces stochastyczy $X(t)$ jest zatem ergodyczny, gdy jego uśrednione parametry na dość długim (dokładniej $\to \infty$) przedziale czasowym są równe średnim wartościom tych parametrów, otrzymanym z wielu realizacji (dokładniej, jeżeli liczba realizacji $\to \infty$) tego procesu (rys. 1).

Strumień nazywamy pojedynczym, jeżeli wykluczona jest możliwość pojawienia się w jednym momencie więcej niż jednego zdarzenia.

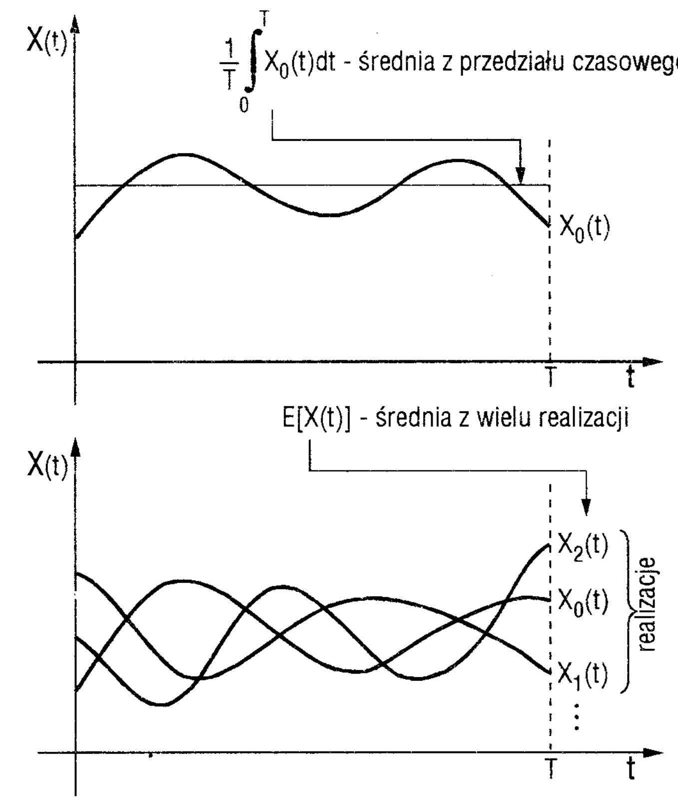
Stacjonarny, pojedynczy, strumień bez następstw nazywamy strumieniem Poissona (strumieniem najprostszym).

Podana została tutaj podstawowa klasyfikacja strumieni, która używana jest tak i w teorii masowej obsługi, jak i w przedmiocie modelowania analitycznego.

Dwa strumienie $v_1(t)$ i $v_2(t)$ nazywane są równoważnymi, jeżeli dla dowolnej liczby całkowitej $n \geq 1$ i dowolnego ciągu liczb nieujemnych $\tau_1, \tau_2, \cdots, \tau_n$ rozkłady dwóch wektorów losowych

$$\{v_1(\tau_k)\text{ , dla } k = 1, 2, \cdots, n\} \text{ i } \{v_2(\tau_k)\text{ , } k = 1, 2, \cdots, n\} \tag{1.22}$$

są identyczne.



Rys. 1 Ergodyczność

Czyli:

a) Stacjonarny i bez następstw strumień $v(t)$, scharakteryzowany jest w pełni, gdy podane jest prawdopodobieństwo k zdarzeń w czasie od 0 do t:

$$P_k(t) = P\{v(t) = k\} \text{ , dla wszystkich } k = 0, 1, \cdots \text{, oraz } t \geq 0 \tag{1.23}$$

które równe jest:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad \text{dla } k \geq 0 \text{ , } t \geq 0 \tag{1.24}$$

b) Strumień najprostszy można w pełni określić podając tylko jedną liczbę λ - intensywność strumienia.

1.3.1 Rozrzedzanie strumienia wejściowego

Operacja rozrzedzania strumieni wejściowych sprowadza się do tworzenia z zadanego strumienia innych strumieni. Operacja ta przypomina strumień wyrobów wpływających do pewnego urządzenia, które z kolei rozdziela te wyroby na różne stanowiska obsługi.

Przedstawmy, bez dokładnej analizy, kilka rezultatów i twierdzeń związanych z tą problematyką.

1) Strumień otrzymany ze strumienia rekurencyjnego za pomocą rekurencyjnej operacji rozrzedzania jest strumieniem rekurencyjnym.

2) Jeżeli wejściowy strumień jest strumieniem Poissona z intensywnością λ, to operacja rozrzedzania określona jest następująco: dowolne zdarzenie ze strumienia wejściowego jest, z prawdopodobieństwem p, odrzucane i z prawdopodobieństwem $q = 1-p$ zatrzymywane, niezależnie od pozostałych zdarzeń. Czyli po takiej operacji rozrzedzony strumień jest także strumieniem Poissona (z intensywnością $q\lambda$).

3) Jeżeli operacja rozrzedzania określona jest następująco: wyrzucamy v_1 pierwszych zdarzeń a jedno następne zostawiamy, potem wyrzucamy następne v_2 zdarzeń, a jedno następne zostawiamy, itd. to operacja rozrzedzania określona jest przez łączne rozkłady ciągu zmiennych losowych $v_1, v_2, \cdots$ i aby określić strumień rozrzedzony wystarczy podać strumień wejściowy i operację rozrzedzania. Podany niżej przykład ilustruje taką operację rozrzedzania, w której pierwsze k zdarzeń odrzuca się, a zdarzenie z numerm $(k+1)$ zostaje, potem następne k zdarzeń odrzuca się, a zdarzenie z numerem $2(k+1)$ zostaje, itd.:

$$\xrightarrow{\text{str. Poissona z } \lambda} \left\{ \begin{array}{c} \text{rekurencyjna operacja} \\ \text{rozrzedzania określona} \\ \text{funkcją } F(z) = z^k \end{array} \right\} \xrightarrow{\text{str. Erlanga rz. } k}$$

Należy tutaj określić warunki przy których funkcja $F(z)$ przyjmuje wartość z^k. Jeżeli rekurencyjna operacja rozrzedzania zadana jest przez ciąg $v_1, v_2, \cdots$ niezależnych zmiennych losowych o jednakowym rozkładzie ($v_1 = v_2 = \cdots$), to oznaczmy:

$$a_k = P\{v_1 = k\}, \text{ dla } k = 0, 1, \cdots \text{ z warunkiem } \sum_{k\geq 0} a_k = 1 \quad (1.25)$$

$$F(z) = \sum_{k\geq 0} a_k z^k \qquad \text{gdzie: } |z| \leq 1 \text{ , } F(1) = 1 \qquad (1.26)$$

w ten sposób została określona rekurencyjna operacja rozrzedzania za pomocą funkcji $F(z)$. Interesująca jest interpretacja probabilistyczna ostatniego równania. Należy tutaj dodać, że w rozwiązywaniu problemów z teorii masowej obsługi, posługujemy się funkcjami tworzącymi w postaci:

$$\sum_{k\geq 0} p_k z^k \qquad \text{lub } \sum_{k\geq 0} p_k \frac{z^k}{k!} \qquad (1.27)$$

i rozpatrzmy strumień zdarzeń przybywających do pewnego systemu obsługi. Niech z będzie taką wielkością, że $0 \leq z \leq 1$ oraz każde zdarzenie przybywające do systemu jest albo czerwone albo niebieskie. Dowolne zdarzenie jest czerwone z prawdopodobieństwem z, niezależnie od tego jakiego koloru są pozostałe zdarzenia. Jeżeli p_k jest prawdopodobieństwem przybycia k (czerwonych i niebieskich) zdarzeń w pewnym przedziale czasu, to $\sum_{k\geq 0} p_k z^k$ jest prawdopodobieństwem tego, że w tym przedziale czasu nie będzie niebieskich zdarzeń. Wracając do operacji rozrzedzania, gdzie k zdarzeń odrzuca się, a tylko jedno następne zostaje, to dla zadanego k wyrażenie $\sum_{k\geq 0} a_k z^k$ przyjmuje postać z^k (prawdopodobieństwo $a_k = p_k$ przybycia dokładnie k zdarzeń, w tym przypadku równe jest 1; patrz wzór (1.25)).

1.3.2 Sumowanie strumieni (superpozycja, nakładanie)

1) Superpozycja strumieni Poissona:

$$\left. \begin{array}{l} \xrightarrow{\text{strumień Poissona z par. } \lambda_1} \\ \xrightarrow{\text{strumień Poissona z par. } \lambda_2} \\ \quad \cdots \\ \xrightarrow{\text{strumień Poissona z par. } \lambda_n} \end{array} \right\{ \quad \right\} \xrightarrow{\text{str. Poissona z par. } \lambda_s}$$

gdzie, parametr λ_s równy jest sumie parametrów charakteryzujących poszczególne strumienie:

$$\lambda_s = \sum_{i=1}^{n} \lambda_i \tag{1.28}$$

2) Superpozycja dowolnych, nierekurentnych strumieni z ograniczonym średnim czasem między zdarzeniami, równym T_i , dla $i = 1$, 2 , $\cdots$:

$$\begin{array}{l} \xrightarrow{\text{str. dowolny z par. } T_1} \\ \xrightarrow{\text{str. dowolny z par. } T_2} \\ \quad \cdots \\ \xrightarrow{\text{str. dowolny z par. } T_i} \\ \quad \cdots \end{array} \left\{ \quad \right\} \xrightarrow{\text{str. Poissona z par. } \lambda_s}$$

gdy $i \to \infty$ daje to strumień Poissona, z intensywnością równą

$$\lambda_s = \sum_{i=1}^{\infty} \frac{1}{T_i} \tag{1.29}$$

1.3.3 Strumień Bernoulliego

W teorii kolejek, jak i w modelowaniu, dość często występują tzw. strumienie Bernoulliego. Jest to ograniczony strumień zdarzeń (zgłoszeń) definiowany następująco : niech $\tau_1, \tau_2, \cdots, \tau_n$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie, jednostajnym w przedziale $[0, T]$, gdzie $T > 0$. Oznaczmy:

$$v_k(t) = \begin{cases} 1 & \text{gdy} \quad \tau_k \geq t \\ \\ 0 & \text{gdy} \quad \tau_k < t \end{cases} \qquad \text{dla } k = 1, 2, \cdots, \text{n} \tag{1.30}$$

oraz

$$v(t) = \nu_1(t) + \nu_2(t) + \cdots + \nu_n(t) \tag{1.31}$$

to strumień $v(t)$, dla $t \geq 0$ nazywamy strumieniem Bernoulliego.

Strumień taki można przedstawić w następujący sposób: niech każde z n niezależnych źródeł, wysyła z prawdopodobieństwem równym 1, w przedziale czasowym $[0, T]$, dla $T > 0$, tylko jedno zdarzenie, przy czym

prawdopodobieństwo tego, że od pewnego ustalonego źródła nadejdzie zdarzenie w przedziale $\Delta \in [0, T]$ równe jest

$$\frac{\Delta}{T}$$

to taki sumaryczny strumień jest strumieniem Bernoulliego.

Wniosek: jeżeli o strumieniu Poissona wiadomo, że w przedziale $[0, T]$, dla $T > 0$ przybyło dokładnie N zdarzeń, to strumień w tym przedziale jest strumieniem Bernoulliego.

1.4 Procesy Markowa

W przedmiocie modelowania analitycznego jak i w teorii masowej obsługi, szczególne miejsce zajmuje pewna klasa procesów stochastycznych, zwanych procesami Markowa (w 1908 roku A. A. Markow opublikował pracę, w której zbadał i określił procesy, znane teraz jako procesy markowskie). Aby zdefiniować procesy Markowa, wróćmy jeszcze raz do określenia procesów stochastycznych.

Rozpatrzmy proces stochastyczny $\{X_t \, , t \in T\}$, w którym zbiór T parametru czasowego t jest zbiorem liczb rzeczywistych nieujemnych $T = \{t \geq 0\}$. Proces stochastyczny $\{X_t \, , t \in T\}$ jest w pełni określony, jeżeli dla każdego układu $t_0, t_1, \cdots, t_m$ znamy wszystkie $(m+1)$ - wymiarowe rozkłady:

$$P\{X_{t_0} = i_0, X_{t_1} = i_1, \cdots, X_{t_m} = i_m\} \tag{1.32}$$

dla $m = 0, 1, 2, \cdots$ oraz $i_0, i_1, \cdots, i_m \in S$. Wyrażenie to interpretowane jest w następujący sposób, jako łączne prawdopodobieństwo tego, że w momencie t_0 proces jest w dyskretnym stanie fazowym i_0, a w momencie t_1 w stanie $i_1, \cdots$, w momencie t_m w stanie i_m.

Dla określenia takiego procesu stochastycznego potrzebna jest zwykle bardzo duża liczba informacji i dlatego zazwyczaj należy ograniczyć się do badania jedynie jednowymiarowych rozkładów procesu, tj.

$$P\{X_t = k\} \qquad \text{dla } t \in T \, , \qquad \text{oraz} \quad k \in S$$

w praktyce warunek ten sprowadzany jest do jeszcze bardziej prostego przypadku, tzw. procesów Markowa, tj. procesów stochastycznych spełniających warunek Markowa.

Proces stochastyczny $\{X_t , t \in T\}$ spełnia warunek Markowa, jeżeli dla $n = 1, 2, \cdots$ oraz dla dowolnych wartości parametru czasowego $t_m \in T$, gdzie $m = 0, 1, \cdots, n$ oraz $t_0 < t_1 < \cdots < t_n$ i dla dowolnych stanów fazowych (przyjmujacych wartości rzeczywiste): $x , y , x_0 , x_1 , \cdots , x_{n-2}$ należących do zbioru S mamy:

$$P\left\{X_{t_n} < y \mid X_{t_{n-1}} = x , X_{t_{n-2}} = x_{n-2} , \cdots , X_{t_0} = x_0\right\} =$$

$$= P\left\{X_{t_n} < y \mid X_{t_{n-1}} = x\right\} \qquad (1.33)$$

tzn. dystrybuanta warunkowa zmiennej losowej X_{t_n} w punkcie y, pod warunkiem, że zmienne losowe $X_{t_{n-1}} , X_{t_{n-2}} , \cdots , X_{t_0}$, przyjęły odpowiednio wartości $x , x_{n-2} , \cdots, x_0$, jest równa dystrybuancie warunkowej zmiennej losowej X_{t_n} w punkcie y , pod jednym warunkiem, że zmienna losowa $X_{t_{n-1}}$ przyjęła wartość x. Proces mający własność Markowa nazywany jest procesem Markowa.

Procesy Markowa o parametrze czasowym dyskretnym (T jest zbiorem przeliczalnym) nazywane są łańcuchami Markowa.

Zdefiniowana przez wyrażenie (1.33) własność procesów Markowa oznacza, że przy ustalonym stanie x w momencie t dalsze zachowanie się procesu niezależne jest od tego, w jaki sposób znalazł się on w stanie x, tzn. jakie stany przyjmował przed momemtem t.

Niech będzie określony proces stochastyczny $\{X_n , n \in N_0\}$, którego zbiór wartości parametru czasowego jest zbiorem N_0– liczby naturalne i zero. Zapis

$$P\{X_n = k\} = d_{nk} \qquad \text{gdzie } n \in N_0 , k \in S \qquad (1.34)$$

określa prawdopodobieństwo d_{nk} z jakim proces w momencie n znajdzie się w dyskretnym stanie fazowym k. Przy tym wielkości d_{nk} spełniają następujące warunki:

$$0 \leq d_{nk} \leq 1 \qquad \text{dla } n \in N_0 , k \in S \qquad (1.35)$$

$$\sum_{k\in S} d_{nk} = 1 \qquad \text{dla } n \in N_0 \qquad (1.36)$$

Wprowadźmy oznaczenia na prawdopodobieństwo warunkowe znalezienia się procesu w stanie fazowym k w momencie n, pod warunkiem, że w

momencie $n - 1$ proces ten znajdował się w stanie fazowym j:

$$P\{X_n = k \mid X_{n-1} = j\} = p_{jk}(n), \text{ dla } j, k \in S , n = 1 , 2 , \cdots \qquad (1.37)$$

Czyli dla określenia łańcucha Markowa należy znać:

a) wszystkie możliwe wartości $p_{jk}(n)$ dla $j, k \in S$, oraz $n = 1, 2, \cdots$,

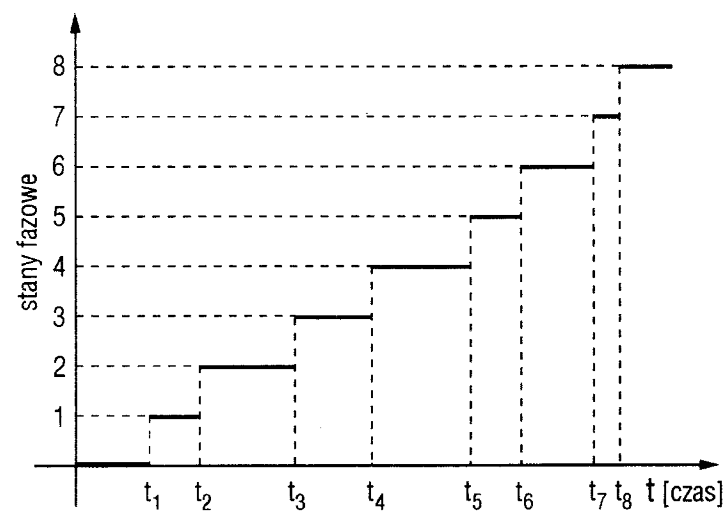
b) wszystkie możliwe wartości d_{0i} dla $i \in S$, czyli wektor rozkładu początkowego procesu (rozklad zmiennej losowej X_0), łańcuch Markowa $\{X_n , n \in N_0\}$ o przestrzeni fazowej S nazywamy jednorodnym, jeżeli prawdopodobieństwo warunkowe $p_{jk}(n)$ przejścia ze stanu fazowego j do stanu k od monentu $n - 1$ do momentu n nie zależy od n, tj.

$$p_{jk}(n) = p_{jk} \qquad \text{dla } n = 1 , 2 , \cdots \qquad \text{oraz } j, k \in S \qquad (1.38)$$

Jednorodny proces Poissona jest najprostszym przykładem procesu Markowa.

Niech $\{X_t , t \in T\}$, gdzie $T = \{t \geq 0\}$, będzie procesem stochastycznym o dyskretnej przestrzeni fazowej $S = \{0 , 1 , \cdots\}$. Niech zmienna losowa X_t oznacza liczbę określonych zdarzeń, które wystąpiły w przedziale czasowym od 0 do t; realizaja takiego procesu przedstawiona jest na rys. 2.

Każda realizacja takiego procesu stochastycznego jest niemalejącą funkcją czasu t, przedziałami stalą, o skokach jednostkowych.



Rys 2. Realizacja procesu stochastycznego

1.5 Proces narodzin i śmierci

Specjalną klasą jednorodnych procesów Markowa $\{X_t\ , t \geq 0\}$ o przestrzeni fazowej S jest tzw. proces narodzin i śmierci. Proces ten zajmuje szczególne miejsce w modelowaniu analitycznym opartym na teorii kolejek, jak również ma i wiele innych zastosowań, jak np. w demografii, teorii niezawodności itp. Ogólnie rzecz ujmując jest to proces z przybywaniem zdarzeń, jak i ich znikaniem, według pewnego probabilistycznego mechanizmu. Każde przybycie (pojawienie się) zdarzenia traktowane jest tutaj jako "narodziny", a z drugiej strony znikanie traktowane jest jako "śmierć". Inaczej rzecz ujmując, w procesie tym przejście ze stanu oznaczonego jako i do stanu $i+1$ traktowane jest jako "narodziny", a przejście do stanu oznaczonego jako $i-1$, to "śmierć". Jeżeli przez λ oznaczymy intensywność przybywania zdarzeń, a przez μ intensywność ich znikania, to proces narodzin i śmierci powinien spełniać następujące warunki :

1) Prawdopodobieństwo, że w małym przedziale czasowym miedzy t i $t+\Delta t$ pojawi się nowe zdarzenie równe jest $\lambda\Delta t + o(\Delta t)$. Warunek ten możemy zapisać również w następujący sposób:

$$P\{\text{nowe zdarzenie między } t \text{ i } t+\Delta t\} = \lambda\Delta t + o(\Delta t) \qquad (1.39)$$

gdzie λ jest stałą, a Δt jest małym przedziałem czasowym i $o(\Delta t)$ jest wielkością nieskończenie mała, taką że:

$$\lim_{\Delta t\to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

2) Prawdopodobieństwo pojawienia się, w przedziale Δt , więcej niż jednego zdarzenia równe jest $o(\Delta t)$:

$$P\{\text{więcej niż jedno zdarzenie między } t \text{ i } \Delta t\} = o(\Delta t) \qquad (1.40)$$

3) Liczby nowych zdarzeń w różnych, nie zachodzących na siebie, przedziałach czasowych, są statystycznie niezależne, czyli jest to proces o przyrostach niezależnych.

4) Prawdopodobieństwo zniknięcia zdarzenia w przedziale Δt , jeżeli wcześniej były pojawienia się, jest równe:

$$P\{\text{zniknięcie między } t \text{ i } t+\Delta t\} = \mu\Delta t\ +\ o(\Delta t) \qquad (1.41)$$

5) Prawdopodobieństwo zniknięcia w przedziale Δt , więcej niż jednego zdarzenia, jeżeli wcześniej one pojawiły się, równe jest:

$$P\{\text{więcej niż jedno zniknięcie między } t \text{ i } t+\Delta t\} = o(\Delta t) \qquad (1.42)$$

Reasumując: proces narodzin i śmierci, jest procesem Markowa, w którym zmiany stanu systemu, w małych przedziałach czasowych, zachodzą wyłącznie do stanów sąsiednich tzn., że każdy przyrost (narodziny) i każde zniknięcie (śmierć) powodują zmianę stanu systemu o jeden.

Przedstawmy graf stanów systemu wg następującego schematu (dla ograniczonej liczby możliwych stanów $= n+1$):

$$H_0 \longleftrightarrow H_1 \longleftrightarrow \ \cdots \ \longleftrightarrow H_{i-1} \longleftrightarrow H_i \longleftrightarrow H_{i+1} \longleftrightarrow \ \cdots \ \longleftrightarrow H_n$$

czyli jeżeli system jest w stanie H_i (potraktujmy stany systemu jako liczbę zdarzeń aktualnie znajdujących się w systemie, i tak H_i oznacza, że w systemie jest i zdarzeń), to każde pojawienie się nowego zdarzenia powoduje zmianę stanu (do przodu) do H_{i+1} (narodziny), a każde zniknięcie (śmierć) przesuwa stan systemu do tyłu, czyli do H_{i-1}.

Załóżmy, że w danym momencie t proces narodzin i śmierci znajduje się w stanie H_i i zobaczmy w jaki sposób , w momencie $t+\Delta t$, proces znowu okaże się w tym samym stanie H_i. Aby trafić do stanu H_i w momencie $t+\Delta t$, w procesie będącym w momencie t w stanie H_i powinno, w małym przedziale czasu Δt, pojawić się j nowych zdarzeń i tyle samo zdarzeń powinno zniknąć lub w procesie będącym w stanie H_{i+j} , w momencie t, przybyło, w czasie Δt, k nowych zdarzeń, oraz zniknęło $j+k$. Również, w momencie t, proces mógłby być w stanie H_{i-j} i w małym przedziale Δt , k zdarzeń zniknęłoby i $j+k$ przybyło. Jednak wiemy o tym, że prawdopodobieństwo przybycia lub zniknięcia więcej niż jednego zdarzenia, w czasie Δt , równe jest $o(\Delta t)$, więc skoncentrujmy się na przypadku pojedynczego pojawienia się zdarzenia lub zniknięcia, na tym małym przedziale czasowym. Czyli proces, w momencie t , może być w stanie H_i i w przedziale czasowym Δt, żadne zdarzenie nie pojawiło się, oraz nie zniknęło lub być w stanie H_{i-1} i w czasie Δt jedno zdarzenie pojawiło sie i żadne nie zniknęło, i w końcu, być w stanie H_{i+1} , w momencie t , i w przedziale Δt żadne zdarzenie

nie pojawiło się, a jedno zniknęło. Zakładamy oczywiście, że $i \geq 1$, więc zapiszmy:

$$P\{\text{system jest w stanie } H_i \text{ w momencie } t+\Delta t\} =$$

$$P\{\text{syst. w } H_i \text{ w } t \text{ i w } \Delta t \text{ żadne zd. nie przybywa i nie znika}\} +$$

$$P\{\text{syst. w } H_i \text{ w } t \text{ i w } \Delta t \text{ jedno zd. przybywa i jedno znika}\} +$$

$$P\{\text{syst. w } H_{i+1} \text{ w } t \text{ i w } \Delta t \text{ żadne zd. nie przybywa a jedno znika}\} +$$

$$P\{\text{syst. w } H_{i-1} \text{ w } t \text{ i w } \Delta t \text{ jedno zd. przybywa i żadne nie znika}\} +$$

$$+ o(\Delta t) \qquad (1.43)$$

Biorąc pod uwagę ten fakt, że procesy pojawiania się (przybywania) zdarzeń jak i ich znikania są niezależne, dla momentu t i $i \geq 1$ możemy zapisać:

$$p_i(t+\Delta t) = p_i(t) \cdot P\{\text{brak przybywań w } \Delta t\} \cdot P\{\text{brak znikań w } \Delta t\}$$

$$+p_i(t) \cdot P\{\text{jedno przybycie w } \Delta t\} \cdot P\{\text{jedno zniknięcie w } \Delta t\}$$

$$+p_{i+1}(t) \cdot P\{\text{jedno zniknięcie w } \Delta t\} \cdot P\{\text{brak przybywań w } \Delta t\}$$

$$+p_{i-1}(t) \cdot P\{\text{jedno przybycie w } \Delta t\} \cdot P\{\text{brak znikań w } \Delta t\} + o(\Delta t)$$

gdzie $p_i(t)$ - jest prawdopodobieństwem, że w momencie t proces znaj dzie się w stanie H_i. Zgodnie z warunkami podanymi dla dowolnego procesu narodzin i śmierci, jeżeli w momencie t proces znajduje się w stanie H_i, w małym przedziale Δt z prawdopodobieństwem $\lambda_i \cdot \Delta t + o(\Delta t)$ przejdzie do stanu H_{i+1}, oraz z prawdopodobieństwem $\mu_i \cdot \Delta t + o(\Delta t)$ wróci do stanu H_{i-1} i z prawdopodobieństwem $1-(\lambda_i+\mu_i)\cdot\Delta t + o(\Delta t)$ pozostanie w stanie H_i, gdzie λ_i i μ_i - intensywności procesu narodzin i śmierci. Czyli równanie dla $p_i(t+\Delta t)$ można zapisać w postaci:

$$p_i(t+\Delta t) = p_i(t)\,[1-\lambda_i\Delta t - o(\Delta t)]\,[1-\mu_i\Delta t - o(\Delta t)] +$$

$$+p_i(t)\,[\lambda_i\Delta t + o(\Delta t)]\,[\mu_i\Delta t + o(\Delta t)]$$

$$+p_{i+1}(t)\,[\mu_{i+1}\Delta t + o(\Delta t)]\,[1-\lambda_{i+1}\Delta t + o(\Delta t)] +$$

$$p_{i-1}(t)\,[\lambda_{i-1}\Delta t + o(\Delta t)]\,[1-\mu_{i-1}\Delta t + o(\Delta t)] + o(\Delta t) \qquad (1.44)$$

Po odpowiednich przekształceniach otrzymujemy :

$$p_i(t+\Delta t) = p_i(t)\,[1-\lambda_i\Delta t - \mu_i\Delta t] + p_{i+1}(t)\,[\mu_{i+1}\Delta t] + p_{i-1}(t)\,[\lambda_{i-1}\Delta t] +$$

$$+ o(\Delta t) \qquad \text{dla } i \geq 1 \qquad (1.45)$$

Dla zobrazowania wszystkich intensywnosci procesu przedstawmy teraz pełny graf stanów systemu:

$$H_0 \overset{\mu_1\ \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2\ \lambda_1}{\longleftrightarrow} \cdots \overset{\mu_i\ \lambda_{i-1}}{\longleftrightarrow} H_i \overset{\mu_{i+1}\ \lambda_i}{\longleftrightarrow} H_{i+1} \overset{\mu_{i+2}\ \lambda_{i+1}}{\longleftrightarrow} \cdots$$

gdzie μ_i i λ_i - intensywności procesu, dla $i = 0, 1, 2, \cdots$. Jeżeli wszystkie μ_i równe są zero, to mamy czysty proces narodzin, gdy zaś wszystkie λ_i równe są zero, jest to czysty proces śmierci.

Rozpatrzmy równanie (1.45) dla $i = 0$, kiedy p_{i-1} nie istnieje, a system może być w stanie H_0, w momencie $t+\Delta t$, jeżeli w momencie t był w stanie H_0 lub H_1 i wtedy:

$$p_0(t+\Delta t) = p_0(t)\,[1-\lambda_0\Delta t] + p_1(t)\,[\mu_1\Delta t] + o(\Delta t) \qquad (1.46)$$

gdyż dla $i = 0$ oraz $i = n$ (gdy liczba stanów jest równa $n+1$) intensywności:

$$\mu_0 = \lambda_{-1} = \lambda_n = 0 \qquad \text{gdy } 0 \leq i \leq n \qquad (1.47)$$

Równania (1.45) oraz (1.46) można przekształcić do następującej postaci:

$$p_i(t+\Delta t) - p_i(t) = -(\lambda_i+\mu_i)\Delta t p_i(t) + \mu_{i+1}\Delta t p_{i+1}(t) + \lambda_{i-1}\Delta t p_{i-1}(t) +$$

$$+o(\Delta t) \qquad \text{dla } i \geq 1$$

$$p_0(t+\Delta t) - p_0(t) = -\lambda_0\Delta t p_0(t) + \mu_1\Delta t p_1(t) + o(t)$$

a po podzieleniu przez Δt i przy $\Delta t \to 0$ otrzymamy:

$$\begin{cases} \frac{dp_i(t)}{dt} = -(\lambda_i+\mu_i)p_i(t) + \mu_{i+1}p_{i+1}(t) + \lambda_{i-1}p_{i-1}(t) & \text{dla } i \geq 1 \\ \frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t) & \text{dla } i = 0 \end{cases} \qquad (1.48)$$

jest to podstawowy układ równań różniczkowych dla procesu narodzin i śmierci.

W modelowaniu analitycznym, jak również w teorii masowej obsługi, często operujemy pojęciem stacjonarności lub stanu ustalonego, kiedy określane są parametry charakteryzujące dowolny proces lub system przy założeniu, że parametry te nie zależą od t. Czyli dla procesu narodzin i śmierci, zamiast prawdopodobieństw

$$p_i(t) = P\{i \text{ zdarzeń w systemie w momencie } t\}$$

(jest to dość złożone zadanie), częściej szukamy rozwiązania dla stanu ustalonego, gdy prawdopodobieństwa

$$p_i = \lim_{t\to\infty} p_i(t), \text{ dla } i = 0,\ 1,\ 2,\ \cdots$$

czyli gdy zależne od czasu bezwzględne prawdopodobieństwa $p_i(t)$ dążą przy $t \to \infty$ do stacjonarnych prawdopodobieństw granicznych charakterystycznych dla danego procesu. W takim przypadku prawdopodobieństwa te niezależne są od początkowego stanu systemu i są wielkościami stałymi.

Wprowadzając tutaj pojęcia bezwzględnych prawdopodobieństw (stacjonarnych prawdopodobieństw stanu) dla procesu narodzin i śmierci korzystamy z dwóch twierdzeń Markowa określających:

a) proces bez następstw (proces o przyrostach niezależnych),

b) ergodyczność procesu (wartość średnia parametru charakteryzującego proces, od nieskończenie wielu realizacji, równa jest wartości średniej od pojedyńczej, nieskończenie długiej realizacji).

Zakładając, że rozwiązania dla stanu ustalonego istnieją, wtedy wartości dla $p_i(t)$ są niezależne od czasu więc:

$$\lim_{t\to\infty} p_i(t) = p_i, \text{ to } \lim_{t\to\infty} p_i'(t) = 0,\ \frac{dp_i(t)}{dt} = 0 \qquad \text{dla } i = 0,\ 1,\ \cdots$$

oraz układ równań różniczkowych(1.48) przyjmuje postać następującego układu równań algebraicznych:

$$\begin{cases} 0 = -(\lambda_i + \mu_i)\, p_i + \mu_{i+1} p_{i+1} + \lambda_{i-1} p_{i-1} & \text{dla } i \geq 1 \\ 0 = -\lambda_0 p_0 + \mu_1 p_1 & \end{cases} \tag{1.49}$$

oraz:

$$\sum_{i\geq 0} p_i = 1 \tag{1.50}$$

Rozwiązanie powyższego układu równań można podzielić na dwie części, w zależności od tego, czy liczba możliwych stanów systemu jest skończona, czy nie.

1.5.1 Skończona przestrzeń stanów systemu

Jeżeli liczba stanów systemu równa jest n, to równania (1.49) i (1.50) powinny być uzupełnione o dodatkowy warunek w postaci:

$$\lambda_{-1} = \mu_0 = \lambda_n = 0$$

i wtedy z równania (1.49) otrzymamy:

$$\underbrace{\mu_i p_i - \lambda_{i-1} p_{i-1}}_{\Gamma_i} = \underbrace{\mu_{i+1} p_{i+1} - \lambda_i p_i}_{\Gamma_{i+1}} \tag{1.51}$$

generalizując, otrzymaliśmy następujące rekurencyjne (iteracyjne) wyrażenie:

$$\Gamma_{i+1} = \Gamma_i = \Gamma_{i-1} = \cdots = \Gamma_1$$

z tego z kolei równania (1.49) mamy:

$$\Gamma_1 = \mu_1 p_1 - \lambda_0 p_0 = 0$$

czyli wszystkie:

$$\mu_i p_i - \lambda_{i-1} p_{i-1} = 0 \tag{1.52}$$

Teraz z kolei można zapisać rekurencyjne równanie dla obliczenia p_i:

$$p_i = \frac{\lambda_{i-1}}{\mu_i} p_{i-1} = \frac{\lambda_{i-1}}{\mu_i}\left(\frac{\lambda_{i-2}}{\mu_{i-1}} p_{i-2}\right) = \frac{\lambda_{i-1}\ \lambda_{i-2}}{\mu_i\ \mu_{i-1}} p_{i-2} = \cdots$$

$$= \frac{\lambda_{i-1}\ \lambda_{i-2} \cdots \lambda_0}{\mu_i\ \mu_{i-1} \cdots \mu_1} p_0 \tag{1.53}$$

oznaczmy dla pewnego $s = 1,\ 2,\ \cdots$:

$$Q_s = \frac{\lambda_0\ \lambda_1\ \cdot\ \cdot\ \cdot\ \lambda_{s-1}}{\mu_1\ \mu_2\ \cdot\ \cdot\ \cdot\ \mu_s} = \prod_{i=1}^{s} \frac{\lambda_{i-1}}{\mu_i} \tag{1.54}$$

zakładamy również, że:

$$Q_0 = 1 \qquad \text{bo } p_0 = 1 \cdot p_0$$

czyli:

$$p_i = Q_i \cdot p_0 \tag{1.55}$$

Korzystając ze wzoru na sumę prawdopodobieństw można obliczyć stacjonarne prawdopodobieństwa dla każdego ze stanów systemu, więc:

$$\sum_{i=0}^{n} p_i = 1 = \sum_{i=0}^{n} Q_i\ p_0 \qquad \text{czyli } p_0 = \frac{1}{\sum_{i=0}^{n} Q_i} \tag{1.56}$$

oraz

$$p_i = \frac{Q_i}{\sum_{k=0}^{n} Q_k}\ , \qquad \text{dla } 0 \leq i \leq n \tag{1.57}$$

w ten oto sposób można obliczyć wszystkie stacjonarne prawdopodobieństw stanów procesu, jako funkcje odpowiednich intensywności λ_i , oraz μ_i , gdzie $0 \leq i \leq n$. Wyrażenia dla obliczania prawdopodobieństw stanu upraszczają się, jeżeli przyjmiemy, że intensywności:

$$\lambda_0 = \lambda_1 = \lambda_2 = \cdot\ \cdot\ \cdot = \lambda\ , \qquad \text{oraz } \mu_0 = \mu_1 = \mu_3 = \cdot\ \cdot\ \cdot = \mu$$

co jest założeniem najczęściej spotykanym w zastosowaniach teorii kolejek.

Powróćmy jeszcze raz do wzoru dla Q_s , to:

$$Q_s = \prod_{i=1}^{s} \frac{\lambda}{\mu} = \left(\frac{\lambda}{\mu}\right)^s = \rho^s \tag{1.58}$$

gdzie ρ - współczynnik obciążenia (wykorzystania) systemu.

1.5.2 Przestrzeń stanów systemu $n \rightarrow \infty$

W przypadku gdy wymiar przestrzeni stanów systemu dąży do nieskończoności oraz dla parametrów intensywności równych λ i μ wyrażenia dla prawdopodobieństw stanów upraszczają się jeszcze bardziej. Wtedy

$$\sum_{i=0}^{\infty} Q_i = \sum_{i=0}^{\infty} \rho^i \tag{1.59}$$

jest to wzór na ciąg geometryczny: $1 + \rho^1 + \rho^2 + \rho^3 + \cdot\ \cdot\ \cdot$, który jest zbieżny, jeżeli $|\rho| < 1$. W teorii kolejek warunek ten pełni funkcję warunku podstawowego dla stacjonarnych procesów. Intuicyjnie można stwierdzić, że dla procesu narodzin i śmierci, przypadek gdy $\lambda > \mu$ ($|\rho| > 1$) oznacza, że dana populacja będzie rozrastać się do nieskończoności.

Korzystając ze wzoru na sumę ciągu geometrycznego wyrażenie (1.59) można przekształcić do postaci:

$$\sum_{i=0}^{\infty} \rho^i = \frac{1}{1-\rho} \qquad \text{dla } \rho < 1 \tag{1.60}$$

więc

$$p_0 = \frac{1}{\frac{1}{1-\rho}} = 1 - \rho \qquad \text{dla } \rho = \frac{\lambda}{\mu} < 1 \tag{1.61}$$

oraz

$$p_i = \frac{\rho^i}{\frac{1}{1-\rho}} = \rho^i(1-\rho) \tag{1.62}$$

Reasumując: prawdopodobieństwa stanów w procesie narodzin i śmierci zależą li tylko od wartości parametrów λ i μ (intensywności procesu).

1.6 Zadania

Zadanie 1. Charakterystyka strumienia Poissona. Podstawowym parametrem charakteryzującym strumień Poissona jest prawdopodobieństwo wystąpienia k zdarzeń w przedziale od 0 do t:

$$P_k(t) = \frac{(\lambda\ t)^k}{k!}\ e^{-\lambda\ t} \qquad \text{dla } k \geq 0,\ t \geq 0$$

drugą charakterystyką jest dystrybuanta pewnej zmiennej losowej, czyli czasu między kolejnymi zdarzeniami, która zadana jest wzorem:

$$F(t) = 1 - e^{-\lambda\, t}$$

Zakładając, że klienci pewnego banku tworzą strumień Poissona, z parametrem:

$$\begin{aligned} \lambda &= 2 \quad \text{(pierwszy wariant)} \\ \lambda &= 4 \quad \text{(drugi wariant)} \\ \lambda &= 6 \quad \text{(trzeci wariant)} \\ \lambda &= 8 \quad \text{(czwarty wariant)} \end{aligned}$$

oznacza to, że w ciągu minuty do banku, przychodzi średnio dwóch, czterech, sześciu lub ośmiu klientów. Obliczyć i przedstawić w formie graficznej (wykresy, histogramy) prawdopodobieństwa przybycia do banku w ciągu minuty $k = 0,\ 1,\ 2, \cdots 16$ klientów (dla $\lambda = 2,\ 4,\ 6,\ 8$) oraz wszystkie dystrybuanty, czyli prawdopodobieństwa tego, że odstępy t między kolejnymi klientami są mniejsze lub równe 0.0, 0.1, 0.2, $\cdots$, 0.9, 1.0 (dla $\lambda = 2,\ 4,\ 6,\ 8$).

Wyniki i wykresy dla zadanego problemu podane są poniżej.
Prawdopodobieństwa przybycia zadanej liczby klientów $P_k(t)$:

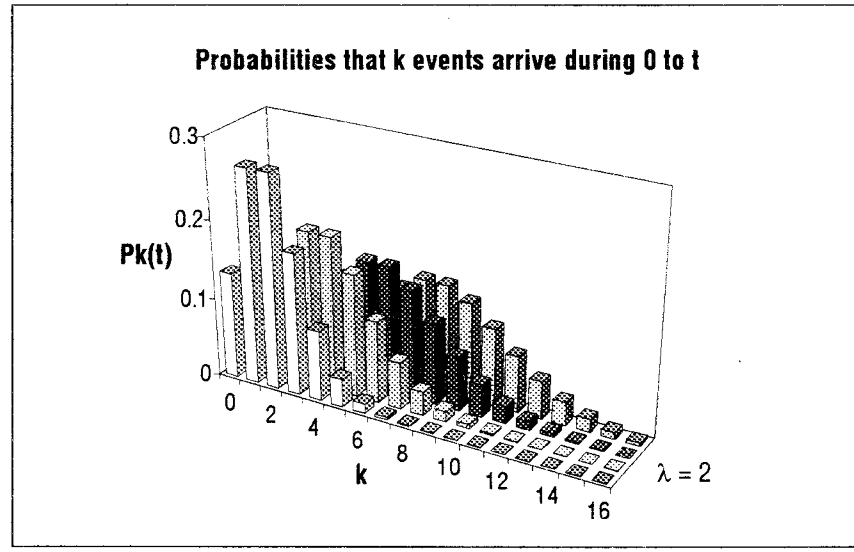
k	$\lambda = 2$	$\lambda = 4$	$\lambda = 6$	$\lambda = 8$
0	0.1353	0.0183	0.0025	0.0003
1	0.2707	0.0733	0.0149	0.0027
2	0.2707	0.1465	0.0446	0.0107
3	0.1804	0.1954	0.0892	0.0286
4	0.0902	0.1954	0.1339	0.0573
5	0.0361	0.1563	0.1606	0.0916
6	0.0120	0.1042	0.1606	0.1221
7	0.0034	0.0595	0.1377	0.1396
8	0.0009	0.0298	0.1033	0.1396
9	0.0002	0.0132	0.0688	0.1241
10	0.0000	0.0053	0.0413	0.0993
11	0.0000	0.0019	0.0225	0.0722
12	0.0000	0.0006	0.0113	0.0481
13	0.0000	0.0002	0.0052	0.0296
14	0.0000	0.0001	0.0022	0.0169
15	0.0000	0.0000	0.0009	0.0090
16	0.0000	0.0000	0.0003	0.0045

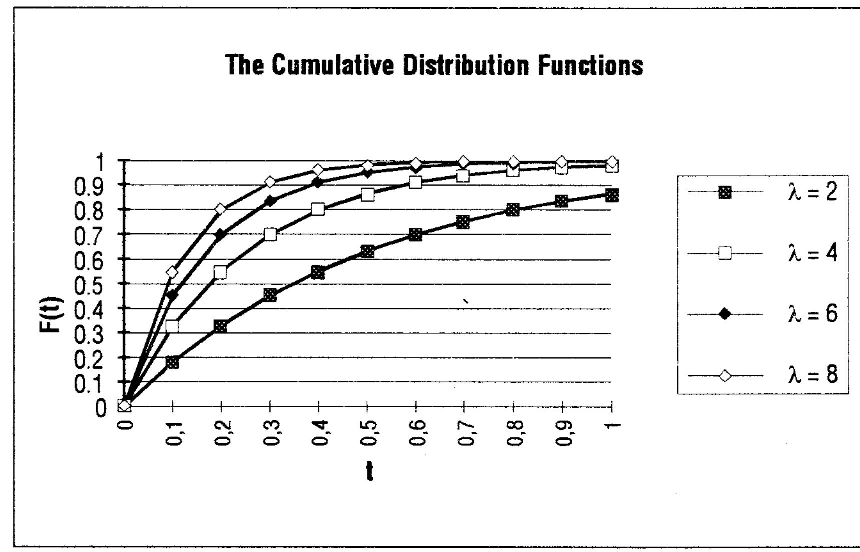
analiza prawdopodobieństw przybycia zadanej liczby klientów pokazuje, że bezpośrednio zależą one od wartości parametru λ. Dla $\lambda = 2$ maksymalne wartości przyjmują prawdopodobieństwa przybycia jednego lub dwóch klientów: $P_1(t) = P_2(t) = 0.2707$, dla $\lambda = 4$ - trzech lub czterech: $P_3(t) = P_4(t) = 0.1954$, dla $\lambda = 6$ - pięciu lub sześciu: $P_5(t) = P_6(t) = 0.1606$, zaś dla $\lambda = 8$ - siedmiu lub ośmiu: $P_7(t) = P_8(t) = 0.1396$. Z kolei dla $\lambda = 2$ prawdopodobieństwa przybycia dziesięciu lub więcej klientów w jednostce czasu, są już bardzo małe, mniejsze niż 10^{-4}.

Dystrybuanty $F(t)$:

t	$\lambda = 2$	$\lambda = 4$	$\lambda = 6$	$\lambda = 8$
0.0	0.0000	0.0000	0.0000	0.0000
0.1	0.1813	0.3297	0.4512	0.5507
0.2	0.3297	0.5507	0.6988	0.7981
0.3	0.4512	0.6988	0.8347	0.9093
0.4	0.5507	0.7981	0.9093	0.9592
0.5	0.6321	0.8647	0.9502	0.9817
0.6	0.6988	0.9093	0.9727	0.9918
0.7	0.7534	0.9392	0.9850	0.9963
0.8	0.7981	0.9592	0.9918	0.9983
0.9	0.8347	0.9727	0.9955	0.9993
1.0	0.8647	0.9817	0.9975	0.9997

Prawdopodobieństwa przybycia do banku zadanej liczby klientów w jednostce czasu $P_k(t)$ i dystrybuanty $F(t)$:





Zadanie 2. Pewne urządzenie może obsłużyć maksymalnie 10 zgłoszeń w ciągu jednej minuty. Zakłada się, że strumień zgłoszeń jest strumieniem Poissona z następującymi parametrami:

$$\begin{aligned}
\lambda &= 3 \quad \text{(wariant pierwszy)}\\
\lambda &= 4 \quad \text{(wariant drugi)}\\
\lambda &= 5 \quad \text{(wariant trzeci)}\\
\lambda &= 6 \quad \text{(wariant czwarty)}\\
\lambda &= 7 \quad \text{(wariant piąty)}\\
\lambda &= 8 \quad \text{(wariant szósty)}\\
\lambda &= 9 \quad \text{(wariant siódmy)}
\end{aligned}$$

oznacza to, że do urządzenia przybywa średnio 3, 4, 5, $\cdots$ zgłoszeń w ciągu minuty.

Prawdopodobieństwo P przybycia dziesięciu lub mniej zgłoszeń (częstotliwość przybywania zgłoszeń ≤ 10) określona jest wzorem:

$$P\{k \leq 10\} = \sum_{k=0}^{10} \frac{(\lambda\ t)^k}{k!} e^{-\lambda\ t}$$

i jest to prawdopodobieństwo tego, że urządzenie nie będzie przeciążone, tzn. że $P \cdot 100\%$ zgłoszeń będzie obsłużone od razu, a $(1-P) \cdot 100\%$ musi czekać w kolejce.

Obciążenie (wykorzystanie) urządzenia można określić jako:

$$\rho = \frac{\text{średnia liczba zgłoszeń na minutę}}{\text{maksymalna liczba obsłużonych zgłoszeń}}$$

Dla $t = 1$ obliczyć i przedstawić w formie wykresów wartości prawdopodobieństw P oraz parametru ρ, dla każdego z parametrów λ strumienia Poissona.

Rozdział 2

Markowowskie modele masowej obsługi (markowowskie modele kolejkowe)

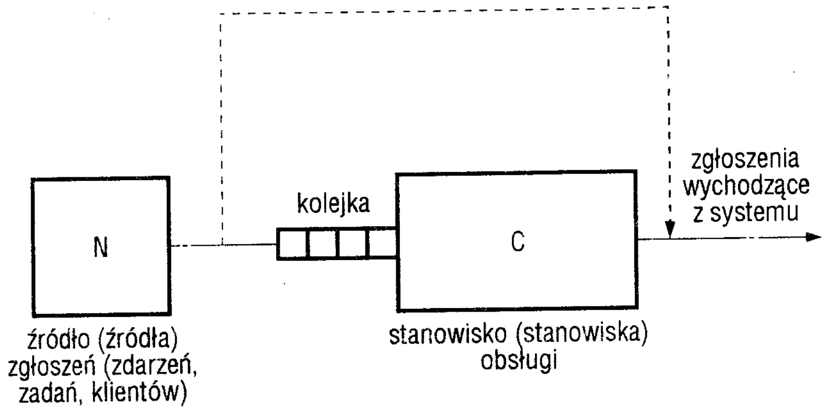
2.1 Wiadomości wstępne

Teoria masowej obsługi, zwana również teorią kolejek (angielska nazwa queueing theory), rozwinęła się na bazie poszukiwań modeli matematycznych (analitycznych), określających zachowanie się pewnych sytemów obsługujących losowo napływające zadania (zgłoszenia, zdarzenia). Pionierskie prace w tej dziedzinie prowadził duński matematyk A. K. Erlang badając obciążenia central (sieci) telefonicznych oraz publikując piewszą prace z tej dziedziny w 1909 roku. W swoich pracach Erlang zauważył, że sieci (systemy) telefoniczne, generalnie można scharakteryzować w następujący sposób :

a) jako systemy ze strumieniem Poissona na wejściu, z wykładniczym czasem obsługi oraz wieloma stanowiskami obsługi,

b) jako systemy ze strumieniem Poissona na wejściu, stałym czasem obsługi i pojedynczym stanowiskiem obsługi.

Inne ważne prace z teorii kolejek związane są z nazwiskiem D. G. Kendalla, który, między innymi usystematyzował systemy masowej obsługi (tzw. notacja Kendalla) i uważany jest za twórcę nauki o masowej obsłudze.



Rys 3. Schemat typowego systemu masowej obsługi

Systemy kolejkowe klasyfikuje się w zależności od :

1) Algorytmu (sposobu) napływu zgłoszeń, czyli charakterystyki strumienia wejściowego. Strumień na wejściu do systemu kolejkowego często charakteryzowany jest przez średnią liczbę zadań (klientów, zgłoszeń) w pewnej jednostce czasu lub przez wartość średnią czasu między wejściem do systemu dwóch kolejnych zgłoszeń. Te dwie wielkości są ściśle ze sobą powiązane i do opisu strumienia wejściowego wystarczy tylko jedna. Algorytm ten może być deterministyczny, gdy zgłoszenia napływają w regularnych odstępach czasu, lub probabilistyczny (stochastyczny), kiedy średni czas między zgłoszeniami charakteryzuje tylko ogólną tendencję dla strumienia wejściowego i jest tylko wartością oczekiwaną (moment pierwszego rzędu) zmiennej losowej - czasu między kolejnymi zadaniami. Wtedy proces napływu zgłoszeń scharakteryzowany jest przez dystrybuanty różnorodnych stochastycznych procesów.

Inną ważną charakterystyką procesu wejściowego jest możliwość grupowego napływu zgłoszeń zamiast napływu pojedynczych zgłoszeń. Ważna jest również reakcja każdego klienta (zgłoszenia) na konieczność czekania w kolejce przed stanowiskiem obsługi. W systemach ze stratą (bez oczekiwania), jeżeli zgłoszenie przychodzące do systemu zastaje

wszystkie stanowiska obsługi zajęte, opuszcza system nie obsłużone. Przykładem może być tutaj automatyczna centrala telefoniczna, gdzie, jeżeli numer jest zajęty, należy powtórzyć próbę jeszcze raz. W systemach z oczekiwaniem, zgłoszenie, w przypadku zajętości wszystkich stanowisk obsługi, ustawia się w kolejce i tutaj liczba miejsc może być ograniczona (też może wystąpić strata zgłoszenia) i nieograniczona.

Inną charakterystyką, wpływającą w znacznym stopniu na algorytm napływu zadań, jest możliwość zmiany tego algorytmu w czasie. Algorytm, w którym wartości parametrów, jak i rozkłady prawdopodobieństw nie ulegają zmianie, jest algorytmem stacjonarnym, w przypadku zmienności w czasie mamy do czynienia z algorytmem niestacjonarnym.

2) Sposobu obsługi zgłoszeń na stanowiskach obsługi. Omawiając proces obsługi zgłoszeń należy zwrócić uwagę na to, że stanowisko obsługi może być zajęte i wtedy możemy mówić np. o średnim czasie potrzebnym na obsługę pojedynczego zgłoszenia lub o liczbie obsłużonych zgłoszeń w jednostce czasu, albo o braku jakichkolwiek zgłoszeń na stanowisku (stanowiskach) obsługi. Algorytm obsługi maże być też tak deterministyczny (czas na obsługę zgłoszenia jest stały), jak i probabilistyczny (tutaj mamy do czynienia z różnym rozkładem czasu obsługi). Obsługa zgłoszeń moze być też grupowa i pojedyńcza, a intensywność obsługi może zależeć od liczby zgłoszeń czekających w kolejce. Parametry charakteryzujące proces obsługi mogą być stacjonarne lub zmienne w czasie.

3) Algorytmu szeregowania zadań (dyscypliny obsługi). Algorytmy te określają reguły wyboru zgłoszeń z kolejki do obsługi na stanowisku obsługi :

a) FIFO (first in, first out) - kolejność wg przybycia, regulamin naturalny,

b) LIFO (last in, first out) - ostatnie zgłoszenie jest obsługiwane najpierw (stos) - jak np. w magazynach, gdy bierzemy "z wierzchu",

c) SIRO (service in random order) - wybór "na chybił trafił", czyli niezależnie od czasu przybycia zgłoszenia,

d) priorytetowe szeregowanie, kiedy zgłoszenia wyższych priorytetów wybierane są do obsługi w pierwszej kolejności, niezależnie od liczby zadań niższych priorytetów :

- priorytet absolutny - z rugowaniem zadań niższego priorytetu, obsługiwane w danej chwili zadanie jest wyrzucane ze stanowiska obsługi,

- priorytet względny - zadanie wyższego priorytetu czeka na zakończenie obsługi, jeżeli na stanowisku jest zgłoszenie niższego priorytetu,

- priorytet dynamiczny - zależny od stanu kolejek lub algorytmu pracy stanowisk obsługi.

4) Pojemności poczekalni (liczby miejsc w kolejce). W systemach z oczekiwaniem i ograniczoną pojemnością kolejki, gdy wszystkie miejsca w poczekalni są zajęte, przychodzące zgłoszenie opuszcza system nie otrzymawszy obsługi, a następne zgłoszenie będzie przyjęte do systemu, tylko wtedy, gdy obsługiwane w danej chwili zgloszenie zakończy obsługę.

5) Liczby równoległych stanowisk (kanałów) obsługi. Jest to liczba identycznych stanowisk, w których zgłoszenia obsługiwane są równolegle. Stanowiska te mogą mieć wspólną kolejkę zgłoszeń, jak również każde stanowisko - oddzielną (np. supermarket i kolejka do kas).

6) Liczby dowolnie powiązanych stanowisk obsługi (sieci kolejkowe). W tym przypadku zgłoszenia obsługiwane są przez szereg niezależnych stanowisk. Przypomina to komisję lekarską, gdzie pacjent badany jest przez specjalistów z różnych dziedzin medycyny. W przypadku sieci kolejkowych zgłoszenie może być wielokrotnie obsługiwane na tym samym stanowisku lub przejść tylko przez część stanowisk.

Dla oznaczenia różnych typów systemów kolejkowych powszechnie używa się notacji zaproponowanej przez Kendalla, gdzie system masowej obsługi opisywany jest przez szereg symboli przedzielonych kreskami ukośnymi, jak:

$$A \;/\; B \;/\; c \;/\; L \;/\; N$$

gdzie:

A - rozkład zmiennej losowej τ_1, czyli czasu między kolejnymi zgłoszeniami (algorytm napływu zgłoszeń),

B - algorytm obsługi zgłoszeń (rozkład czasu obsługi τ_2),

c - liczba identycznych, równoległych stanowisk obsługi,

L - pojemność poczekalni,

N - wymiar źródła zgłoszeń.

Jeżeli L i N w notacji zostaną pominiente, to znaczy, że są nieskończenie wielkie.

Czyli, dla pełnego opisu systemu kolejkowego należy podać:

a) rozkłady prawdopodobieństwa zmiennych losowych τ_1 oraz τ_2 ,

b) określić wielkości c, N, L (czy są ograniczone, czy nie),

c) dyscyplinę obsługi,

d) czy zmienne τ_1 , τ_2 są zależne, czy niezależne.

W notacji Kendalla literom A i B odopowiadają nastepujące rozkłady:

D - rozkład regularny (deterministyczny) o stałych odstępach czasu,

M - rozkład wykładniczy o gęstości:

$$f(t) = \begin{cases} \lambda\, e^{-\lambda t} & \text{dla} \quad t \geq 0 \\ 0 & \text{dla} \quad t < 0 \end{cases} \qquad \text{gdzie } \lambda > 0 \tag{2.1}$$

E_k - rozkład Erlanga rzędu k , z funkcją gęstości:

$$f_k(t) = \begin{cases} \frac{\lambda}{(k-1)\,!}(\lambda\, t)^{k-1} e^{-\lambda\, t} & \text{dla} \quad t \geq 0 \\ 0 & \text{dla} \quad t < 0 \end{cases} \tag{2.2}$$

gdzie : $\lambda > 0$, a $k = 0, 1, \ldots$

G - rozkład dowolny (general),

GI - rozkład dowolny ze zmiennymi losowymi niezależnymi od siebie (general independent). Czyli np. system typu $M \;/\; E_2 \;/\; 3$ oznacza rozkład wejścia wykładniczy, czas obsługi oznacza rozkład Erlanga rzędu drugiego, 3 równoległe stanowiska obsługi.

Badając modele kolejkowe staramy się określić właściwości procesu stochastycznego $\{N(t)\ ,\ t > 0\}$, gdzie $N(t)$ dla zadanego t , jest zmienną losową oznaczającą liczbę zgłoszeń (klientów) znajdujących się w systemie w momencie t. Przestrzeń fazowa tego procesu $S = \{0,\ 1,\ 2,\ \ldots\ ,\ c\ +\ L\}$, zawiera $c + L + 1$ stanów fazowych (stany systemu), a przy nieograniczonej liczbie miejsc w kolejce przestrzeń fazowa S jest zbiorem przeliczalnym nieskończonym. Proces tego typu, w przypadku gdy zmienne losowe τ_1 i τ_2 (odstęp czasowy między kolejnymi zgłoszeniami i czas obsługi) mają rozkład wykładniczy, z parametrami λ i μ, jest procesem Markowa, a tego typu modele nazywane są markowowskimi. Ogólnie zakłada się, że zgłoszenia nadchodzą pojedynczo, niezależnie od siebie i wg regulaminu naturalnego FIFO.

Odwrotność parametrów λ i μ , czyli $\frac{1}{\lambda}$ i $\frac{1}{\mu}$ interpretowana jest jako średni czas między zgłoszeniami oraz jako średni czas obsługi pojedynczego zgłoszenia, zaś λ nazywane jest intensywnością napływu zgłoszeń, a μ intensywnością ich obsługi. Pamiętając o tym, że interesują nas parametry takiego systemu, po bardzo długim czasie działania (stan ustalony), wprowadzamy nowe oznaczenia parametrów systemu, które będą jakby przypisane do tego stacjonarnego stanu:

n - liczba zgłoszeń znajdujących się w systemie,
v - liczba zgłoszeń oczekujących w kolejce,
l - liczba zgłoszeń na stanowiskach (kanałach) obsługi,
w - czas oczekiwania w kolejce,
s - czas obsługi zgłoszenia,
q - czas pobytu zgłoszenia w systemie,
ρ - obciążenie (wykorzystanie) stanowiska obsługi.

2.2 Model systemu obsługi typu M/M/1 ze stratą (blokadą)

Systemy tego typu, opisywane markowskim procesem narodzin i śmierci, posiadają tylko dwa stany: H_0 - zero zgłoszeń w systemie i H_1 - jedno zgłoszenie na stanowisku obsługi (jedno zgłoszenie w systemie). Mamy więc system z jednym stanowiskiem obsługi, z blokadą (zero miejsc w poczekalni), więc stan H_1 określa max. liczbę zgłoszeń w systemie i jeżeli w tym czasie przyjdzie następne - to jest ono stracone (system ze stratą):

$$H_0 \;\overset{\mu_1 \qquad \lambda_0}{\longleftrightarrow}\; H_1$$

jeżeli przyjąć, że $\lambda_0 = \lambda$ i $\mu_1 = \mu$, to w sytemie tym:

$\frac{1}{\lambda}$ - średni czas między kolejnymi zgłoszeniami,
$\frac{1}{\mu}$ - średni czas obsługi.
Jeżeli oznaczymy:

$$F(t) = 1 - e^{-\lambda\, t}$$

jako dystrybuantę czasu między kolejnymi zgłoszeniami, a przez:

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

dystrybuantę czasu obsługi, to proces przybywania zgłoszeń, jak i ich obsługę można przedstawić w następujący sposób:

$$\underset{F(t)}{\underbrace{\overset{\text{czas do przyb. pierwszego zgł.}}{|\,-\,-\,-\,-\,-\,-}}}\quad \underset{\text{obsługa pierw. i ew. strata następnych zgł.}}{\overset{G(t)}{\overbrace{|\,-\,-\,-\,-\,-\,-\,-\,-\,-\,-}}}\quad \underset{F(t)}{\underbrace{\overset{\text{do kolejn. zgłoszenia}}{\cdots\,-\,-}}}\quad \underset{\text{obsługa i strata}}{\overset{G(t)}{\overbrace{|\,-\,-\,-\,-\,|}}}\ \xrightarrow{t}$$

Dla zadanego systemu kolejkowego (stan ustalony), w którym mogą wystąpić tylko stany H_0 i H_1, opierając się na wzorach dla procesu narodzin i śmierci, można od razu napisać:

$$Q_0 = 1, \qquad Q_1 = \frac{\lambda_0}{\mu_1} = \frac{\lambda}{\mu} = \rho \tag{2.3}$$

oraz

$$p_0 = \frac{1}{Q_0 + Q_1} = \frac{1}{1+\rho} \tag{2.4}$$

$$p_1 = \frac{Q_1}{Q_0 + Q_1} = \frac{\rho}{1+\rho} \tag{2.5}$$

W systemie M/M/1 ze stratą, oprócz prawdopodobieństw stanów, używa się i następujących dodatkowych charakterystyk:

a) prawdopodobieństwo obsługi zgłoszenia (stan systemu H_0):

$$p_{obs} = p_0 = \frac{1}{1+\rho} \tag{2.6}$$

b) prawdopodobieństwo straty zgłoszenia (stan systemu H_1):

$$p_{bl} = p_1 = \frac{\rho}{1+\rho} \tag{2.7}$$

oznacza to, że w systemie (tj. na stanowisku obsługi) jest tylko jedno zgłoszenie.

2.3 Wielostanowiskowy M/M/c system obs[łu]gi ze stratą

System taki charakteryzuje się tym , że mamy tutaj c identycznych stanowisk(kanałów) obsługi. Oznacza to, że w tym samym czasie, nawet c zgłoszeń może być obsługiwane.

Stany systemu:

$$H_0 \underset{\longleftrightarrow}{\mu_1\ \lambda_0} H_1 \underset{\longleftrightarrow}{\mu_2\ \lambda_1} H_2 \underset{\longleftrightarrow}{\mu_3\ \lambda_2} \cdots \underset{\longleftrightarrow}{\mu_c\ \lambda_{c-1}} H_c$$

gdzie:

H_0 - oznacza brak zgłoszeń w systemie (w tym konkretnym przypadku na stanowiskach obsługi),

H_1- jedno zgłoszenie (zdarzenie, zadanie, klient) obsługiwane (obojętnie na jakim stanowisku obsługi),

H_2 - dwa zgłoszenia obsługiwane, itd.,

H_c - wszystkie stanowiska obsługi zajęte, blokada systemu.

W analizie tego i podobnych systemów zakładamy, że zgłoszenia przychodzą z zewnątrz i źródło zgłoszeń jest bardzo duże (w teorii nieskończenie wielkie), a oznacza to, że fakt przyjęcia zgłoszeń do obsługi nie zmienia nic w charakterystyce (intensywności) strumienia wejściowego. Niech

$$\lambda_0 = \lambda_1 = \cdots = \lambda_{c-1} = \lambda$$

a $\frac{1}{\lambda}$ - to średni czas między kolejnymi zgłoszeniami; μ_1 - intensywność obsługi w systemie, gdy napłynęło tylko jedno zgłoszenie i:

$$\mu_1 = \mu$$

a $\frac{1}{\mu}$ - średni czas obsługi, następnie μ_2 reprezentuje intensywność obsługi w systemie, gdy dwa zgłoszenia są na stanowiskach obsługi i jest to intensywność przejścia ze stanu H_2 do H_1, oznacza to, że jedno albo drugie zgłoszenie zakończy obsługę i system przejdzie do stanu H_1. Więc:

$$\mu_2 = 2\mu$$

i tak dalej

$$\mu_3 = 3\mu , \cdots , \mu_i = i\mu , \cdots , \mu_c = c\mu \qquad (2.8)$$

więc wyrażenie na Q_i (dla $i = 1, 2, \cdots, c$) przyjmuje postać:

$$Q_i = \frac{\lambda_0\ \lambda_1 \cdots \lambda_{i-1}}{\mu_1\ \mu_2 \cdots \mu_i} = \frac{\lambda^i}{i!\ \mu^i} = \frac{\rho^i}{i!} \qquad \text{oraz } Q_0 = 1 \qquad (2.9)$$

należy tutaj przypomnieć, że dla stanów granicznych (ustalonych) systemu, przyjmuje się $\rho < 1$, ale ρ może być ≥ 1.

Stacjonarne prawdopodobienstwa stanów systemu oblicza się więc ze wzoru:

$$p_i = \frac{Q_i}{\sum_{j=0}^{c} Q_j} = \frac{\frac{\rho^i}{i\,!}}{\sum_{j=0}^{c} \frac{\rho^j}{j\,!}} \qquad (2.10)$$

przy

$$\sum_{i=0}^{c} p_i = 1$$

Podstawowe parametry (charakterystyki) wielostanowiskowego systemu z blokadą:

a) prawdopodobieństwo blokady systemu (wszystkie stanowiska zajęte):

$$p_{bl} = p_c = \frac{\frac{\rho^c}{c\,!}}{\sum_{j=0}^{c} \frac{\rho^j}{j\,!}} \qquad (2.11)$$

wyrażenie to interpretowane jest w następujący sposób: jeżeli do systemu przyjdzie nowe zgłoszenie, a system jest w stanie H_c, to każde nowe zgłoszenie będzie stracone (odrzucone), póki system będzie w tym stanie,

b) prawdopodobieństwo obsługi zgłoszenia:

$$p_{obs} = 1 - p_{bl} = 1 - p_c \qquad (2.12)$$

czyli zgłoszenie będzie obsłużone, jeżeli system nie będzie w stanie H_c, czyli w stanach od H_0 do H_{c-1}, bo

$$\sum_{i=0}^{c-1} p_i + p_c = 1$$

c) średnia liczba zgłoszeń na stanowiskach obsługi (zajętych stanowisk):

$$\overline{l} = \sum_{i=0}^{c} i\ p_i = \sum_{i=0}^{c} i\ \frac{\frac{\rho^i}{i\,!}}{\sum_{j=0}^{c} \frac{\rho^j}{j\,!}} \tag{2.13}$$

wzór ten można wyprowadzić i w trochę inny sposób, mianowicie wprowadz pojęcie absolutnej przepustowości i obliczając średnią liczbę zgłoszeń obsłużonych w jednostce czasu, czyli niech

$$A = \lambda\ p_{obs}$$

oznacza średnią liczbę zgłoszeń przybyłych i przyjętych na obsługę w jednostce czasu, a pojedyncze stanowisko jest w stanie obsłużyć, w tej jednostce czasu, μ zgłoszeń, czyli średnia liczba zajętych stanowisk równa jest $\frac{A}{\mu}$, więc

$$\overline{l} = \frac{A}{\mu} = \frac{\lambda\ p_{obs}}{\mu} = \rho\ p_{obs} = \rho(1 - p_c) \tag{2.14}$$

Przykład: W wielostanowiskowym systemie masowej obsługi ze stratą parametr $c = 3$, a $\lambda = 1$ i $\mu = 2$. Obliczyć średnią liczbę zgłoszeń na stanowiskach obsługi korzystając ze wzorów (2.13) i (2.14).

Na początek narysujmy graf stanów systemu:

$$H_0 \ \underset{\longleftrightarrow}{\mu_1 \quad \lambda_0}\ H_1 \ \underset{\longleftrightarrow}{\mu_2 \quad \lambda_1}\ H_2 \ \underset{\longleftrightarrow}{\mu_3 \quad \lambda_2}\ H_3$$

gdzie:

H_0 - brak zgłoszeń w systemie ,
H_1 - jedno zgłoszenie jest obsługiwane,
H_2 - dwa zgłoszenia na stanowiskach obsługi,
H_3 - trzy zgłoszenia na stanowiskach (wszystkie stanowiska zajęte: blokada następnych zgłoszeń)
$\lambda_0 = \lambda_1 = \lambda_2 = \lambda = 1$ - intensywność strumienia wejściowego,
$\mu_1 = \mu = 2$; $\mu_2 = 2\mu$; $\mu_3 = 3\mu$ - intensywność obsługi.

Obliczmy wartość parametru ρ (obciążenie stanowiska obsługi):

$$\rho = \frac{\lambda}{\mu} = \frac{1}{2}$$

oraz wszystkie stacjonarne prawdopodobieństwa stanów:

$$p_0 = \frac{1}{\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{3!}} = \frac{1}{1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{48}} = \frac{1}{\frac{79}{48}} = \frac{48}{79}$$

$$p_1 = \frac{\frac{\rho^1}{1!}}{\frac{79}{48}} = \frac{\frac{1}{2}}{\frac{79}{48}} = \frac{24}{79}$$

$$p_2 = \frac{\frac{\rho^2}{2!}}{\frac{79}{48}} = \frac{\frac{1}{8}}{\frac{79}{48}} = \frac{6}{79}$$

$$p_3 = \frac{\frac{\rho^3}{3!}}{\frac{79}{48}} = \frac{\frac{1}{48}}{\frac{79}{48}} = \frac{1}{79}$$

więc

$$\overline{l} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 = 0 + 1 \cdot \frac{24}{79} + 2 \cdot \frac{6}{79} + 3 \cdot \frac{1}{79} = \frac{39}{79}$$

porównajmy teraz ten rezultat z wynikiem otrzymanym z wyrażenia (2.14):

$$\overline{l} = \rho(1 - p_c) = \frac{1}{2}(1 - \frac{1}{79}) = \frac{1}{2} \cdot \frac{78}{79} = \frac{39}{79}$$

2.4 System obsługi typu M/M/1/L z oczekiwaniem (kolejką)

Jest to system z pojedynczym stanowiskiem obsługi i maksymalną pojemnością poczekalni (liczba miejsc w kolejce) równą m ($L = m$). Zacznijmy, jak zwykle, od analizy grafu stanów systemu:

$$H_0 \ \underset{\longleftrightarrow}{\mu_1\ \lambda_0}\ H_1 \ \underset{\longleftrightarrow}{\mu_2\ \lambda_1}\ H_2 \ \underset{\longleftrightarrow}{\mu_3\ \lambda_2}\ \cdots \ \underset{\longleftrightarrow}{\mu_{m+1}\ \lambda_m}\ H_{m+1}$$

gdzie:

H_0 - brak zgłoszeń w systemie,
H_1 - jedno zgłoszenie (na stanowisku obsługi), kolejka pusta,
H_2 - dwa zgłoszenia w systemie: jedno na stanowisku obsługi i jedno w kolejce,

H_{m+1} - $m+1$ zgłoszeń w systemie: jedno na stanowisku obsługi i m w poczekalni; gdy system znajduje się w tym stanie, to każde nowe zgłoszenie jest tracone (blokada systemu), aż do czasu zakończenia obsługi zgłoszenia, które aktualnie jest na stanowisku obsługi.

Jeżeli założymy, że:

$$\lambda_0 = \lambda_1 = \lambda_2 = \cdot\cdot\cdot = \lambda_m = \lambda$$

oraz

$$\mu_1 = \mu_2 = \mu_3 = \cdot\cdot\cdot = \mu_{m+1} = \mu$$

to

$$Q_i = \frac{\lambda^i}{\mu^i} = \rho^i \qquad \text{dla } i = 1,\ 2,\ \cdot\cdot\cdot, m+1 \qquad \text{oraz } Q_0 = 1 \tag{2.15}$$

$$p_i = \frac{\rho^i}{\sum_{j=0}^{m+1} \rho^j} \qquad \text{dla } i = 0,\ 1,\ \cdot\cdot\cdot, m+1 \tag{2.16}$$

oczywiście

$$\sum_{i=0}^{m+1} p_i = 1$$

Pamiętając o tym, że w stanie ustalonym $\rho < 1$, i korzystając ze wzoru na sumę cząstkową postępu geometrycznego ($m+2$ wyrazów) powyższe wyrażenie dla p_i można uprościć:

$$\sum_{j=0}^{m+1} \rho^j = \frac{1-\rho^{m+2}}{1-\rho}$$

więc

$$p_i = \frac{\rho^i(1-\rho)}{1-\rho^{m+2}} \tag{2.17}$$

dla $i = 0$ otrzymamy:

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}} \tag{2.18}$$

oraz

$$p_i = p_0 \cdot \rho^i \tag{2.19}$$

Podstawowe charakterystyki systemu:

a) prawdopodobieństwo blokady systemu (stan systemu H_{m+1}):

$$p_{bl} = p_{m+1} = \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}} \tag{2.20}$$

czyli blokada następuje wtedy, gdy w poczekalni wszystkie miejsca są zajęte i przychodzące zgłoszenie zostaje odrzucone,

b) prawdopodobieństwo obsługi zgłoszenia (brak zgłoszeń na stanowisku obsługi lub wolne miejsca w poczekalni - stany systemu od H_0 do H_m):

$$p_{obs} = 1 - p_{bl} = 1 - \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}} = \frac{1-\rho^{m+2}-\rho^{m+1}+\rho^{m+2}}{1-\rho^{m+2}} = \frac{1-\rho^{m+1}}{1-\rho^{m+2}} \tag{2.21}$$

c) średnia liczba zgłoszeń w kolejce (kolejka zaczyna się od stanu H_2):

$$\overline{v} = 1 \cdot p_2 + 2 \cdot p_3 + \cdot\cdot\cdot\ m \cdot p_{m+1} =$$

$$= \sum_{v=1}^{m} v\ p_{v+1} = \sum_{v=1}^{m} v\ \frac{\rho^{v+1}(1-\rho)}{1-\rho^{m+2}} = \frac{(1-\rho)\rho^2}{1-\rho^{m+2}} \sum_{v=1}^{m} v\ \rho^{v-1}$$

dla uproszczenia tego wyrażenia rozpatrzmy oddzielnie następującą sumę:

$$\sum_{v=1}^{m} v\ \rho^{v-1} = \sum_{v=1}^{m} (\rho^v)' = \left(\sum_{v=1}^{m} \rho^v\right)'$$

rozwijając dalej sumę zawartą w nawiasach otrzymamy:

$$\rho^1 + \rho^2 + \cdot\cdot\cdot + \rho^m =$$

$$= \rho(1 + \rho + \rho^1 + \cdot\cdot\cdot + \rho^{m-1}) = \rho\left(\frac{1-\rho^m}{1-\rho}\right) = \frac{\rho-\rho^{m+1}}{1-\rho}$$

obliczając następnie pochodną tego wyrażenia mamy:

$$\left(\frac{\rho-\rho^{m+1}}{1-\rho}\right)' = \frac{[1-(m+1)\rho^m](1-\rho)-(\rho-\rho^{m+1})(-1)}{(1-\rho)^2} =$$

$$= \frac{1-\rho-(m+1)\rho^m+(m+1)\rho^{m+1}+\rho-\ \rho^{m+1}}{(1-\rho)^2} =$$

$$\frac{1 - m\rho^m + m\rho^{m+1} - \rho^m}{(1-\rho)^2} = \frac{1 + \rho^m\,[m\rho - (m+1)]}{(1-\rho)^2} \tag{2.22}$$

więc wyrażenie na średnią liczbę zgłoszeń w kolejce przyjmie następującą postać:

$$\overline{v} = \frac{(1-\rho)\rho^2}{1-\rho^{m+2}}\left[\frac{1 + \rho^m\,[m\rho - (m+1)]}{(1-\rho)^2}\right] = \frac{\rho^2\,[1 + \rho^m(m\rho - (m+1))]}{(1-\rho^{m+2})(1-\rho)} \tag{2.23}$$

d) średnia liczba zgłoszeń na stanowisku obsługi (stany systemu $H_1, H_2, \cdots, H_{m+1}$):

$$\overline{l} = 1\cdot p_1 + 1\cdot p_2 + \cdots + 1\cdot p_{m+1} =$$

$$= 1\cdot(1-p_0) = 1 - \frac{1-\rho}{1-\rho^{m+2}} = \frac{\rho(1-\rho^{m+1})}{1-\rho^{m+2}} \tag{2.24}$$

e) średnia liczba zgłoszeń znajdujących się w systemie (stany od H_1 do H_{m+1}):

$$\overline{n} = 1\cdot p_1 + 2\cdot p_2 + \cdots + (m+1)\cdot p_{m+1} =$$

$$= \sum_{i=1}^{m+1} i\ p_i = \sum_{i=1}^{m+1} i\ p_0\rho^i = p_0\rho\sum_{i=1}^{m+1} i\ \rho^{i-1} =$$

$$= p_0\ \rho\sum_{i=1}^{m+1}\frac{d}{d\rho}\left(\rho^i\right) = p_0\ \rho\ \frac{d}{d\rho}\left(\sum_{i=1}^{m+1}\rho^i\right) = p_0\ \rho\ \frac{d}{d\rho}\left(\frac{\rho - \rho^{m+2}}{1-\rho}\right) =$$

$$= p_0\ \rho\ \frac{[1-(m+2)\rho^{m+1}]\,(1-\rho) - (\rho - \rho^{m+2})\,(-1)}{(1-\rho)^2} =$$

$$= p_0\ \rho\ \frac{1 - \rho - (m+2)(\rho^{m+1} - \rho^{m+2}) + \rho - \rho^{m+2}}{(1-\rho)^2} =$$

$$= \frac{1 - (m+2)\rho^{m+1} + (m+1)\rho^{m+2}}{(1-\rho)^2}\cdot\frac{(1-\rho)\rho}{1-\rho^{m+2}} =$$

$$= \frac{\rho\,[1 - (m+2)\rho^{m+1} + (m+1)\rho^{m+2}]}{(1-\rho)(1-\rho^{m+2})} \tag{2.25}$$

średnią liczbę zgłoszeń w systemie, można również otrzymać i z następującego wyrażenia:

$$\overline{n} = \overline{v} + \overline{l}$$

f) średni czas oczekiwania w kolejce:

$$\overline{w} = p_1\cdot\frac{1}{\mu} + p_2\cdot\frac{2}{\mu} + \cdots + p_m\cdot\frac{m}{\mu} + p_{m+1}\cdot 0 \tag{2.26}$$

poszczególne wyrazy tej sumy interpretowane są następująco:

$p_1\cdot\frac{1}{\mu}$ - w systemie i na stanowisku obsługi jest tylko jedno zgłoszenie, więc przychodzące zgłoszenie czeka $\frac{1}{\mu}$,

$p_2\cdot\frac{2}{\mu}$ - w systemie są dwa zgłoszenia, więc przychodzące zgłoszenie czeka $2\cdot\frac{1}{\mu}$, itd.,

$p_{m+1}\cdot 0$ - w systemie jest jedno zgłoszenie na stanowisku obsługi i m zgłoszeń w kolejce (poczekalni), więc zgłoszenie opuszcza system nie obsłużone, więc:

$$\overline{w} = \sum_{i=1}^{m}\frac{i}{\mu}p_i = \sum_{i=1}^{m}\frac{1}{\mu}\cdot i\ \cdot\frac{\rho^i\,(1-\rho)}{1-\rho^{m+2}} = \frac{\rho\,(1-\rho)}{\mu(1-\rho^{m+2})}\sum_{i=1}^{m} i\cdot\rho^{i-1}$$

i korzystając ze wzoru (2.22), otrzymamy:

$$\overline{w} = \frac{\rho\,(1-\rho)}{\mu\,(1-\rho^{m+2})}\cdot\frac{1+\rho^m\,[m\rho-(m+1)]}{(1-\rho)^2} = \frac{\rho\,[1+\rho^m\,(m\rho-(m+1))]}{\mu\,(1-\rho)(1-\rho^{m+2})} \tag{2.27}$$

średni czas oczekiwania w kolejce można również obliczyć z tzw. formuły Little'a, wiążącej średni czas oczekiwania w kolejce i średnią liczbę zgłoszeń w kolejce lub średni czas pobytu zgłoszenia w systemie i średnią liczbę zgłoszeń w systemie:

$$\overline{v} = \lambda\,\overline{w} \tag{2.28}$$

$$\overline{n} = \lambda\,\overline{q} \tag{2.29}$$

g) średni czas pobytu zgłoszenia w systemie:

$$\overline{q} = \underbrace{p_1\cdot\frac{1}{\mu} + p_2\cdot\frac{2}{\mu} + \cdots + p_m\cdot\frac{m}{\mu}}_{\overline{w}} + p_{obs}\cdot\frac{1}{\mu} \tag{2.30}$$

gdzie: $\overline{w}$ średni czas oczekiwania w kolejce, a $p_{obs}\cdot\frac{1}{\mu}$ interpretowane jest jako średni czas na obsługę zgłoszenia na stanowisku obsługi (część zgłoszeń jest tracona), więc:

$$\overline{q} = \frac{\rho}{\mu}\left[\frac{1+\rho^m(m\rho-(m+1))}{(1-\rho^{m+2})(1-\rho)}\right] + \frac{1-\rho^{m+1}}{1-\rho^{m+2}}\cdot\frac{1}{\mu} =$$

$$= \frac{1}{\mu(1-\rho^{m+2})}\left[\frac{\rho\,[1+\rho^m(m\rho-(m+1))]}{1-\rho}+\frac{(1-\rho)(1-\rho^{m+1})}{1-\rho}\right]=$$

$$= \frac{1}{\mu(1-\rho^{m+2})}\left[\frac{\rho+m\rho^{m+2}-(m+1)\rho^{m+1}+1-\rho^{m+1}-\rho+\rho^{m+2}}{1-\rho}\right]=$$

$$\frac{1}{\mu}\left[\frac{1-(m+2)\rho^{m+1}+(m+1)\rho^{m+2}}{(1-\rho^{m+2})(1-\rho)}\right] \tag{2.31}$$

średni czas pobytu zgłoszenia w systemie, można również obliczyć ze wzoru Little'e:

$$\overline{q}=\frac{\overline{n}}{\lambda}$$

h) średni czas obsługi zgłoszenia na stanowisku obsługi (część zgłoszeń odchodzi nie obsłużona, czyli jest tracona):

$$\overline{s}=p_{obs}\cdot\frac{1}{\mu}=\frac{1-\rho^{m+1}}{1-\rho^{m+2}}\cdot\frac{1}{\mu} \tag{2.32}$$

Analiza systemu kolejkowego typu M/M/1/L znacząco upraszcza się, przy założeniu, że liczba miejsc w poczekalni (w kolejce) $\rightarrow\infty$ i gdy analizę przeprowadza się dla stanu ustalonego (zakładamy, że $\rho<1$). Wtedy:

1) prawdopodobieństwa stanów systemu:

$$p_0=1-\rho \qquad \text{(ze wzoru (2.18))} \tag{2.33}$$

$$p_i=\rho^i(1-\rho) \qquad \text{dla } i=1,\ 2,\ \cdot\ \cdot\ \cdot \qquad \text{(ze wzoru (2.17))} \tag{2.34}$$

2) prawdopodobieństwo blokady systemu (ze wzoru (2.20)):

$$p_{bl}=0 \tag{2.35}$$

3) prawdopodobieństwo obsługi zgłoszenia (ze wzoru (2.21)):

$$p_{obs}=1 \tag{2.36}$$

4) średnia liczba zgłoszeń w kolejce (ze wzoru (2.23)):

$$\overline{v}=\frac{\rho^2}{1-\rho} \tag{2.37}$$

5) średnia liczba zgłoszeń na stanowisku obsługi (ze wzoru (2.24)):

$$\overline{l}=\rho \tag{2.38}$$

6) średnia liczba zgłoszeń w systemie (ze wzoru (2.25)):

$$\overline{n}=\frac{\rho}{1-\rho} \tag{2.39}$$

lub

$$\overline{n}=\overline{v}+\overline{l}=\frac{\rho^2}{1-\rho}+\rho=\frac{\rho^2+\rho-\rho^2}{1-\rho}=\frac{\rho}{1-\rho}$$

7) średni czas oczekiwania w kolejce (ze wzoru (2.27)):

$$\overline{w}=\frac{\rho}{\mu(1-\rho)} \tag{2.40}$$

lub z formuły Little'a :

$$\overline{w}=\frac{\overline{v}}{\lambda}=\frac{\rho^2}{\lambda(1-\rho)}=\frac{\frac{\lambda}{\mu}\,\rho}{\lambda(1-\rho)}=\frac{\rho}{\mu(1-\rho)}$$

8) średni czas obsługi (ze wzoru (2.32)):

$$\overline{s}=\frac{1}{\mu} \tag{2.41}$$

9) średni czas pobytu zgłoszenia w systemie (ze wzoru (2.31)):

$$\overline{q}=\frac{1}{\mu(1-\rho)} \tag{2.42}$$

lub z formuły Little'e:

$$\overline{q}=\frac{\overline{n}}{\lambda}=\frac{\rho}{\lambda(1-\rho)}=\frac{\frac{\lambda}{\mu}}{\lambda(1-\rho)}=\frac{1}{\mu(1-\rho)}$$

Przykład: Zadany jest system kolejkowy typu M/M/1/L z tylko jednym miejscem w poczekalni. Obliczyć stacjonarne prawdopodobieństwa stanów systemu oraz średnią liczbę zgłoszeń w kolejce dla $\lambda=1$ i $\mu=2$.

Zacznijmy od analizy grafu stanów systemu:

$$H_0 \overset{\mu_1 \qquad \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2 \qquad \lambda_1}{\longleftrightarrow} H_2$$

kolejne stany sytemu należy interpretować następująco:

H_0–brak zgłoszeń w systemie,
H_1–jedno zgłoszenie na stanowisku obsługi, kolejka pusta,
H_2–jedno zgłoszenie na stanowisku obsługi i jedno w kolejce.
Zakładając, że:

$$\lambda_0 = \lambda_1 = \lambda$$

oraz

$$\mu_1 = \mu_2 = \mu$$

można od razu obliczyć współczynnik obciążenia stanowiska obsługi, oraz wszystkie prawdopodobieństwa stanów systemu:

$$\rho = \frac{\lambda}{\mu} = \frac{1}{2}$$

a

$$p_i = \frac{\rho^i(1-\rho)}{1-\rho^{m+2}} \qquad \text{dla } i = 0,\ 1,\ 2, \cdot \cdot \cdot$$

$$p_0 = \frac{1 \cdot \left(1 - \frac{1}{2}\right)}{1 - \left(\frac{1}{2}\right)^3} = \frac{4}{7}$$

$$p_1 = \frac{\left(\frac{1}{2}\right)^1 \cdot \left(1 - \frac{1}{2}\right)}{1 - \frac{1}{8}} = \frac{2}{7}$$

$$p_2 = \frac{\left(\frac{1}{2}\right)^2 \cdot \left(1 - \frac{1}{2}\right)}{1 - \frac{1}{8}} = \frac{1}{7}$$

oczywiście

$$p_0 + p_1 + p_2 = \frac{4}{7} + \frac{2}{7} + \frac{1}{7} = 1$$

Średnią liczbę zgłoszeń znajdujących się w kolejce obliczmy korzystając ze wzoru (2.23):

$$\overline{v} = \frac{\left(\frac{1}{2}\right)^2 \cdot \left[1 + \left(\frac{1}{2}\right)^1 \cdot \left(1 \cdot \frac{1}{2} - (1+1)\right)\right]}{\left[1 - \left(\frac{1}{2}\right)^3\right]\left(1 - \frac{1}{2}\right)} = \frac{\frac{1}{4} \cdot \left[1 + \frac{1}{2} \cdot \left(-\frac{3}{2}\right)\right]}{\frac{7}{8} \cdot \frac{1}{2}} = \frac{1}{7}$$

lub ze wzoru ogólnego:

$$\overline{v} = 0 \cdot p_1 + 1 \cdot p_2 = 1 \cdot \frac{1}{7} = \frac{1}{7}$$

2.5 Wielostanowiskowy system typu M/M/c/L z oczekiwaniem (kolejką)

System takiego typu składa się z c identycznych stanowisk obsługi oraz poczekalni (kolejki) posiadającej m miejsc($L = m$). Jeżeli c zgłoszeń znajduje się na stanowiskach obsługi i w tym samym czasie w kolejce jest m innych zgłoszeń (poczekalnia pełna), to każde następne przychodzące zgłoszenie jest tracone. Oznacza to, że system zachowuje się jak system ze stratą, gdy w systemie znajduje się dokładnie $m + c$ zgłoszeń. Przeanalizujmy graf stanów takiego systemu:

$$H_0 \overset{\mu_1\ \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2\ \lambda_1}{\longleftrightarrow} \cdot\cdot\cdot \overset{\mu_c\ \lambda_{c-1}}{\longleftrightarrow} H_c \overset{\mu_{c+1}\ \lambda_c}{\longleftrightarrow} \cdot\cdot\cdot \overset{\mu_{m+c}\ \lambda_{m+c-1}}{\longleftrightarrow} H_{m+c}$$

gdzie:

H_0–brak zgłoszeń w systemie,
H_1–jedno zgłoszenia na dowolnym stanowisku obsługi, kolejka pusta,
H_2–dwa zgłoszenia na dowolnych stanowiskach obsługi, kolejka pusta,
H_c–c zgłoszeń na stanowiskach obsługi (wszystkie stanowiska zajęte), kolejka pusta,
H_{c+1} – c zgłoszeń na stanowiskach obsługi i jedno w kolejce,
H_{m+c} – c zgłoszeń na stanowiskach obsługi i m zgłoszeń w kolejce (poczekalnia pełna).
Załóżmy, że:

$$\lambda_0 = \lambda_1 = \lambda_2 = \cdot\ \cdot\ \cdot = \lambda_{m+c-1} = \lambda$$

oraz

$$\mu_1 = \mu$$

jest to intensywność obsługi, gdy na dowolnym stanowisku obsługi (w systemie) znajduje się tylko jedno zgłoszenie,

$$\mu_2 = 2 \cdot \mu$$

intensywność obsługi, gdy na stanowiskach obsługi (dowolnych) są dwa zgłoszenia, itd.

$$\mu_c = c \cdot \mu$$

intensywność obsługi, gdy wszystkie stanowiska są zajęte, a kolejka jest pusta,

$$\mu_{c+1} = c \cdot \mu \ , \quad \cdot \quad \cdot \quad \cdot \ , \ \mu_{m+c} = c \cdot \mu$$

intensywność obsługi, gdy zgłoszenia znajdują się i w poczekalni, oczywiście przy zajętych wszystkich stanowiskach obsługi.

Następnym etapem powinno być obliczenie stacjonarnych prawdopodobie stanów i tak, jeżeli liczba zgłoszeń zawarta jest w następującym przedziale $0 \leq i \leq c$, to

$$Q_i = \frac{\lambda_0 \ \lambda_1 \cdot \ \cdot \ \cdot \ \lambda_{i-1}}{\mu_1 \ \mu_2 \cdot \ \cdot \ \cdot \ \mu_i} = \frac{\lambda^i}{i! \ \mu^i} = \frac{\rho^i}{i!} \tag{2.43}$$

Z kolei, jeżeli liczba zgłoszeń znajdujących się w systemie należy do przedziału $c+1 \leq i \leq c+m$, to

$$Q_i = \frac{\lambda^i}{\underbrace{\mu_1 \ \mu_2 \cdot \ \cdot \ \cdot \ \mu_c}_{c! \ \mu^c} \cdot \underbrace{\mu_{c+1} \cdot \ \cdot \ \cdot \ \mu_i}_{(c \ \mu)^{i-c}}} = \frac{\lambda^i}{c! \ \mu^c \ c^{i-c} \ \mu^{i-c}} =$$

$$= \frac{\lambda^i}{c! \ \mu^i \ c^{i-c}} = \frac{\lambda^i}{c! \ \mu^i \ \frac{c^i}{c^c}} = \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^i \qquad \text{gdzie } \frac{\rho}{c} < 1 \tag{2.44}$$

Znając wartość współczynnika Q_i, można od razu napisać wyrażenie dla stacjonarnego prawdopodobieństwa stanu p_i:

$$p_i = p_0 \cdot Q_i \tag{2.45}$$

pamiętając o tym, że:

$$\sum_{k=0}^{m+c} p_k = 1 = \sum_{k=0}^{m+c} p_0 Q_k \qquad \text{czyli } p_0 = \frac{1}{\sum\limits_{k=0}^{m+c} Q_k} \tag{2.46}$$

więc podstawowe wyrażenie dla p_i równe jest:

$$p_i = \frac{Q_i}{\sum\limits_{k=0}^{m+c} Q_k} \qquad \text{przy } Q_0 = 1 \tag{2.47}$$

Podstawowe charakterystyki systemu:

a) prawdopodobieństwo blokady systemu i straty przychodzącego zgłoszenia (stan H_{m+c}):

$$p_{bl} = p_{m+c} \tag{2.48}$$

dokładniej, jest to prawdopodobieństwo obsługi zgłoszenia znajdującego się na ostatnim miejscu w poczekalni (w kolejce) i jeżeli to zgłoszenie zajmuje ostatnie miejsce w poczekalni, to każde następne przychodzące zgłoszenie jest odrzucane (blokada),

b) prawdopodobieństwo obsługi zgłoszenia (stany systemu od H_0 do H_{m+c-1}):

$$p_{obs} = 1 - p_{bl} = 1 - p_{m+c} = \sum_{i=0}^{m+c-1} p_i \tag{2.49}$$

oznacza to, że przychodzące zgłoszenie zostanie przyjęte na obsługę, jeżeli przynajmniej jest wolne jedno miejsce w poczekalni,

c) średnia liczba zgłoszeń na stanowiskach obsługi (zajętych stanowisk), czyli stany systemu od H_1 do H_{m+c}:

$$\overline{l} = 1 \cdot p_1 + 2 \cdot p_2 + \cdot \ \cdot \ \cdot + c \cdot p_c + c \cdot p_{c+1} + c \cdot p_{c+2} + \cdot \ \cdot \ \cdot + c \cdot p_{c+m} =$$

$$= \sum_{k=1}^{c} k \cdot p_k + \sum_{k=c+1}^{c+m} c \cdot p_k$$

pierwsza część powyższego wyrażenia odnosi się do przypadku, gdy kolejka jest pusta, a zgłoszenia znajdują się tylko na stanowiskach obsługi, druga zaś część, to zgłoszenia znajdują sie i w kolejce (poczekalni), więc wszystkie stanowiska obsługi są permanentnie zajęte, więc:

$$\overline{l} = \sum_{k=1}^{c} k \ \frac{\rho^k}{k!} \ p_0 + \sum_{k=c+1}^{c+m} c \ \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^k p_0 = p_0 \left[\sum_{k=1}^{c} \frac{\rho^k}{(k-1)!} + \frac{c^{c+1}}{c!} \sum_{k=c+1}^{c+m} \left(\frac{\rho}{c}\right)^k\right] \tag{2.50}$$

d) średnia liczba zgłoszeń znajdujących się w kolejce (stany systemu od H_{c+1} do H_{c+m}):

$$\overline{v} = 1 \cdot p_{c+1} + 2 \cdot p_{c+2} + \cdot \ \cdot \ \cdot + m \cdot p_{c+m} = \sum_{v=1}^{m} v \cdot p_{c+v} =$$

$$= \sum_{v=1}^{m} v \ p_0 \ \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^{c+v} = p_0 \ \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^{c+1} \sum_{v=1}^{m} v \ \left(\frac{\rho}{c}\right)^{v-1}$$

do tego wyrażenia można zastosować, tak jak i dla systemu M/M/1/L, wzór na sumę szeregu geometrycznego:

$$\sum_{v=1}^{m} v\ \alpha^{v-1} = \left(\sum_{v=1}^{m} \alpha^{v}\right)' = \frac{1+\alpha^m\left[m\alpha-(m+1)\right]}{(1-\alpha)^2}$$

gdzie

$$\alpha = \frac{\rho}{c} < 1$$

teraz podstawiając to wyrażenie do wzoru na $\overline{v}$ otrzymamy:

$$\overline{v} = p_0\ \frac{c^c}{c!}\ \alpha^{c+1}\left[\frac{1+\alpha^m\left[m\alpha-(m+1)\right]}{(1-\alpha)^2}\right] \tag{2.51}$$

e) średnia liczba zgłoszeń znajdujących sie w systemie (stany systemu od H_1 do H_{c+m}):

$$\overline{n} = 1\cdot p1 + 2\cdot p_2 + \cdots + c\cdot p_c + (c+1)\cdot p_{c+1} + \cdots + (c+m)p_{c+m} =$$

$$= \sum_{k=1}^{c} k\cdot p_k + \sum_{k=c+1}^{c+m} k\cdot p_k = p_0\sum_{k=1}^{c} k\cdot\frac{\rho^k}{k!} + p_0\sum_{k=c+1}^{c+m} k\cdot\frac{c^c}{c!}\left(\frac{\rho}{c}\right)^k =$$

$$= p_0\left[\sum_{k=1}^{c}\frac{\rho^k}{(k-1)!} + \frac{c^c}{c!}\sum_{k=c+1}^{c+m} k\cdot\left(\frac{\rho}{c}\right)^k\right] \tag{2.52}$$

przeanalizujmy oddzielnie drugą część tego wyrażenia

$$\frac{c^c}{c!}\sum_{k=c+1}^{c+m} k\cdot\left(\frac{\rho}{c}\right)^k = \frac{c^c}{c!}\cdot\frac{\rho^c}{c^c}\sum_{k=c+1}^{c+m} k\cdot\left(\frac{\rho}{c}\right)^{k-c} =$$

dla $i = k - c$, oraz $k = i + c$

$$= \frac{\rho^c}{c!}\sum_{i=1}^{m}(i+c)\left(\frac{\rho}{c}\right)^i = \frac{\rho^c}{c!}\left[\sum_{i=1}^{m} i\cdot\left(\frac{\rho}{c}\right)^i + c\sum_{i=1}^{m}\left(\frac{\rho}{c}\right)^i\right] =$$

$$= \frac{\rho^c}{c!}\left[\frac{\rho}{c}\sum_{i=1}^{m} i\cdot\left(\frac{\rho}{c}\right)^{i-1} + c\sum_{i=1}^{m}\left(\frac{\rho}{c}\right)^i\right] =$$

następnie, postępując podobnie jak w punkcie d) oraz stosując wzór na sumę wyrazów postępu geometrycznego, otrzymamy:

$$= \frac{\rho^c}{c!}\left[\alpha\ \frac{1+\alpha^m\left[m\alpha-(m+1)\right]}{(1-\alpha)^2} + c\ \alpha\ \frac{1-\alpha^m}{(1-\alpha)}\right]$$

więc, wyrażenie dla średniej liczby zgłoszeń w systemie przyjmuje następującą postać:

$$\overline{n} = p_0\left\{\sum_{k=1}^{c}\frac{\rho^k}{(k-1)!} + \frac{\rho^c}{c!}\left[\alpha\ \frac{1+\alpha^m\left[m\alpha-(m+1)\right]}{(1-\alpha)^2} + c\ \alpha\ \frac{1-\alpha^m}{(1-\alpha)}\right]\right\} \tag{2.53}$$

średnią liczbę zgłoszeń można również otrzymać z wyrażenia:

$$\overline{n} = \overline{l} + \overline{v} = p_0\left[\sum_{k=1}^{c} k\cdot\frac{\rho^k}{k!} + \frac{c^c}{c!}\sum_{k=c+1}^{c+m} c\left(\frac{\rho}{c}\right)^k + \frac{c^c}{c!}\sum_{k=1}^{m} k\left(\frac{\rho}{c}\right)^{c+k}\right] =$$

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

$$\left.\left. + m\left(\frac{\rho}{c}\right)^{c+m}\right]\right\} =$$

$$= p_0\left\{\sum_{k=1}^{c}\frac{\rho^k}{(k-1)!} + \frac{c^c}{c!}\left[(c+1)\left(\frac{\rho}{c}\right)^{c+1} + (c+2)\left(\frac{\rho}{c}\right)^{c+2} + \cdots\right.\right. =$$

$$\left.\left. + (c+m)\left(\frac{\rho}{c}\right)^{c+m}\right]\right\} =$$

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

f) średni czas oczekiwania w kolejce (kolejka zaczyna się gdy od stanu H_c):

$$\overline{w} = p_c\cdot\frac{1}{c\mu} + p_{c+1}\cdot\frac{2}{c\mu} + \cdots + p_{c+m-1}\cdot\frac{m}{c\mu} + p_{c+m}\cdot 0$$

gdzie, poszczególne wyrazy sumy, oznaczają:

$p_c \cdot \frac{1}{c\mu}$—na stanowiskach obsługi, znajduje się dokładnie c zgłoszeń i każde następne zgłoszenie czeka na zwolnienie pojedynczego (dowolnego) stanowiska $\frac{1}{c\mu}$,

$p_{c+1} \cdot \frac{2}{c\mu}$—w kolejce jest jedno zgłoszenie, więc przychodzące zgłoszenie czeka na zwolnienie dwóch, dowolnych stanowisk obsługi, jednego dla zgłoszenia już znajdującego się w kolejce, drugiego dla siebie, itd.,

$p_{c+m-1} \cdot \frac{m}{c\mu}$—w kolejce jest $m-1$ zgłoszeń, więc przychodzące zgłoszenie czeka $\frac{m}{c\mu}$,

$p_{c+m} \cdot 0$—na stanowiskach obsługi znajduje się c zgłoszeń, a w kolejce jest m, czyli przychodzące zgłoszenie opuszcza system nie obsłużone (blokada systemu), i teraz wyrażenie na średni czas oczekiwania w kolejce, po przekształceniach, przyjmie następującą postać:

$$\overline{w} = \sum_{i=1}^{m} i \cdot \frac{1}{c\mu} \cdot p_{c+i-1} = p_0 \frac{1}{c\mu} \sum_{i=1}^{m} i \cdot \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^{c+i-1} =$$

$$= p_0 \frac{c^c}{c \cdot c!} \frac{1}{\mu} \left(\frac{\rho}{c}\right)^c \sum_{i=1}^{m} i \cdot \left(\frac{\rho}{c}\right)^{i-1}$$

i, postępując podobnie jak przy obliczaniu parametru $\overline{v}$, otrzymamy:

$$\overline{w} = p_0 \frac{c^{c-1}}{c!} \frac{\alpha^c}{\mu} \left[\frac{1 + \alpha^m [m\alpha - (m+1)]}{(1-\alpha)^2}\right] \tag{2.54}$$

gdzie

$$\alpha = \frac{\rho}{c} < 1$$

średni czas oczekiwania w kolejce, można również obliczyć bezpośrednio korzystając z formuły Little'a:

$$\overline{w} = \frac{\overline{v}}{\lambda} = p_0 \frac{c^c}{c!} \frac{\alpha^{c+1}}{\lambda} \left[\frac{1 + \alpha^m [m\alpha - (m+1)]}{(1-\alpha)^2}\right] =$$

$$= p_0 \frac{c^c}{c!} \alpha^c \frac{\frac{\rho}{c}}{\lambda} \left[\frac{1 + \alpha^m [m\alpha - (m+1)]}{(1-\alpha)^2}\right] =$$

$$= p_0 \frac{c^c}{c!} \alpha^c \frac{\frac{\lambda}{\mu}}{c\lambda} \left[\frac{1 + \alpha^m [m\alpha - (m+1)]}{(1-\alpha)^2}\right] =$$

$$= p_0 \frac{c^{c-1}}{c!} \frac{\alpha^c}{\mu} \left[\frac{1 + \alpha^m [m\alpha - (m+1)]}{(1-\alpha)^2}\right]$$

g) średni czas pobytu zgłoszenia w systemie: parametr ten najprościej można obliczyć korzystając ze wzoru Little'a:

$$\overline{q} = \frac{\overline{n}}{\lambda} = \frac{1}{\lambda} p_0 \left[\sum_{k=1}^{c} \frac{\rho^k}{(k-1)!} + \frac{c^c}{c!} \sum_{k=c+1}^{c+m} k \left(\frac{\rho}{c}\right)^k\right] \tag{2.55}$$

h) średni czas obsługi zgłoszenia na stanowisku obsługi (część zgłoszeń jest tracona):

$$\overline{s} = p_{obs} \cdot \frac{1}{\mu} = \frac{1}{\mu} \sum_{i=0}^{c+m-1} p_i = \frac{1}{\mu} p_0 \left[\sum_{i=0}^{c} \frac{\rho^i}{i!} + \sum_{i=c+1}^{c+m-1} \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^i\right] =$$

$$= \frac{1}{\mu} p_0 \left[\sum_{i=0}^{c} \frac{\rho^i}{i!} + \frac{c^c}{c!} \sum_{i=c+1}^{c+m-1} \left(\frac{\rho}{c}\right)^i\right] \tag{2.56}$$

lub prościej

$$\overline{s} = \frac{1}{\mu}(1 - p_{c+m}) = \frac{1}{\mu}\left[1 - p_0 \frac{c^c}{c!} \left(\frac{\rho}{c}\right)^{c+m}\right] \tag{2.57}$$

lub z formuły Little'a:

$$\overline{s} = \frac{\overline{l}}{\lambda} = \frac{1}{\lambda} p_0 \left[\sum_{i=1}^{c} \frac{\rho^i}{(i-1)!} + \frac{c^c}{c!} \sum_{i=c+1}^{c+m} \left(\frac{\rho}{c}\right)^i\right] \tag{2.58}$$

Analiza systemu tego typu upraszcza się, przy założeniu, że liczba miejsc w poczekalni $\to \infty$ (i oczywiście $\frac{\rho}{c} < 1$). Wtedy:

1) prawdopodobieństwa stanów systemu (wzory (2.43), (2.44), (2.45), (2.46), (2.47)):

$$p_0 = \frac{1}{\sum_{i=0}^{c} \frac{\rho^i}{i!} + \frac{c^c}{c!} \sum_{i=c+1}^{\infty} \left(\frac{\rho}{c}\right)^i}$$

przekształcając, tę drugą sumę otrzymamy

$$\frac{c^c}{c!} \sum_{i=c+1}^{\infty} \left(\frac{\rho}{c}\right)^i = \frac{c^c}{c!} \cdot \frac{\rho^c}{c^c} \sum_{i=c+1}^{\infty} \left(\frac{\rho}{c}\right)^{i-c}$$

jeżeli $k = i - c$, to

$$\frac{\rho^c}{c!}\sum_{k=1}^{\infty}\left(\frac{\rho}{c}\right)^k = \frac{\rho^c}{c!}\sum_{k=1}^{\infty}\alpha^k \qquad \text{gdzie } \alpha = \frac{\rho}{c} < 1$$

i po zastosowaniu wzoru na sumę wyrazów postępu geometrycznego, otrzymamy:

$$\frac{\rho^c}{c!}\cdot\frac{\alpha}{1-\alpha}$$

ostatecznie

$$p_0 = \frac{1}{\sum\limits_{k=0}^{c}\left(\frac{\rho^i}{i!}\right) + \frac{\rho^c\alpha}{c!(1-\alpha)}} \tag{2.59}$$

oraz

$$p_i = p_0 \cdot Q_i \tag{2.60}$$

2) prawdopodobieństwo blokady systemu (ze wzoru (2.48)):

$$p_{bl} = 0 \tag{2.61}$$

3) prawdopodobieństwo obsługi zgłoszenia (ze wzoru (2.49)):

$$p_{obs} = 1 \tag{2.62}$$

4) średnia liczba zgłoszeń na stanowiskach obsługi (ze wzoru (2.50)):

$$\overline{l} = p_0\left[\sum_{i=1}^{c}\frac{\rho^i}{(i-1)!} + \frac{c^{c+1}}{c!}\cdot\frac{\rho^c}{c^c}\sum_{i=c+1}^{\infty}\left(\frac{\rho}{c}\right)^{i-c}\right]$$

postępując dalej podobnie jak w punkcie 1) otrzymamy:

$$\overline{l} = \frac{\rho\left[\sum\limits_{i=1}^{c}\left(\frac{\rho^{i-1}}{(i-1)!}\right) + \frac{\rho^{c-1}}{(c-1)!}\cdot\frac{\alpha}{(1-\alpha)}\right]}{\sum\limits_{i=0}^{c}\left(\frac{\rho^i}{i!}\right) + \frac{\rho^c}{c!}\cdot\frac{\alpha}{(1-\alpha)}} =$$

$$= \frac{\rho\left[\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^{c-1}}{(c-1)!} + \frac{\rho^{c-1}}{(c-1)!}\cdot\frac{\alpha}{(1-\alpha)}\right]}{\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^{c-1}}{(c-1)!} + \frac{\rho^c}{c!} + \frac{\rho^c}{c!}\cdot\frac{\alpha}{(1-\alpha)}} =$$

$$= \frac{\rho\left[\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^{c-1}}{(c-1)!} + \frac{\rho}{c}\cdot\frac{\rho^{c-1}}{(c-1)!}\cdot\frac{1}{(1-\alpha)}\right]}{\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^{c-1}}{(c-1)!} + \frac{\rho^c}{c!}\left(\frac{1-\alpha+\alpha}{(1-\alpha)}\right)} = \rho \tag{2.63}$$

5) średnia liczba zgłoszeń w kolejce (ze wzoru (2.51)):

$$\overline{v} = p_0\,\frac{c^c}{c!}\,\alpha^{c+1}\,\frac{1}{(1-\alpha)^2} = p_0\,\frac{c^c}{c!}\,\frac{\alpha^{c+1}}{(1-\alpha)^2} \tag{2.64}$$

6) średnia liczba zgłoszeń znajdujących się w systemie (ze wzoru (2.53)):

$$\overline{n} = p_0\left\{\sum_{i=1}^{c}\left(\frac{\rho^i}{(i-1)!}\right) + \frac{\rho^c}{c!}\left[\frac{\alpha}{(1-\alpha)^2} + \frac{c\,\alpha}{(1-\alpha)}\right]\right\} \tag{2.65}$$

7) średni czas oczekiwania w kolejce (ze wzoru (2.54)):

$$\overline{w} = p_0\,\frac{c^{c-1}}{c!}\,\frac{\alpha^c}{\mu\,(1-\alpha)^2} \tag{2.66}$$

8) średni czas pobytu zgłoszenia w systemie (ze wzoru Little'a):

$$\overline{q} = \frac{\overline{n}}{\lambda} = \frac{p_0}{\lambda}\left\{\sum_{i=1}^{c}\left(\frac{\rho^i}{(i-1)!}\right) + \frac{\rho^c}{c!}\left[\frac{\alpha}{(1-\alpha)^2} + \frac{c\,\alpha}{(1-\alpha)}\right]\right\} \tag{2.67}$$

9) średni czas obsługi (ze wzoru (2.57)):

$$\overline{s} = \frac{1}{\mu} \tag{2.68}$$

Przykład: Zadany jest system kolejkowy typu M/M/c/L, w którym: $c = 2$, $L = 2$, $\lambda = 1$, $\mu = 2$. Obliczyć średni czas obsługi zgłoszenia, średni czas pobytu zgłoszenia w kolejce oraz średni czas pobytu w systemie.

Analiza grafu stanów systemu:

$$H_0 \overset{\mu_1 \qquad \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2 \qquad \lambda_1}{\longleftrightarrow} H_2 \overset{\mu_3 \qquad \lambda_2}{\longleftrightarrow} H_3 \overset{\mu_4 \qquad \lambda_3}{\longleftrightarrow} H_4$$

gdzie:

H_0- brak zgłoszeń w systemie,

H_1- jedno zgłoszenie na stanowisku obsługi (dowolnym),

H_2- dwa zgłoszenia na stanowiskach obsługi (wszystkie stanowiska zajęte),

H_3- jedno zgłoszenie w poczekalni (wszystkie stanowiska zajęte),

H_4- dwa zgłoszenia w poczekalni, blokada systemu.

Obliczmy wielkości Q_i dla $0 \leq i \leq 2$ (ze wzoru (2.43)):

$$Q_i = \frac{\rho^i}{i!}$$

$$Q_0 = 1$$

$$Q_1 = \frac{\rho^1}{1!} = \frac{\lambda}{\mu} = \frac{1}{2}$$

$$Q_2 = \frac{\rho^2}{2!} = \frac{\left(\frac{1}{2}\right)^2}{2} = \frac{1}{8}$$

oraz dla $3 \leq i \leq 4$ (ze wzoru (2.44)):

$$Q_i = \frac{c^c}{c!}\left(\frac{\rho}{c}\right)^i$$

$$Q_3 = \frac{2^2}{2!}\left(\frac{\frac{1}{2}}{2}\right)^3 = 2 \cdot \left(\frac{1}{4}\right)^3 = \frac{1}{32}$$

$$Q_4 = \frac{2^2}{2!}\left(\frac{1}{4}\right)^4 = \frac{2}{256} = \frac{1}{128}$$

stąd

$$p_0 = \frac{1}{\sum_{i=0}^{c+m} Q_i} = \frac{1}{1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128}} = \frac{1}{\frac{213}{128}} = \frac{128}{213}$$

oraz, gdy

$$p_i = p_0 \cdot Q_i$$

$$p_1 = \frac{128}{213} \cdot \frac{1}{2} = \frac{64}{213}$$

$$p_2 = \frac{128}{213} \cdot \frac{1}{8} = \frac{16}{213}$$

$$p_3 = \frac{128}{213} \cdot \frac{1}{32} = \frac{4}{213}$$

$$p_4 = \frac{128}{213} \cdot \frac{1}{128} = \frac{1}{213}$$

Sprawdźmy teraz sumę prawdopodobieństw:

$$\sum_{i=0}^{c+m} p_i = 1, \qquad \text{więc } \frac{128}{213} + \frac{64}{213} + \frac{16}{213} + \frac{4}{213} + \frac{1}{213} = \frac{213}{213} = 1$$

Obliczmy teraz średni czas obsługi (ze wzoru (2.56)):

$$\overline{s} = \frac{1}{2} \cdot \frac{128}{213}\left[\frac{\left(\frac{1}{2}\right)^0}{1} + \frac{\left(\frac{1}{2}\right)^1}{1} + \frac{\left(\frac{1}{2}\right)^2}{2} + \frac{2^2}{1 \cdot 2}\left(\frac{\frac{1}{2}}{2}\right)^3\right] =$$

$$= \frac{64}{213}\left(1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{32}\right) = \frac{106}{213}$$

lub ze wzoru (2.57):

$$\overline{s} = \frac{1}{2}\left(1 - \frac{1}{213}\right) = \frac{106}{213}$$

lub ze wzoru (2.58):

$$\overline{s} = \frac{1}{1} \cdot \frac{128}{213}\left\{\frac{\left(\frac{1}{2}\right)^1}{1} + \frac{\left(\frac{1}{2}\right)^2}{1} + \frac{2^3}{1 \cdot 2}\left[\left(\frac{\frac{1}{2}}{2}\right)^3 + \left(\frac{\frac{1}{2}}{2}\right)^4\right]\right\} =$$

$$= \frac{128}{213}\left(\frac{1}{2} + \frac{1}{4} + \frac{4}{64} + \frac{4}{256}\right) = \frac{128}{213}\left(\frac{64 + 32 + 8 + 2}{128}\right) = \frac{106}{213}$$

jeszcze inny sposób obliczenia średniego czasu wymaga najpierw obliczenia $\overline{l}$ (patrz punkt c):

$$\overline{l} = 1 \cdot p_1 + 2 \cdot p_2 + 2 \cdot p_3 + 2 \cdot p_4 = \frac{64}{213} + \frac{32}{213} + \frac{8}{213} + \frac{2}{213} = \frac{106}{213}$$

$$\overline{s} = \frac{\overline{l}}{\lambda} = \frac{106}{1 \cdot 213} = \frac{106}{213}$$

Średni czas w kolejce (ze wzoru (2.54)):

$$\overline{w} = \frac{128}{213} \cdot \frac{2^1}{2} \cdot \frac{\left(\frac{1}{4}\right)^2}{2} \left\{ \frac{1 + \left(\frac{1}{4}\right)^2 \left[2 \cdot \frac{1}{4} - (2+1)\right]}{\left(1 - \frac{1}{4}\right)^2} \right\} =$$

$$= \frac{128}{213} \cdot \frac{1}{32} \left[\frac{1 + \frac{1}{16}(-\frac{5}{2})}{\frac{9}{16}} \right] = \frac{4}{213} \cdot \frac{3}{2} = \frac{6}{213}$$

Średni czas pobytu zgłoszenia w systemie (ze wzoru (2.55)):

$$\overline{q} = \frac{1}{1} \cdot \frac{128}{213} \left\{ \frac{\left(\frac{1}{2}\right)^1}{1} + \frac{\left(\frac{1}{2}\right)^2}{1} + \frac{2^2}{2} \left[3 \cdot \left(\frac{1}{4}\right)^3 + 4 \cdot \left(\frac{1}{4}\right)^4 \right] \right\} =$$

$$= \frac{128}{213} \left[\frac{64 + 32 + 12 + 4}{128} \right] = \frac{128}{213} \cdot \frac{112}{128} = \frac{112}{213}$$

czas ten można również obliczyć z wyrażenia:

$$\overline{q} = \overline{w} + \overline{s} = \frac{6}{213} + \frac{106}{213} = \frac{112}{213}$$

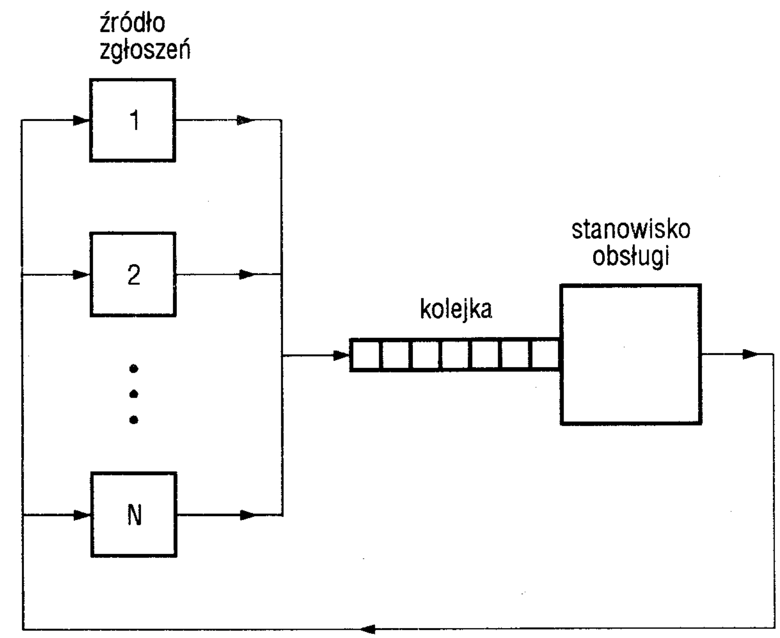
2.6 Jednostanowiskowy system kolejkowy ze skończonym wymiarowo źródłem zgłoszeń (typu M/M/1/N)

W tego typu systemach intensywność napływu zgłoszeń zależy od stanu systemu, tj. od liczby zgłoszeń znajdujących się w kolejce i na stanowisku obsługi. Klasycznym przykładem takiego systemu jest tzw. problem konserwatora, gdzie jeden konserwator naprawia (obsługuje) N obrabiarek. Innym przykładem, bliżej związanym z modelowaniem systemów komputerowych, jest wielodostępny system komputerowy, w którym z jednoprocesorowym komputerem połaczonych jest N niezależnyc terminali:

$$H_0 \overset{\mu_1 \quad \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2 \quad \lambda_1}{\longleftrightarrow} H_2 \overset{\mu_3 \quad \lambda_2}{\longleftrightarrow} \cdots \overset{\mu_{N-1} \quad \lambda_{N-2}}{\longleftrightarrow} H_{N-1} \overset{\mu_N \quad \lambda_{N-1}}{\longleftrightarrow} H_N$$

gdzie:

H_0 - zero niesprawnych (zepsutych) obrabiarek,
H_1 - jedna niesprawna obrabiarka,
H_2 - dwie niesprawne obrabiarki, itd.,
H_N - wszystkie obrabiarki zepsute.



Rys. 4. Przykład jednostanowiskowego systemu obsługi ze skończonym źródłem

Graf stanów systemu (na przykładzie problemu konserwatora):

Przeanalizujmy intensywność napływu zgłoszeń (sygnałów o niesprawnościa obrabiarek) w takim systemie:

$$\lambda_0 = N\,\lambda \tag{2.69}$$

oznacza to, że od N obrabiarek, może przyjść sygnał o awarii, gdzie $\frac{1}{\lambda}$ - wartość średnia przedziału czasu, w którym pojedyncza obrabiarka pracuje bez naprawy (średni czas między dwiema kolejnymi naprawami). Kolejne intensywności można obliczyć w następujący sposób:

$$\lambda_1 = (N-1)\ \lambda, \quad \cdots, \quad \lambda_i = (N-i)\lambda, \quad \cdots, \quad \lambda_{N-1} = \lambda \qquad (2.70)$$

Przy pojedynczym stanowisku obsługi:

$$\mu_1 = \mu_2 = \cdots = \mu_N = \mu \qquad (2.71)$$

gdzie μ—intensywność obsługi.

Znając intensywności strumienia wejściowego, oraz intensywność obsługi, możemy obliczyć wartość parametru Q_i (dla $0 \le i \le N$):

$$Q_i = \frac{\lambda_0\ \lambda_1 \cdots \lambda_{i-1}}{\mu_1\ \mu_2 \cdots \mu_i} = \frac{N\lambda\ (N-1)\lambda \cdots (N-i+1)\lambda}{\mu^i} = \frac{N!}{(N-i)!}\ \rho^i \qquad (2.72)$$

potem

$$p_0 = \frac{1}{\sum_{k=0}^{N} Q_k} \qquad (2.73)$$

oraz

$$p_i = p_0\ Q_i \qquad \text{dla } 0 \le i \le N \qquad (2.74)$$

Podstawowe charakterystyki systemu:

a) prawdopodobieństwo zajętości stanowiska obsługi (konserwatora), tj. stany systemu od H_1 do H_N, gdy konserwator zajęty jest naprawą:

$$p_z = p_1 + p_2 + \cdots + p_N = 1 - p_0 \qquad (2.75)$$

b) średnia liczba obrabiarek wymagających naprawy (łącznie z naprawianą):

$$\overline{n} = 1 \cdot p_1 + 2 \cdot p_2 + \cdots + N \cdot p_N = \sum_{i=1}^{N} i \cdot p_i =$$

$$= \sum_{i=1}^{N} i\ \frac{N!}{(N-i)!}\ \rho^i\ p_0 = p_0\ N! \sum_{i=1}^{N} \frac{i}{(N-i)!}\ \rho^i \qquad (2.76)$$

c) średnia liczba naprawianych obrabiarek:

$$\overline{l} = 0 \cdot p_0 + 1 \cdot p_1 + 1 \cdot p_2 + \cdots + 1 \cdot p_N = 1 \cdot p_z = 1 - p_0 \qquad (2.77)$$

d) średnia liczba obrabiarek oczekujących naprawy w kolejce (kolejka zaczyna się od stanu H_2):

$$\overline{v} = 1 \cdot p_2 + 2 \cdot p_3 + \cdots + (N-1) \cdot p_N = \sum_{i=1}^{N-1} i \cdot p_{i+1} =$$

$$= \sum_{i=1}^{N-1} i \cdot p_0\ \frac{N!}{(N-i-1)!} \rho^{i+1} = p_0\ N! \sum_{i=1}^{N-1} \frac{i\ \rho^{i+1}}{(N-i-1)!} \qquad (2.78)$$

wielkość tę można też obliczyć trochę inaczej:

$$\overline{v} = \overline{n} - \overline{l} = p_0\ N! \sum_{i=1}^{N} \frac{i\ \rho^i}{(N-i)!} - 1 + p_0$$

Dla obliczenia parametrów czasowych systemu konieczne jest obliczenie średniej intensywności napływu zgłoszeń. Jeżeli w danej chwili w systemie jest i jednostek, to od $(N-i)$, znajdujących się poza systemem, mogą napływać następne zgłoszenia, z intensywnością λ każde, czyli łączna intensywność dopływu nowych zgłoszeń równa jest w tym przypadku $(N-1)\lambda$. Sumując to ostatnie wyrażenie po wszystkich wartościach i oraz wprowadzając czynnik wag, równy p_i, otrzymamy:

$$\lambda' = \sum_{i=0}^{N} (N-i)\ \lambda\ p_i =$$

$$= \lambda\ N \sum_{i=0}^{N} p_i - \lambda \sum_{i=0}^{N} i\ p_i = \lambda(N - \overline{n}) \qquad (2.79)$$

e) średni czas oczekiwania w kolejce do naprawy (ze wzoru Little'a):

$$\overline{w} = \frac{\overline{v}}{\lambda'} = \frac{\overline{v}}{\lambda(N-\overline{n})} \qquad (2.80)$$

f) średni czas awarii obrabiarki (naprawa plus oczekiwanie w kolejce):

$$\overline{q} = \frac{\overline{n}}{\lambda'} = \frac{\overline{n}}{\lambda(N-\overline{n})} \qquad (2.81)$$

lub jako

$$\overline{q} = \overline{w} + \frac{1}{\mu}$$

Przykład: Dany jest system kolejkowy typu M/M/1/N, w którym $N = 4, \lambda = 1, \mu = 2$. Obliczyć podstawowe charakterystyki systemu.

Pierwszym etapem rowziązania jest obliczenie prawdopodobieństw stanów systemu:

$$Q_i = \frac{N!}{(N-i)!}\,\rho^i \qquad \text{dla } 0 \le i \le N \qquad \text{gdzie: } \rho = \frac{\lambda}{\mu} = \frac{1}{2}$$

$$Q_0 = 1$$

$$Q_1 = \frac{4!}{3!}\,\rho^1 = 4 \cdot \frac{1}{2} = 2$$

$$Q_2 = \frac{4!}{2!}\,\rho^2 = 12 \cdot \frac{1}{4} = 3$$

$$Q_3 = \frac{4!}{1!}\,\rho^3 = 24 \cdot \frac{1}{8} = 3$$

$$Q_4 = \frac{4!}{0!}\,\rho^4 = 24 \cdot \frac{1}{16} = \frac{3}{2}$$

oraz

$$p_0 = \frac{1}{\sum\limits_{k=0}^{4} Q_k} = \frac{1}{1+2+3+3+\frac{3}{2}} = \frac{2}{21}$$

więc

$$p_i = p_0\, Q_i$$

$$p_1 = \frac{2}{21} \cdot 2 = \frac{4}{21}$$

$$p_2 = \frac{2}{21} \cdot 3 = \frac{6}{21}$$

$$p_3 = \frac{2}{21} \cdot 3 = \frac{6}{21}$$

$$p_4 = \frac{2}{21} \cdot \frac{3}{2} = \frac{3}{21}$$

$$\sum_{i=0}^{4} p_i = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} + \frac{6}{21} + \frac{3}{21} = \frac{21}{21} = 1$$

Średnia liczba obrabiarek wymagających naprawy (ze wzoru (2.76)):

$$\overline{n} = \frac{2}{21}\,4!\left[\frac{1}{3!}\cdot\left(\frac{1}{2}\right)^1 + \frac{2}{2!}\cdot\left(\frac{1}{2}\right)^2 + \frac{3}{1!}\cdot\left(\frac{1}{2}\right)^3 + \frac{4}{0!}\cdot\left(\frac{1}{2}\right)^4\right] =$$

$$= \frac{48}{21}\left(\frac{1}{12} + \frac{1}{4} + \frac{3}{8} + \frac{4}{16}\right) = \frac{48}{21}\left(\frac{4+12+18+12}{48}\right) = \frac{46}{21}$$

sprawdzenie

$$\overline{n} = 1 \cdot \frac{4}{21} + 2 \cdot \frac{6}{21} + 3 \cdot \frac{6}{21} + 4 \cdot \frac{3}{21} = \frac{4+12+18+12}{21} = \frac{46}{21}$$

Średnia liczba obrabiarek będących w naprawie (ze wzoru (2.77)):

$$\overline{l} = 1 - p_0 = 1 - \frac{2}{21} = \frac{19}{21}$$

Średnia liczba obrabiarek oczekujących w kolejce do naprawy (ze wzoru (2.78)):

$$\overline{v} = \frac{2}{21}\,4!\left[\frac{1}{2!}\cdot\left(\frac{1}{2}\right)^2 + \frac{2}{1!}\cdot\left(\frac{1}{2}\right)^3 + \frac{3}{0!}\cdot\left(\frac{1}{2}\right)^4\right] =$$

$$= \frac{48}{21}\left(\frac{1}{8} + \frac{2}{8} + \frac{3}{16}\right) = \frac{48}{21}\cdot\frac{9}{16} = \frac{27}{21}$$

sprawdzenie

$$\overline{v} = 1 \cdot p_2 + 2 \cdot p_3 + 3 \cdot p_4 = 1 \cdot \frac{6}{21} + 2 \cdot \frac{6}{21} + 3 \cdot \frac{3}{21} = \frac{27}{21}$$

również

$$\overline{v} = \overline{n} - \overline{l} = \frac{46}{21} - \frac{19}{21} = \frac{27}{21}$$

Średnia intensywność napływu zgłoszeń (ze wzoru (2.79)):

$$\lambda' = 1 \cdot \left(4 - \frac{46}{21}\right) = \frac{38}{21}$$

Średni czas oczekiwania w kolejce (ze wzoru (2.80)):

$$\overline{w} = \frac{\frac{27}{21}}{\frac{38}{21}} = \frac{27}{38}$$

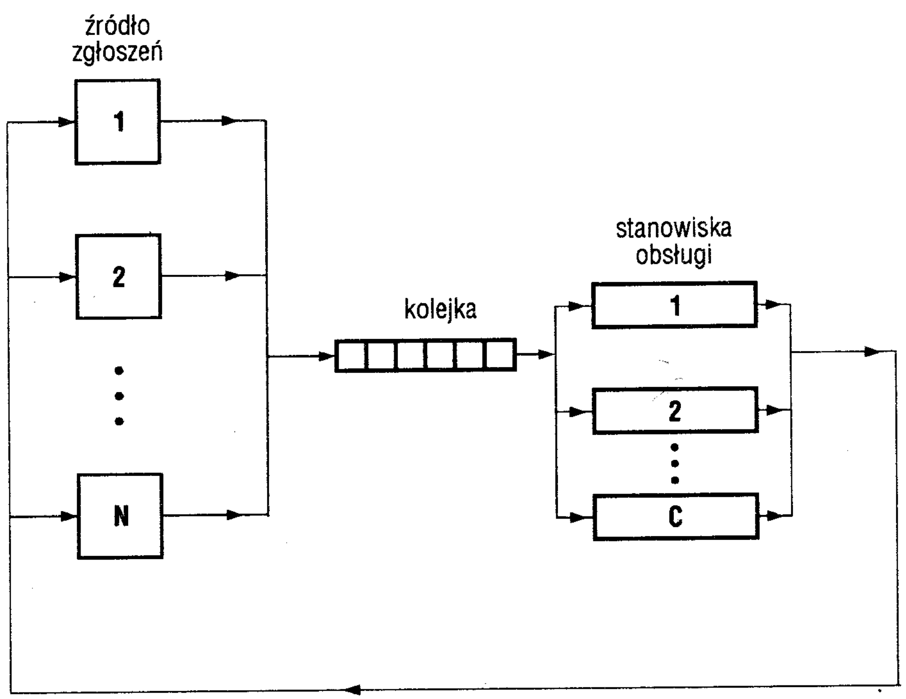
Średni czas naprawy i oczekiwania na naprawę (ze wzoru (2.81)):

$$\overline{q} = \frac{\frac{46}{21}}{\frac{38}{21}} = \frac{46}{38}$$

lub jako

$$\overline{q} = \overline{w} + \frac{1}{\mu} = \frac{27}{38} + \frac{1}{2} = \frac{46}{38}$$

2.7 Wielostanowiskowy system kolejkowy ze skończonym wymiarowo źródłem zgłoszeń (typu M/M/c/N)



Rys. 5. Wielostanowiskowy system kolejkowy typu M/M/c/N

Przykładem tego typu systemu może być, jak dla systemu M/M/1/N, problem obsługi technicznej hali obrabiarek, gdzie c konserwatorów obsługuje N obrabiarek, a przy tym, w danym momencie, pojedynczy konserwator może naprawiać tylko jedną obrabiarkę. Zadany system kolejkowy można przedstawić w taki sposób jak na rys. 5.

Graf stanów systemu:

$$H_0 \overset{\mu_1\ \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2\ \lambda_1}{\longleftrightarrow} \cdots \overset{\mu_c\ \lambda_{c-1}}{\longleftrightarrow} H_c \overset{\mu_{c+1}\ \lambda_c}{\longleftrightarrow} \cdots \overset{\mu_N\ \lambda_{N-1}}{\longleftrightarrow} H_N$$

gdzie:

H_0 - zero obrabiarek do naprawy,

H_1 - jedna obrabiarka w naprawie,

H_c - c obrabiarek w naprawie (wszyscy konserwatorzy zajęci),

H_{c+1} - c obrabiarek w naprawie, jedna czeka w kolejce,

H_N - c obrabiarek w naprawie, $N-c$ w kolejce, czyli wszystkie obrabiarki zepsute.

Intensywności napływu zgłoszeń:

$$\lambda_0 = N \cdot \lambda$$

czyli od każdej z N obrabiarek może przyjść sygnał o awarii, a $\frac{1}{\lambda}$ - średni czas pracy bez awarii. Podobnie:

$$\lambda_1 = (N-1)\lambda, \cdots, \lambda_i = (N-i)\lambda, \cdots, \lambda_{N-1} = \lambda$$

oraz

$$\mu_1 = \mu,\ \mu_2 = 2\mu, \cdots, \mu_c = c\mu,\ \mu_{c+1} = c\mu, \cdots, \mu_N = c\mu$$

Dla obliczenia stacjonarnych prawdopodobieństw stanu systemu konieczne jest obliczenie parametru Q, który dla $0 \leq i \leq c$ równy jest:

$$Q_i = \frac{\lambda_0 \cdot \lambda_1 \cdot \cdots \lambda_{i-1}}{\mu_1 \cdot \mu_2 \cdot \cdots \mu_i} = \frac{N\lambda \cdot (N-1)\lambda \cdot \cdots (N-i+1)\lambda}{\mu \cdot 2\mu \cdot \cdots i\mu} =$$

$$= \frac{\frac{N!}{(N-i)!}\lambda^i}{i!\ \mu^i} = \frac{N!}{(N-i)!\ i!}\ \rho^i \qquad (2.82)$$

a dla $c+1 \le i \le N$

$$Q_i = \frac{\lambda_0 \cdot \lambda_1 \cdot \cdot \cdot \lambda_{i-1}}{\underbrace{\mu_1 \cdot \mu_2 \cdot \cdot \cdot \mu_c}_{c!\ \mu^c} \cdot \underbrace{\mu_{c+1} \cdot \cdot \cdot \mu_i}_{(c\ \mu)^{i-c}}} = \frac{\frac{N!}{(N-i)!}\ \lambda^i}{c!\ \mu^c\ c^{i-c}\ \mu^{i-c}} = \frac{N!}{(N-i)!\ c!\ c^{i-c}}\ \rho^i$$

$$= \frac{N!}{(N-i)!\ c!\ \frac{c^i}{c^c}}\ \rho^i = \frac{c^c\ N!}{(N-i)!\ c!}\ \left(\frac{\rho}{c}\right)^i = \frac{c^c\ N!}{(N-i)!\ c!}\ \alpha^i \qquad (2.83)$$

gdzie

$$\alpha = \frac{\rho}{c} < 1$$

stąd

$$p_0 = \frac{1}{\sum\limits_{k=0}^{N} Q_k} = \frac{1}{N! \sum\limits_{k=0}^{c} \frac{\rho^k}{(N-k)!\ k!} + \frac{N!\ c^c}{c!} \sum\limits_{k=c+1}^{N} \frac{\alpha^k}{(N-k)!}} \qquad (2.84)$$

oraz

$$p_i = p_0\ Q_i \qquad (2.85)$$

Podstawowe charakterystyki systemu:

a) prawdopodobieństwo zajętości wszystkich stanowisk obsługi (konserwatorów) - stany systemu od H_c do H_N:

$$p_z = 1 - \sum_{i=0}^{c-1} p_i = 1 - N! \sum_{i=0}^{c-1} \frac{\rho^i}{(N-i)!\ i!} \qquad (2.86)$$

b) średnia liczba obrabiarek wymagających naprawy (naprawianych i czekających w kolejce) - stany systemu od H_1 do H_N:

$$\overline{n} = 1 \cdot p_1 + 2 \cdot p_2 + \cdot \cdot \cdot + N \cdot p_N = \sum_{i=1}^{c} i \cdot p_i + \sum_{i=c+1}^{N} i \cdot p_i =$$

$$= p_0\ N! \left[\sum_{i=1}^{c} \frac{\rho^i}{(N-i)!\ (i-1)!} + \frac{c^c}{c!} \sum_{i=c+1}^{N} \frac{i \cdot \alpha^i}{(N-i)!} \right] \qquad (2.87)$$

c) średnia liczba obrabiarek oczekujących w kolejce do naprawy (stany systemu od H_{c+1} do H_N - gdy wszyscy konserwatorzy zajęci są naprawą):

$$\overline{v} = 1 \cdot p_{c+1} + 2 \cdot p_{c+2} + \cdot \cdot \cdot + (N-c) \cdot p_N = \sum_{i=c+1}^{N} (i-c)\ p_i =$$

$$= \sum_{i=c+1}^{N} (i-c)\ \frac{N!\ c^c}{(N-i)!\ c!}\ \alpha^i\ p_0 = p_0\ \frac{N!\ c^c}{c!} \sum_{i=c+1}^{N} \frac{(i-c)\ \alpha^i}{(N-i)!} \qquad (2.88)$$

d) średnia liczba naprawianych obrabiarek (konserwatorów zajętych naprawą):

$$\overline{l} = 1 \cdot p_1 + 2 \cdot p_2 + \cdot \cdot \cdot + c \cdot p_c + c \cdot p_{c+1} + c \cdot p_{c+2} + \cdot \cdot \cdot + c \cdot p_N =$$

$$= \sum_{i=1}^{c} i \cdot p_i + \sum_{i=c+1}^{N} c \cdot p_i =$$

$$= p_0\ N! \left[\sum_{i=1}^{c} \frac{\rho^i}{(N-i)!\ (i-1)!} + \frac{c^{c+1}}{c!} \sum_{i=c+1}^{N} \frac{\alpha^i}{(N-i)!} \right] \qquad (2.89)$$

lub prościej

$$\overline{l} = \overline{n} - \overline{v}$$

Następnym etapem powinno być obliczenie parametrów czasowych dla zadanego systemu i tutaj, postępując podobnie jak przy analizie systemu M/M/1/N, należy obliczyć średnią intensywność napływu sygnałów o awarii do systemu, tj. średnią intensywność zgłoszeń:

$$\lambda' = \sum_{i=0}^{N} (N-i)\ \lambda\ p_i =$$

$$= N\ \lambda \sum_{i=0}^{N} p_i - \lambda \sum_{i=0}^{N} i \cdot p_i = \lambda\ (N - \overline{n}) \qquad (2.90)$$

e) średni czas oczekiwania w kolejce do naprawy:

$$\overline{w} = \frac{\overline{v}}{\lambda'} = \frac{\overline{v}}{\lambda\ (N - \overline{n})} \qquad (2.91)$$

f) średni czas awarii (oczekiwania na naprawę i naprawy):

$$\overline{q} = \frac{\overline{n}}{\lambda'} = \frac{\overline{n}}{\lambda\ (N - \overline{n})} \qquad (2.92)$$

lub

$$\overline{q} = \overline{w} + \frac{1}{\mu}$$

2.8 Cykliczne systemy kolejkowe (model Palma)

System tego typu daje się również opisać tzw. procesem narodzin i śmierci, chociaż graf stanów systemu wygląda trochę inaczej:

$$\xrightarrow{\lambda_n} H_0 \xrightarrow{\lambda_0} H_1 \xrightarrow{\lambda_1} \cdots \xrightarrow{\lambda_{n-1}} H_n \xrightarrow{\lambda_n}$$

Podstawowy układ równań, z niewiadomymi prawdopodobieństwami p_i, dla takiego systemu wygląda następująco:

$$\left\{ \begin{array}{lcl} \lambda_n\, p_n & = & \lambda_0\, p_0 \\ \lambda_0\, p_0 & = & \lambda_1\, p_1 \\ \ldots & \ldots & \ldots \\ \lambda_{n-1}\, p_{n-1} & = & \lambda_n\, p_n \end{array} \right. \tag{2.93}$$

skąd

$$p_n = \frac{\lambda_0}{\lambda_n}\, p_0$$

$$p_1 = \frac{\lambda_0}{\lambda_1}\, p_0$$

$$p_2 = \frac{\lambda_1}{\lambda_2}\, p_1 = \frac{\lambda_1}{\lambda_2} \cdot \frac{\lambda_0}{\lambda_1}\, p_0 = \frac{\lambda_0}{\lambda_2}\, p_0$$

itd., czyli

$$p_i = \frac{\lambda_0}{\lambda_i}\, p_0 \tag{2.94}$$

Dalej, wychodząc ze wzoru na sumę prawdopodobieństw, można obliczyć p_0:

$$\sum_{i=0}^{n} p_i = 1 = \sum_{i=0}^{n} \frac{\lambda_0}{\lambda_i}\, p_0 = p_0 \sum_{i=0}^{n} \frac{\lambda_0}{\lambda_i}$$

czyli

$$p_0 = \frac{1}{\lambda_0 \sum\limits_{i=0}^{n} \frac{1}{\lambda_i}} \tag{2.95}$$

i dalej

$$p_i = \frac{\lambda_0}{\lambda_i}\, p_0 = \frac{\lambda_0}{\lambda_i} \cdot \frac{1}{\lambda_0 \sum\limits_{k=0}^{n} \frac{1}{\lambda_k}} = \frac{1}{\lambda_i \sum\limits_{k=0}^{n} \frac{1}{\lambda_k}} \tag{2.96}$$

Wprowadzając nowy parametr $\overline{t_i}$ (dla $0 \leq i \leq n$) czyli wartość średnią przedziału czasu miedzy przejściem systemu ze stanu H_i do stanu H_{i+1}, inaczej jest to średni czas pobytu systemu w stanie H_i, który równy jest:

$$\overline{t_i} = \frac{1}{\lambda_i}$$

można zapisać nowe wyrażenie dla prawdopodobieństwa p_i:

$$p_i = \frac{\overline{t_i}}{\sum\limits_{k=0}^{n} \overline{t_k}} \qquad \text{dla } 0 \leq i \leq n \tag{2.97}$$

Jest to system z kumulacją awarii. Załóżmy, że mamy N urządzeń, które po pewnym czasie ulegają awariom i kiedy liczba awarii równa jest np. n, wtedy rozpoczyna się remont (np. sieć kabli telekomunikacyjnych, jakiś ciąg technologiczny itp.). Wtedy, przy założeniu że $\frac{1}{\lambda}$ - jest to wartość średnia przedziału czasu, miedzy awariami pojedyńczego urządzenia otrzymamy:

$$\lambda_0 = N\, \lambda$$

$$\lambda_1 = (N-1)\, \lambda$$

$$\lambda_{n-1} = (N-n-1)\, \lambda$$

$$\lambda_n = [(N-n)\, \lambda + \mu_n] \tag{2.98}$$

gdzie, μ_n - intensywność naprawy n awarii. Problem ten, jest często zaliczany do teorii niezawodności.

2.9 Metoda faz Erlanga

Innymi systemami, które można opisywać procesem narodzin i śmierci, są systemy, w których obok rozkładów wykładniczych, występują tzw. rozkłady Erlanga lub wyłącznie rozkłady Erlanga. Jako przykład może tu służyć system obsługi fazowej, w którym każde zgłoszenie przechodzi

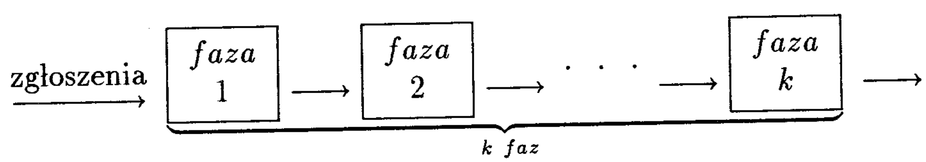
k faz, przy czym czas obsługi w każdej z tych faz ma jednakowy rozkład wykładniczy z parametrem μ (system typu M/E$_k$/1). Przy obsłudze fazowej, między kolejnymi fazami, kolejki nie występują i we wszystkich k fazach znajdować się może, co najwyżej jedno zgłoszenie. Taka adaptacja systemu oparta jest na twierdzeniu, że suma k niezależnych zmiennych losowych, o jednakowym rozkładzie wykładniczym z parametrem μ, ma rozkład Erlanga stopnia k.

Przypomnijmy, że zmienna losowa ma rozkład Erlanga stopnia k z parametrem $\lambda > 0$, gdy jej funkcja gęstości f ma postać:

$$f(t) = \frac{\lambda^k \, t^{k-1}}{(k-1)!} e^{-\lambda \, t} \qquad \text{dla } t \geq 0 \tag{2.99}$$

jeżeli $k = 1$, to otrzymamy gęstość rozkładu wykładniczego.

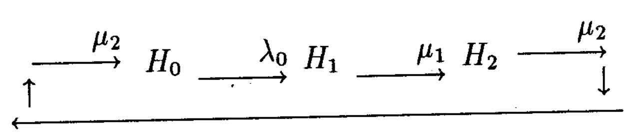
Metodę faz, dla obsługi pojedynczego zgłoszenia, można zilustrować w następujący sposób:



Zastosowanie tej metody sprowadza się do aproksymowania praktycznie spotykanych rozkładów, kombinacją (średnie ważone) rozkładów Erlanga. Pozwala to na badanie systemów o praktycznie dowolnie skomplikowanych rozkładach czasu obsługi i odstępów czasu między zgłoszeniami.

Przykład: Dany jest system obsługi składający się z pojedynczego stanowiska z blokadą, do którego dopływa strumień zgłoszeń od źródła o pojemności N. Zgłoszenia przychodzą do systemu w sposób niezależny od siebie, wg rozkładu wykładniczego z parametrem λ. Czas obsługi zgłoszeń na stanowisku obsługi ma rozkład Erlanga stopnia 2 (system typu M/E$_2$/1/N).

Graf stanów systemu:



gdzie:

H_0 - brak zgłoszeń na stanowisku obsługi,
H_1 - stanowisko obsługi zajęte, pierwsza faza obsługi,
H_2 - stanowisko obsługi zajęte, druga faza obgsługi.

Jest to inna wersja następującego grafu:

$$H_0 \; \overset{\mu_1 \qquad \lambda_0}{\longleftrightarrow} \; H_1$$

w którym stan H_1 zamieniony został przez obsługę w dwóch fazach.

Następnym etapem powinno być obliczenie intensywności:

$$\lambda_0 = N \cdot \lambda$$

ponieważ zgłoszenia przychodzą od N niezależnych źródeł,

$$\mu_1 = \mu_2 = \mu$$

Podstawowy układ równań dla stacjonarnych prawdopodobieństw stanów p_i przyjmuje następującą postać:

$$\begin{cases} N \, \lambda \, p_0 & = & \mu \, p_2 \\ N \cdot \lambda \, p_0 & = & \mu \, p_1 \\ \mu \, p_1 & = & \mu \, p_2 \end{cases} \tag{2.100}$$

plus równanie sumy prawdopodobieństw

$$p_0 + p_1 + p_2 = 1$$

i w tym przypadku rozwiązaniami są parametry:

$$p_0 = \left[1 + 2 \, \frac{N \, \lambda}{\mu}\right]^{-1} = p_{obs} \tag{2.101}$$

$$p_1 = p_2 = \frac{N \, \lambda}{\mu} \, p_0 \tag{2.102}$$

$$p_{bl} = p_1 + p_2 = \frac{2 \, N \, \lambda}{\mu} \cdot \frac{1}{\frac{\mu + 2 \, N \, \lambda}{\mu}} = \frac{2 \, N \, \lambda}{\mu + 2 \, N \, \lambda} \tag{2.103}$$

2.10 Zadania

Zadanie 1. W laboratorium, otwartym przez 8 godzin, zainstalowano dwa pełne zestawy komputerowe, przypadające na grupę 5 studentów. Przeciętnie do laboratorium przychodzi 2 dziennie studentów, którzy pracują średnio po 4 godziny. Wybrać odpowiedni model obsługi, narysować graf stanów systemu, oraz obliczyć:

a) intensywności przejść między stanami,

b) stacjonarne prawdopodobieństwa stanów,

c) średnią długość kolejki do laboratorium.

Rozwiązanie:

1) Model: wielostanowiskowy system ze skończonym źródłem, typu M/M/c/N, gdzie $N = 5$, $c = 2$, $\lambda = 2$, $\mu = \frac{8}{4} = 2$, $\rho = \frac{\lambda}{\mu} = \frac{2}{2} = 1$.

2) Graf stanów systemu:

$$H_0 \underset{\longleftrightarrow}{\mu_1\ \lambda_0} H_1 \underset{\longleftrightarrow}{\mu_2\ \lambda_1} H_2 \underset{\longleftrightarrow}{\mu_3\ \lambda_2} H_3 \underset{\longleftrightarrow}{\mu_4\ \lambda_3} H_4 \underset{\longleftrightarrow}{\mu_5\ \lambda_4} H_5$$

3) Intensywności przejść:

$$\begin{aligned} \lambda_0 &= N\,\lambda &&= 5\,\lambda & \qquad \mu_1 &= 1\,\mu \\ \lambda_1 &= (N-1)\,\lambda &&= 4\,\lambda & \mu_2 &= 2\,\mu \\ \lambda_2 &= (N-2)\,\lambda &&= 3\,\lambda & \mu_3 &= 2\,\mu \\ \lambda_3 &= (N-3)\,\lambda &&= 2\,\lambda & \mu_4 &= 2\,\mu \\ \lambda_4 &= (N-4)\,\lambda &&= 1\,\lambda & \mu_5 &= 2\,\mu \end{aligned}$$

4) Stacjonarne prawdopodobieństwa stanów. Tutaj pierwszym etapem powinno być obliczenie współczynnika Q_i:

$$Q_i = \frac{\lambda_0\,\lambda_1 \cdot \;\cdot\; \cdot \lambda_{i-1}}{\mu_1\,\mu_2 \cdot \;\cdot\; \cdot \mu_i}$$

który

$$Q_0 = 1$$

$$Q_1 = \frac{\lambda_0}{\mu_1} = \frac{5\,\lambda}{\mu} = 5\,\rho = 5$$

$$Q_2 = \frac{\lambda_0\,\lambda_1}{\mu_1\,\mu_2} = \frac{5\lambda \cdot 4\lambda}{\mu \cdot 2\mu} = 10\,\rho^2 = 10$$

$$Q_3 = \frac{\lambda_0\,\lambda_1\,\lambda_2}{\mu_1\,\mu_2\,\mu_3} = \frac{5\lambda \cdot 4\lambda \cdot 3\lambda}{\mu \cdot 2\mu \cdot 3\mu} = 15\,\rho^3 = 15$$

$$Q_4 = \frac{\lambda_0\,\lambda_1\,\lambda_2\,\lambda_3}{\mu_1\,\mu_2\,\mu_3\,\mu_4} = \frac{5\lambda \cdot 4\lambda \cdot 3\lambda \cdot 2\lambda}{\mu \cdot 2\mu \cdot 2\mu \cdot 2\mu} = 15\,\rho^4 = 15$$

$$Q_5 = \frac{\lambda_0\,\lambda_1\,\lambda_2\,\lambda_3\,\lambda_4}{\mu_1\,\mu_2\,\mu_3\,\mu_4\,\mu_5} = \frac{5\lambda \cdot 4\lambda \cdot 3\lambda \cdot 2\lambda \cdot \lambda}{\mu \cdot 2\mu \cdot 2\mu \cdot 2\mu \cdot 2\mu} = 7\frac{1}{2}\,\rho^5 = 7\frac{1}{2}$$

więc

$$p_0 = \frac{1}{\sum\limits_{k=0}^{N} Q_k} = \frac{1}{1+5+10+15+15+7.5} = \frac{2}{107}$$

$$p_1 = p_0 \cdot Q_1 = \frac{2}{107} \cdot 5 = \frac{10}{107}$$

$$p_2 = p_0 \cdot Q_2 = \frac{2}{107} \cdot 10 = \frac{20}{107}$$

$$p_3 = p_0 \cdot Q_3 = \frac{2}{107} \cdot 15 = \frac{30}{107}$$

$$p_4 = p_0 \cdot Q_4 = \frac{2}{107} \cdot 15 = \frac{30}{107}$$

$$p_5 = p_0 \cdot Q_5 = \frac{2}{107} \cdot 7.5 = \frac{15}{107}$$

oraz

$$p_0 + p_1 + p_2 + p_3 + p_4 + p_5 = \frac{2+10+20+30+30+15}{107} = \frac{107}{107} = 1$$

4) Średnia długość kolejki do laboratorium:

$$\overline{v} = \sum_{i=c+1}^{N} (i-c)\,p_i = 1 \cdot p_3 + 2 \cdot p_4 + 3 \cdot p_5 =$$

$$= 1 \cdot \frac{30}{107} + 2 \cdot \frac{30}{107} + 3 \cdot \frac{15}{107} = \frac{30+60+45}{107} = \frac{135}{107} = 1\frac{28}{107}$$

Zadanie 2. W gabinecie dentystycznym pracuje trzech specjalistów, a miejsca w poczekalni są tylko 2. W ciągu godziny przychodzi średnio

4 pacjentów i na każdego z nich potrzeba średnio pół godziny. Wybrać odpowiedni model obsługi, narysować graf stanów systemu oraz obliczyć:

a) intensywności przejść miedzy stanami systemu,

b) stacjonarne prawdopodobieństwa stanów systemu,

c) prawdopodobieństwo blokady systemu (odmowy przyjęcia pacjenta),

d) średnią liczbę zajętych foteli.

Rozwiązanie:

1) Model: system z ograniczoną kolejką, typu M/M/c/L, gdzie: $c = 3$, $L = 2$, $\lambda = 4$, $\mu = 2$.

2) Graf stanów systemu:

$$H_0 \overset{\mu_1\ \lambda_0}{\longleftrightarrow} H_1 \overset{\mu_2\ \lambda_1}{\longleftrightarrow} H_2 \overset{\mu_3\ \lambda_2}{\longleftrightarrow} H_3 \overset{\mu_4\ \lambda_3}{\longleftrightarrow} H_4 \overset{\mu_5\ \lambda_4}{\longleftrightarrow} H_5$$

3) Intensywności:

$$\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda$$

$$\mu_1 = \mu;\ \mu_2 = 2\mu;\ \mu_3 = 3\mu;\ \mu_4 = 3\mu;\ \mu_5 = 3\mu$$

4) Obliczanie współczynników Q_i:

$$Q_i = \frac{\lambda_0\ \lambda_1 \cdot\ \cdot\ \cdot \lambda_{i-1}}{\mu_1\ \mu_2 \cdot\ \cdot\ \cdot \mu_i}$$

więc

$$Q_0 = 1$$

$$Q_1 = \frac{\lambda_0}{\mu_1} = \frac{4}{2} = 2$$

$$Q_2 = \frac{\lambda_0\ \lambda_1}{\mu_1\ \mu_2} = \frac{4 \cdot 4}{2 \cdot 4} = 2$$

$$Q_3 = \frac{\lambda_0\ \lambda_1\ \lambda_2}{\mu_1\ \mu_2\ \mu_3} = \frac{4 \cdot 4 \cdot 4}{2 \cdot 4 \cdot 6} = \frac{8}{6}$$

$$Q_4 = \frac{\lambda_0\ \lambda_1\ \lambda_2\ \lambda_3}{\mu_1\ \mu_2\ \mu_3\ \mu_4} = \frac{4 \cdot 4 \cdot 4 \cdot 4}{2 \cdot 4 \cdot 6 \cdot 6} = \frac{32}{36}$$

$$Q_5 = \frac{\lambda_0\ \lambda_1\ \lambda_2\ \lambda_3\ \lambda_4}{\mu_1\ \mu_2\ \mu_3\ \mu_4\ \mu_5} = \frac{4 \cdot 4 \cdot 4 \cdot 4 \cdot 4}{2 \cdot 4 \cdot 6 \cdot 6 \cdot 6} = \frac{128}{216}$$

5) Prawdopodobieństwa stanów:

$$p_0 = \frac{1}{\sum\limits_{k=0}^{5} Q_k} = \frac{1}{1 + 2 + 2 + \frac{8}{6} + \frac{32}{36} + \frac{128}{216}} = \frac{216}{1688} = \frac{27}{211}$$

$$p_1 = p_0 \cdot Q_1 = \frac{27}{211} \cdot 2 = \frac{54}{211}$$

$$p_2 = p_0 \cdot Q_2 = \frac{27}{211} \cdot 2 = \frac{54}{211}$$

$$p_3 = p_0 \cdot Q_3 = \frac{27}{211} \cdot \frac{8}{6} = \frac{36}{211}$$

$$p_4 = p_0 \cdot Q_4 = \frac{27}{211} \cdot \frac{32}{36} = \frac{24}{211}$$

$$p_5 = p_0 \cdot Q_5 = \frac{27}{211} \cdot \frac{128}{216} = \frac{16}{211}$$

oraz

$$p_0 + p_1 + p_2 + p_3 + p_4 + p_5 = \frac{27 + 54 + 54 + 36 + 24 + 16}{211} = \frac{211}{211} = 1$$

6) Prawdopodobieństwo odmowy przyjęcia pacjenta:

$$p_{bl} = p_5 = \frac{16}{211} = 0.076$$

7) Średnia liczba zajętych foteli:

$$\overline{l} = \sum_{k=1}^{3} k \cdot p_k + \sum_{k=4}^{5} 3 \cdot p_k = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 3 \cdot p_4 + 3 \cdot p_5 =$$

$$= 1 \cdot \frac{54}{211} + 2 \cdot \frac{54}{211} + 3 \cdot \frac{36}{211} + 3 \cdot \frac{24}{211} + 3 \cdot \frac{16}{211} =$$

$$= \frac{54 + 108 + 108 + 72 + 48}{211} = \frac{390}{211} = 1\frac{179}{211} = 1.848$$

Zadanie 3. Wielostanowiskowy system obsługi z ograniczoną długością kolejki typu M/M/c/L.

Dane: $c = 5$, $L = m = 20$.

Dla zadanego obciążenia systemu kolejkowego $\rho = 0.25, 0.5, \cdots, 4.75$ obliczyć i przedstawić wykresy następuj ących parametrów systemu:

a) prawdopodobieństw możliwych stanów systemu (tylko dla $\rho = 1.0$, oraz $\rho = 4.0$),

b) średniej liczby zgłoszeń na stanowiskach obsługi - $\overline{l}$,

c) średniej liczby zgłoszeń w kolejce - $\overline{v}$,

d) średniego czasu pobytu zgłoszenia w systemie (dla λ =1.0) - $\overline{q}$.

Wyniki.

Prawdopodobieństwa wszystkich możliwych stanów systemu:

a) dla $\rho = 1.0$

H_i	p_i	H_i	p_i
0	3.6782 $e-01$	13	7.8467 $e-09$
1	3.6782 $e-01$	14	1.5694 $e-09$
2	1.8391 $e-01$	15	3.1387 $e-10$
3	6.1303 $e-02$	16	6.2774 $e-11$
4	1.5326 $e-02$	17	1.2555 $e-11$
5	3.0651 $e-03$	18	2.5110 $e-12$
6	6.1303 $e-04$	19	5.0219 $e-13$
7	1.2260 $e-04$	20	1.0044 $e-13$
8	2.4521 $e-05$	21	2.0088 $e-14$
9	4.9042 $e-06$	22	4.0175 $e-15$
10	9.8084 $e-07$	23	8.0351 $e-16$
11	1.9617 $e-07$	24	1.6070 $e-16$
12	3.9234 $e-08$	25	3.2140 $e-17$

analizując prawdopodobieństwa stanów systemu dla małej wartości parametru obciążenie systemu, czyli dla $\alpha = 0.2$ ($\rho = 1.0$), należy zauważyć, że prawdopodobieństwo pojawienia się zgłoszenia w kolejce, gdy wszystkie stanowiska obsługi są zajęte, jest znikomo małe i równe $p_6 = 6.1303\ e-04$, a prawdopodobieństwo tego, że w kolejce znajdzie się wieksza liczba zgłoszeń jest jeszcze mniejsze. Prawdopodobieństwo blokady systemu, gdy wszystkie stanowiska obsługi są zajęte, a w kolejce znajduje się dokładnie 20 zgłoszeń, jest bliskie zero, tzn. $p_{25} = 3.2140\ e-17$, a z największym prawdopodobieństwem wyst ąpią stany, gdy w systemie jest zero lub jedno zgłoszenie $p_0 = p_1 = 0.36782$,

b) dla $\rho = 4.00$ ($\alpha = 0.8$)

H_i	p_i	H_i	p_i
0	1.3054 $e-02$	13	1.8688 $e-02$
1	5.2215 $e-02$	14	1.4951 $e-02$
2	1.0443 $e-01$	15	1.1961 $e-02$
3	1.3924 $e-01$	16	9.5685 $e-03$
4	1.3924 $e-01$	17	7.6548 $e-03$
5	1.1139 $e-01$	18	6.1238 $e-03$
6	8.9113 $e-02$	19	4.8991 $e-03$
7	7.1291 $e-02$	20	3.9192 $e-03$
8	5.7033 $e-02$	21	3.1354 $e-03$
9	4.5626 $e-02$	22	2.5083 $e-03$
10	3.6501 $e-02$	23	2.0067 $e-03$
11	2.9201 $e-02$	24	1.6053 $e-03$
12	2.3361 $e-02$	25	1.2843 $e-03$

analiza systemu, przy obciążeniu bliskim optymalnemu ($\alpha = 0.8$) pokazuje, że z najwyższym prawdopodobieństwem, wystąpi taki stan systemu, w którym będą trzy lub cztery zgłoszenia na stanowiskach obsługi, tzn. $p_3 = p_4 = 0.13924$, zaś blokada systemu, gdy wszystkie stanowiska obsługi są zajęte i kolejka jest pełna, wystąpi z prawdopodobieństwem $p_{25} = 1.2843\ e-03$.

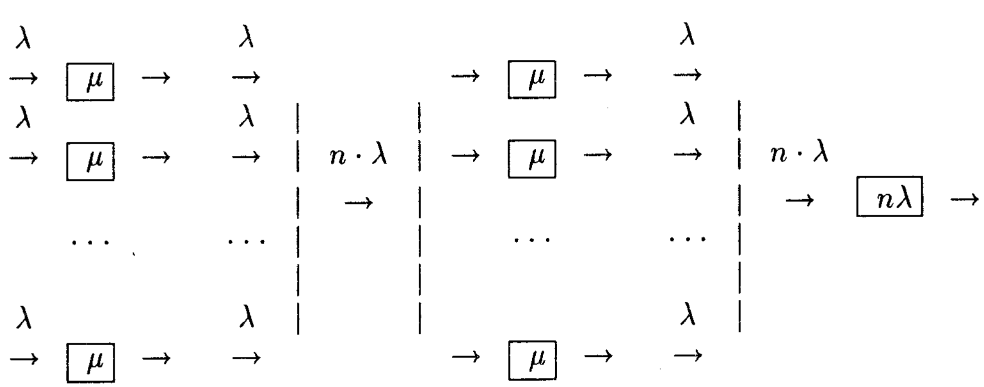
Średnie liczby zgłoszeń na stanowiskach obsługi i w kolejce oraz średni czas pobytu zgłoszenia w systemie:

ρ	$\overline{l}$	$\overline{v}$	$\overline{q}$	ρ	$\overline{l}$	$\overline{v}$	$\overline{q}$
0.00	0.000	0.000	0.000	2.50	2.500	0.130	2.630
0.25	0.250	0.000	0.250	2.75	2.750	0.218	2.968
0.50	0.500	0.000	0.500	3.00	3.000	0.354	3.354
0.75	0.750	0.000	0.750	3.25	3.250	0.561	3.811
1.00	1.000	0.001	1.001	3.50	3.500	0.877	4.377
1.25	1.250	0.003	1.253	3.75	3.749	1.361	5.110
1.50	1.500	0.009	1.509	4.00	3.995	2.099	6.094
1.75	1.750	0.020	1.770	4.25	4.233	3.203	7.436
2.00	2.000	0.040	2.040	4.50	4.454	4.756	9.210
2.25	2.250	0.074	2.324	4.75	4.643	6.733	11.376

analiza podstawowych parametrów systemu pozwala stwierdzić, że średnia liczba zgłoszeń w kolejce, czy średni czas reakcji systemu (pobytu zgłoszenia w systemie), zaczyna szybko wzrastać, gdy obciążenie systemu ρ przekracza 3.50, czyli dla $\alpha > 0.75$, a dla $\alpha < 0.45$ kolejka praktycznie nie istnieje (patrz wykresy na następnych stronach).

Zadanie 4. Koncentracja zasobów w systemie komputerowym:

a) b) c)



gdzie instalacja składa się z n identycznych jednostek, które można traktować jako:

a) n identycznych systemów typu M/M/1,

b) jako system typu M/M/n, gdzie n niezależnych strumieni łączy się w jeden z intensywnością $n\lambda$ i n równoległymi stanowiskami obsługi,

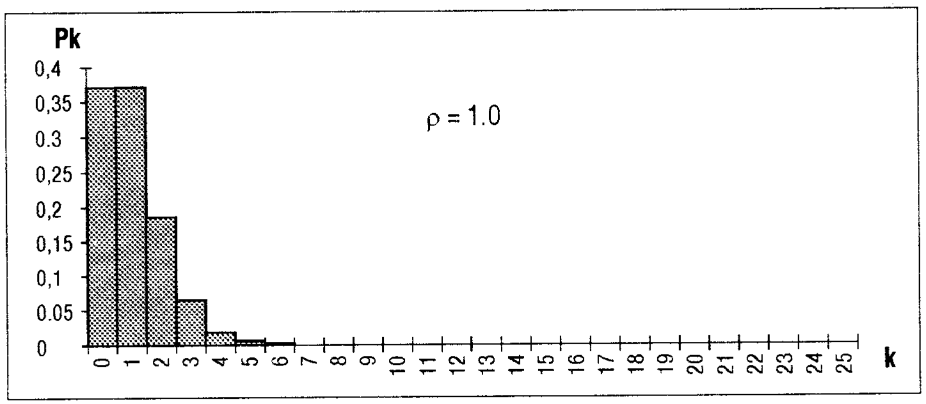
c) tak jak w b) lecz n równoległych stanowisk łączy się w jedno, obsługujące zgłoszenia n razy szybciej.

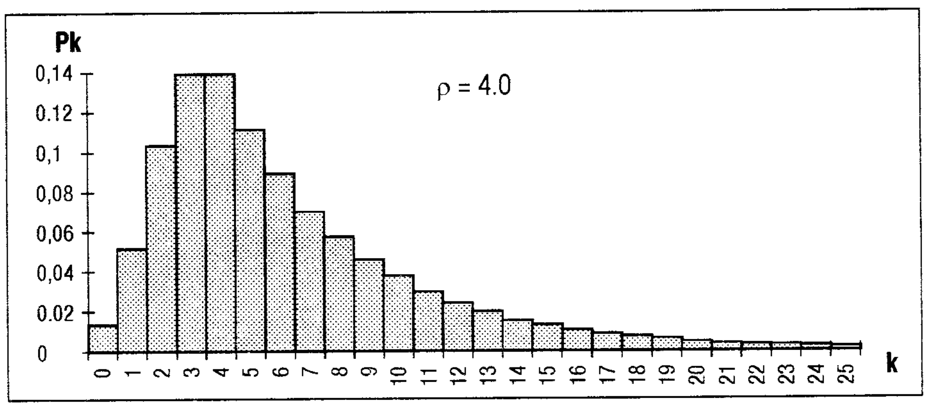
Dane: $\lambda = 0.95$, $\mu = 1.0,\ 1.2,\ 1.5,\ 2.0,\ 3.5,\ 5.0$, $n = 4$.

Obliczyć i przedstawić w graficznej formie średni czas pobytu zgłoszenia w systemie - $\overline{q}$ i średnią liczbę zgłoszeń w systemie - $\overline{n}$, dla każdej z tych strategii rozdziału zasobów.

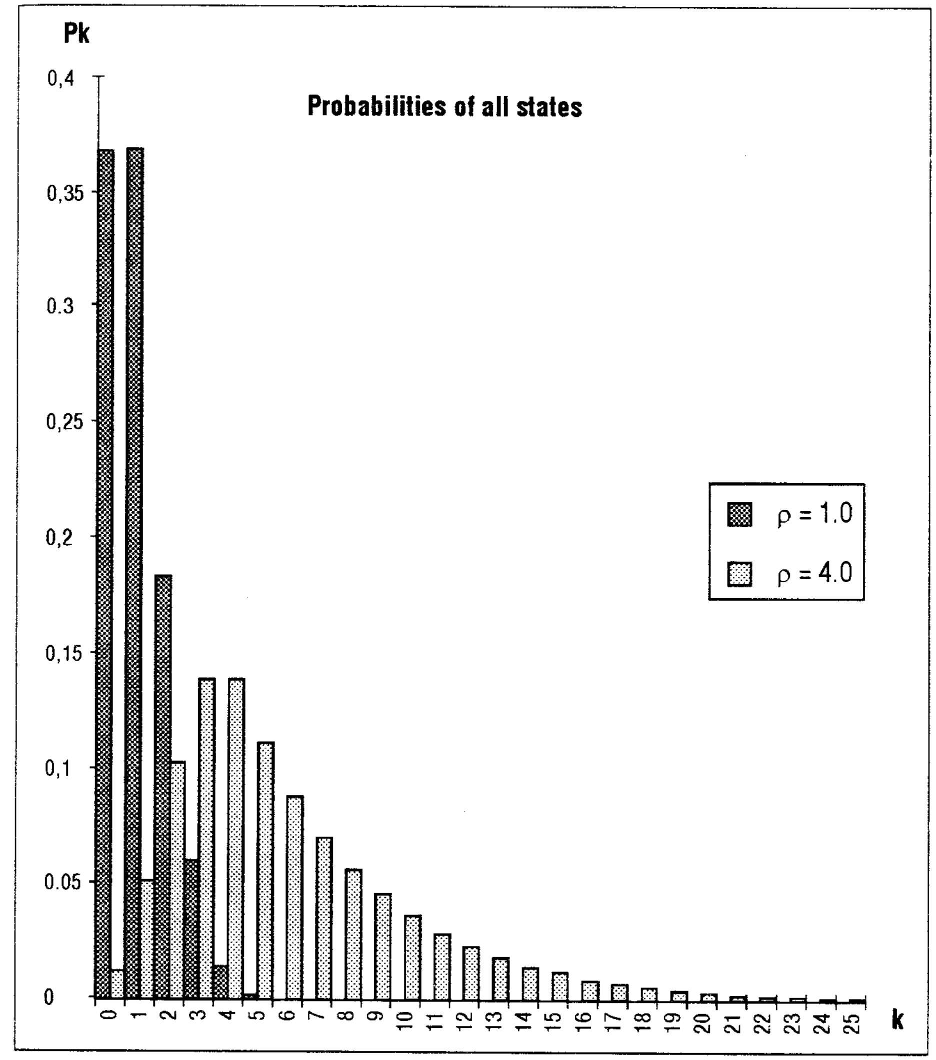
Rezultaty i wykresy: dwie strony dalej.

Wykresy prawdopodobieństw stanów systemu dla różnych wartości współczynnika obciążenia systemu, czyli dla $\rho = 1.0$ i $\rho = 4.0$:

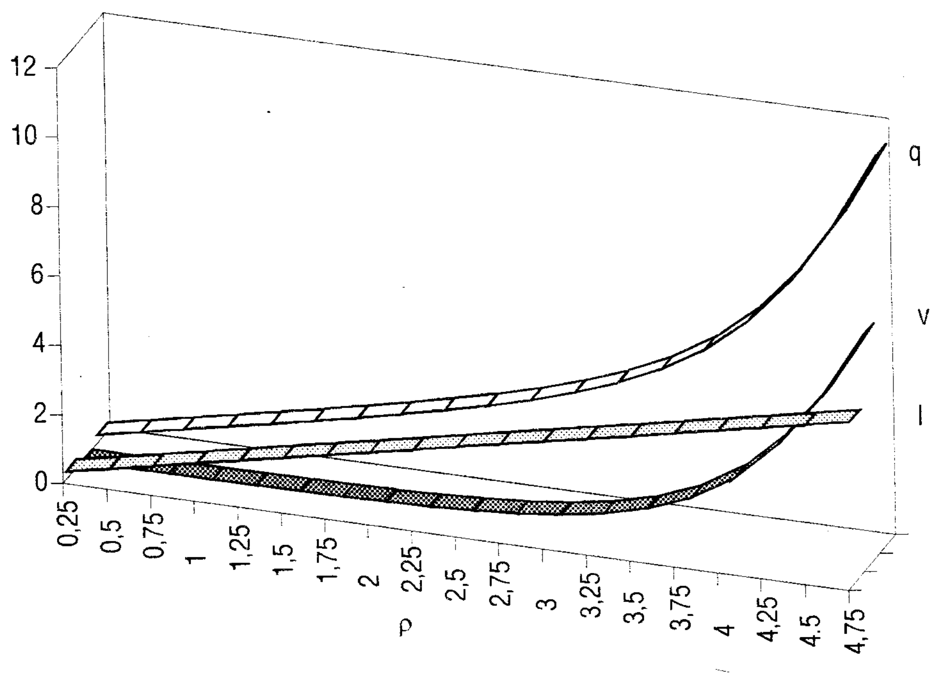




Porównawczy wykres prawdopodobieństw stanów dla $\rho = 1.0$ i $\rho = 4.0$:

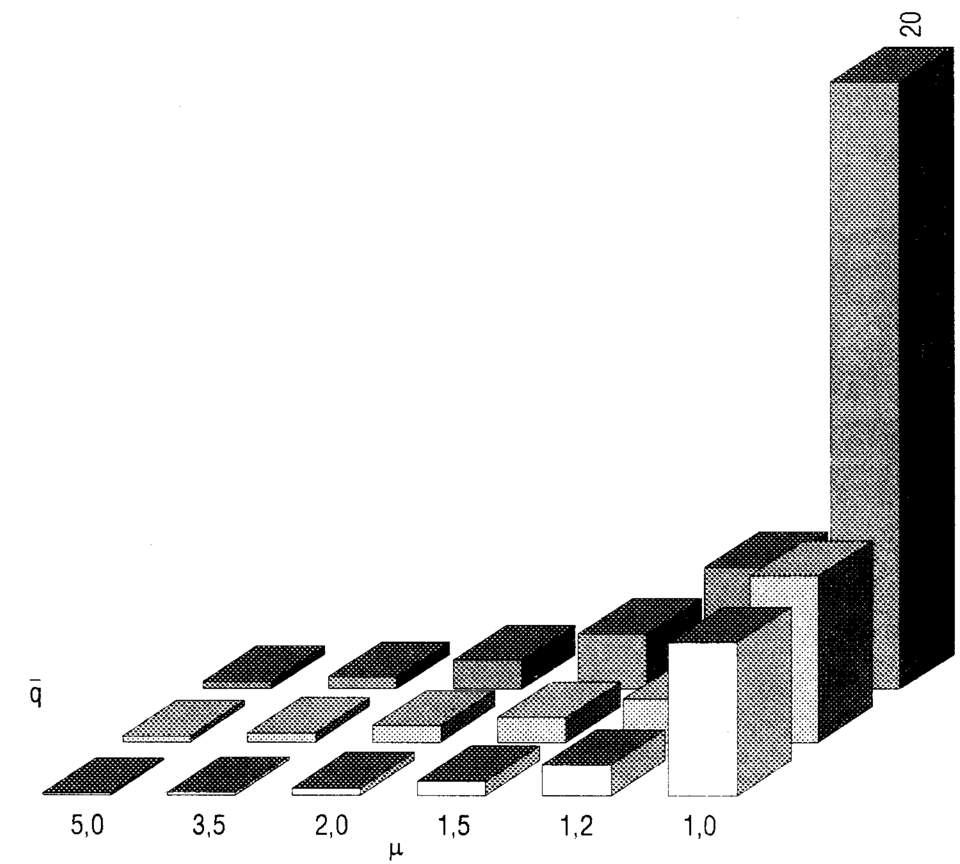


Wykresy średniej liczby zgłoszeń na stanowiskach obsługi - $\overline{l}$, średniej długości kolejki - $\overline{v}$, średniego czasu pobytu zgłoszenia w systemie - $\overline{q}$, w zależności od wartości parametru obciążenia systemu - ρ:



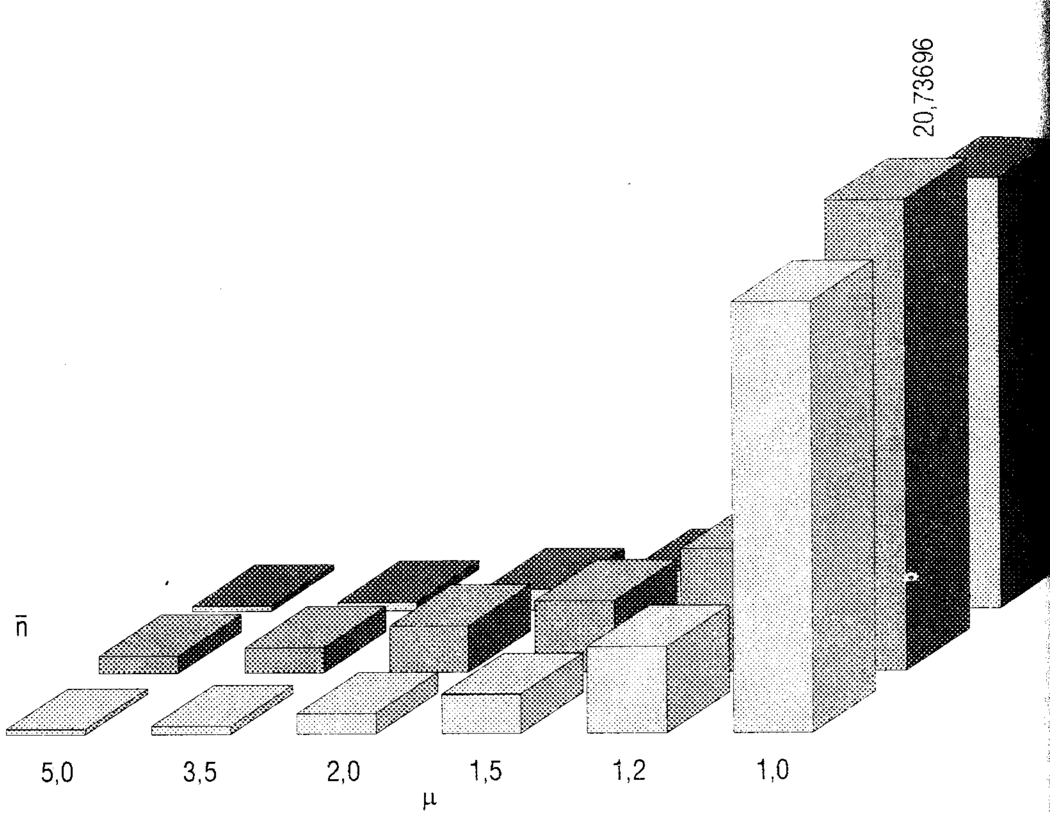
Średnia czas pobytu zgłoszenia w systemie $\overline{q}$:

μ	ρ	$a)$	$b)$	$c)$
1.0	0.9500	20.0000	5.4571	5.0000
1.2	0.7917	4.0000	1.4148	1.0000
1.5	0.6333	1.8182	0.8172	0.4545
2.0	0.4750	0.9524	0.5358	0.2381
3.5	0.2714	0.3922	0.2883	0.0980
5.0	0.1900	0.2469	0.2005	0.0617



tak jak można było oczekiwać podstawowy parametr czasowy: $\overline{q_a} > \overline{q_b} > \overline{q_c}$ dla dowolnej wartości parametru ρ i jest to podstawowy zysk otrzymany dzięki koncentracji zasobów systemu komputerowego. Przy intuicyjnej interpretacji otrzymamy, że w przypadku a) pewne stanowiska obsługi mogą być wolne, gdy w tym samym czasie inne są zajęte obsługą, jak również mogą mieć dodatkowo i zgłoszenia w kolejkach, w przypadku b) taka sytuacja jest nie możliwa (wspólna kolejka), ale przy braku zgłoszeń w kolejce, jedne stanowiska mogą być zajęte obsługą, a inne w tym samym czasie są wolne i nie wykorzystują swojej mocy obliczeniowej. Przypadek c) wolny jest od tych niedostatków i w najpełniejszy sposób wykorzystuje swoje zasoby.

Średnia liczba zgłoszeń w systemie $\overline{n}$:



μ	ρ	$a)$	$b)$	$c)$
1.0	0.9500	19.0000	20.7370	19.0000
1.2	0.7917	3.8000	5.3764	3.8000
1.5	0.6333	1.7273	3.1054	1.7273
2.0	0.4750	0.9048	2.0360	0.9048
3.5	0.2714	0.3725	1.0957	0.3725
5.0	0.1900	0.2346	0.7619	0.2346

Zadanie 5. Systemy obsługi ze stratą:

1) System typu M/M/1 z blokadą. Dla zadanego parametru obciążenia stanowiska obsługi: $\rho = 0.0,\ 0.1,\ 0.2,\ \cdots,\ 1.0$ obliczyć i przedstawić graficznie prawdopodobieństwa obsługi oraz straty zgłoszenia.

2) System typu M/M/c z blokadą. Dla $c = 3$, oraz $\rho = 0.0,\ 0.25,\ 0.50,\ \cdots,\ 3.0$ obliczyć i przedstawić graficznie prawdopodobieństwa blokady i obsługi zgłoszenia oraz średnią liczbę zgłoszeń na stanowiskach obsługi.

Zadanie 6. System obsługi typu M/M/1/L z kolejką.

Dla $L = m = 5$ i dla $\rho = 0.0,\ 0.1,\ \cdots,\ 0.9$ obliczyć i przedstawić w formie wykresów:

a) prawdopodobieństwa stanów systemu (wykres dla $\rho = 0.7$),
b) prawdopodobieństwo blokady systemu,
c) prawdopodobieństwo obsługi zgłoszenia,
d) średnią liczbę zgłoszeń na stanowisku obsługi,
e) średnią liczbę zgłoszeń w kolejce,
f) średni czas oczekiwania w kolejce,
g) średni czas pobytu zgłoszenia w systemie (dla $\mu = 2$).

Zadanie 7. Jednostanowiskowy system kolejkowy typu M/M/1/N ze skończenie wymiarowym źródłem.

Dla $\mu = 2,\ N = 10$, oraz $\rho = 0.0,\ 0.1,\ \cdots,\ 1.0$ obliczyć i przedstawić graficznie:

a) prawdopodobieństwa stanów systemu (wykres dla $\rho = 0.8$),

b) prawdopodobieństwo zajętości konserwatora,
c) średnią liczbę obrabiarek wymagających naprawy (zepsutych),
d) średnią liczbę obrabiarek naprawianych,
e) średną liczbę obrabiarek oczekujących naprawy,
f) średni czas oczekiwania w kolejce do naprawy.

Zadanie 8. Wielostanowiskowy system obsługi ze skończenie wymiarowym źródłem zgłoszeń typu M/M/c/N.

Dla $N = 15$, $c = 3$, oraz $\rho = 0.0,\ 0.3,\ 0.6,\ \cdots,\ 3.0$ obliczyć i przedstawić graficznie:
a) prawdopodobieństwa stanów systemu (grafika dla $\rho = 2.7$),
b) prawdopodobieństwo zajętości wszystkich stanowisk,
c) średnią liczbę obrabiarek oczekujących naprawy (w kolejce),
d) średnią liczbę naprawianych obrabiarek,
e) średnią liczbę obrabiarek wymagających naprawy (w kolejce i naprawianych).

Rozdział 3

Niemarkowowskie modele systemów kolejkowych

3.1 System obsługi typu M/G/1 z nieograniczoną kolejką

System typu M/G/1 nie jest systemem typu markowowskiego (nie jest procesem Markowa), a poprzednie metody badania systemów kolejkowych korzystały z tej własności jednorodnego łańcucha Markowa, która polegała na tym, że wybierając dowolny ciąg momentów czasowych $\{t_k\}$, z przestrzeni parametru czasowego T, to odpowiadający mu ciąg zmiennych losowych $\{X_{t_k}, t_k \in T\}$ stanowił jednorodny łańcuch Markowa, włożony do procesu $\{X_t, t \in T\}$ i są to, tzw. punkty regeneracyjne procesu.

Inaczej rzecz ujmując, każdy punkt przestrzeni parametru czasowego T jest punktem regeneracyjnym procesu, więc gdyby, dla systemu M/G/1 udało się znaleźć ciąg punktów regeneracyjnych, to badanie takiego procesu (systemu) można zastąpić badaniem łańcucha włożonego do procesu. Należy tylko znaleźć dla niemarkowowskich systemów takiego rodzaju punkty.

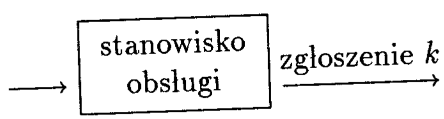
Okazało się, że losowe momenty opuszczania systemu przez obsłużone zgłoszenia są właśnie takimi punktami regeneracyjnymi procesu. Tworzą one przeliczalny ciąg $\{t_k\}$ dla $k \in N_1$, gdzie N_1 - liczby naturalne $1, 2, \cdot$

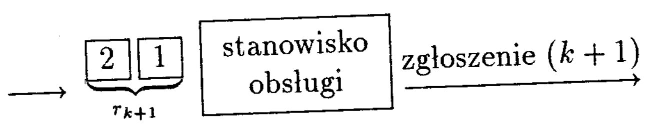
Jeżeli badanym procesem jest proces $\{N(t),\ t \geq 0\}$, którego przestrzenią fazową jest $S = \{0, 1, \cdots\}$, to można oznaczyć przez $N_k = N(t_k + 0)$ liczbę zgłoszeń w systemie, natychmiast po opuszczeniu systemu przez zgłoszenie z numerem k. Oczywiste jest, iż N_{k+1} zależy tylko od N_k, a nie zależy od stanu systemu w momentach poprzedzających t_k. Zatem ciąg zmiennych losowych $\{N_k\}$ stanowi łańcuch Markowa o przestrzeni stanów fazowych S i o wartościach parametru czasowego t_k, włożony w proces $\{N(t),\ t \geq 0\}$.

W systemie masowej obsługi typu M/G/1 czas między kolejnymi zgłoszeniami ma rozkład wykładniczy o parametrze λ, a czas obsługi, niezależny od odstępów czasu między kolejnymi zgłoszeniami (od procesu napływu zgłoszeń), jest zmienną losową o dystrybuancie $B(t)$. Zakładamy, że zmienna ta ma skończoną wartość średnią, równą $\frac{1}{\mu}$ i wariancję σ^2.

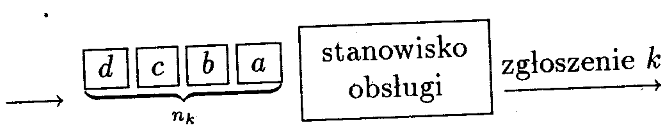
Oznaczmy przez: n_k - liczbę zgłoszeń w systemie w chwili zakończenia odsługi (i opuszczenia systemu) przez k-te - zgłoszenie, a przez r_{k+1} - liczbę zgłoszeń, które przybyły do systemu w trakcie obsługi zgłoszenia z numerem $(k+1)$. Jest rzeczą oczywistą, że liczby $r_1, r_2, \cdots$ są od siebie niezależne, a zależność między wielkościami n_k i n_{k+1}, bezpośrednio po opuszczeniu systemu, przez $(k+1)$ zgłoszenie, można zilustrować w następujący sposób:

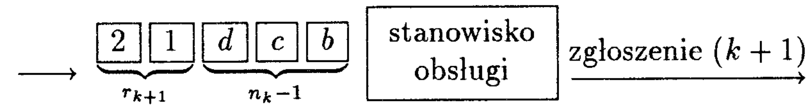
a) dla $n_k = 0$:





b) dla $n_k > 0$





czyli

$$n_{k+1} = n_k + r_{k+1} - 1 + \delta \tag{3.1}$$

gdzie, δ funkcja logiczna:

$$\delta = \begin{cases} 0 & \text{jeżeli} \quad n_k \neq 0 \\ 1 & \text{jeżeli} \quad n_k = 0 \end{cases}$$

Pogrupujmy wyrazy i podnieśmy do kwadratu obie strony powyższego równania:

$$n_{k+1}^2 = (n_k + \delta)^2 + 2(n_k + \delta)(r_{k+1} - 1) + (r_{k+1} - 1)^2 =$$

$$= n_k^2 + (r_{k+1} - 1)^2 + \delta^2 + 2\, n_k(r_{k+1} - 1) + 2\, n_k\ \delta + 2(r_{k+1} - 1)\delta \tag{3.2}$$

Równanie to rozpatrzmy w tzw. stanie statystycznej równowagi granicznej i obliczmy wartość średnią obu stron, gdy:

$$E\,[n_k] = E\,[n_{k+1}] = \overline{n} \tag{3.3}$$

z powyższego równania oraz z definicji funkcji logicznej mamy

$$n_k\delta = 0$$

oraz

$$E\,[n(r-1)] = \overline{n}(\overline{r} - 1)$$

$$E\,[\delta(r-1)] = \overline{\delta}(\overline{r} - 1)$$

to otrzymamy:

$$\overline{n^2} = \overline{n^2} + \overline{r^2} - 2\overline{r} + 1 + \overline{\delta^2} + 2\,\overline{n}(\overline{r} - 1) + 2\,\overline{n}\,\overline{\delta} + 2\,\overline{\delta}\,(\overline{r} - 1) \tag{3.4}$$

Teraz, kolejno, należy określić wielkości występujące w tym równaniu, gdzie $\overline{r}$ jest wartością średnią liczby zgłoszeń, napływających do systemu, w średnim czasie na obsługę:

$$\overline{r} = \lambda \cdot \frac{1}{\mu} = \rho \tag{3.5}$$

Pamiętając o tym, że wejściowym strumieniem jest strumień Poissona, więc prawdopodobieństwo, że w określonym przedziale czasowym (w naszym przypadku jest to czas obsługi zgłoszenia τ_2) nadejdzie dokładnie r zgłoszeń, określone jest następującym wzorem:

$$\frac{(\lambda\ \tau_2)^r}{r!} e^{-\lambda\ \tau_2}$$

Uwzględniając to wyrażenie można określić wartość średnią kwadratu liczby zgłoszeń przychodzących do systemu w określonym czasie obsługi τ_2, jako:

$$\sum_{r=0}^{\infty} r^2 \cdot p_r(\tau_2) = \sum_{r=0}^{\infty} r^2 \cdot \frac{e^{-\lambda\ \tau_2} \cdot (\lambda\ \tau_2)^r}{r^2} = \cdot\ \cdot\ \cdot = \lambda\ \tau_2(1 + \lambda\ \tau_2)$$

oraz, pamiętając o tym, że τ_2 jest zmienną losową o rozkładzie $B(t)$ otrzymamy:

$$E(r^2) = \overline{r^2} = \int_0^{\infty} \lambda\ \tau_2(1 + \lambda\ \tau_2)\ dB(\tau_2) =$$

$$= \lambda \int_0^{\infty} \tau_2\ dB(\tau_2) + \lambda^2 \int_0^{\infty} (\tau_2)^2 dB(\tau_2)$$

gdzie pierwsza część, to wzór na moment pierwszego rzędu (m_1), druga zaś, to moment drugiego rzędu (m_2), czyli

$$\overline{r^2} = \lambda\ \overline{\tau_2} + \lambda^2\ \overline{\tau_2^2}$$

z kolei

$$m_2 - m_1^2 = \sigma^2$$

to

$$\overline{r^2} = \rho + \lambda^2 \left(\sigma^2 + (\overline{\tau_2})^2\right) = \rho + \rho^2 + \lambda^2\sigma^2 \tag{3.6}$$

Następnie, jeśli korzystamy z określenia funkcji logicznej δ, to można zapisać:

$$\overline{\delta^2} = E(\delta^2) = E(\delta) = \overline{\delta}$$

i dalej, jeśli rozważyć wartości średnie obu stron równania dla n_{k+1} (patrz wzór (3.1)):

$$\overline{n_{k+1}} = \overline{n_k} + \overline{r_{k+1}} - 1 - \overline{\delta}$$

to dla stanu ustalonego otrzymamy:

$$\overline{\delta} = 1 - \overline{r} = 1 - \rho \tag{3.7}$$

Ostatecznie wyrażenie dla obliczenia wartości średniej $\overline{n}$ wygląda następująco (ze wzoru (3.4)):

$$0 = \overline{r^2} - 2\overline{r} + 1 + \overline{\delta^2} + 2\ \overline{n}\ \overline{r} - 2\ \overline{n} + 2\ \overline{\delta}\,(\overline{r} - 1)$$

$$2\ \overline{n} - 2\ \overline{n}\ \overline{r} = \overline{r^2} - 2\ \overline{r} + 1 + \overline{\delta} + 2\ \overline{\delta}\,\overline{r} - 2\ \overline{\delta}$$

$$2\ \overline{n}(1 - \overline{r}) = \overline{r^2} - 2\ \overline{r} + 1 + 2\ \overline{\delta}\,\overline{r} - \overline{\delta} = \overline{r^2} + 2\ \overline{r}(\overline{\delta} - 1) - (\overline{\delta} - 1) =$$

$$= \overline{r^2} + (\overline{\delta} - 1)(2\ \overline{r} - 1) = \rho + \rho^2 + \lambda^2\sigma^2 + (1 - \rho - 1)(2\ \rho - 1) =$$

$$= \rho^2 + \lambda^2\sigma^2 - 2\ \rho^2 + 2\ \rho$$

stąd

$$\overline{n} = \frac{\rho^2 + \lambda^2\sigma^2 + 2\ \rho(1 - \rho)}{2(1 - \rho)} = \frac{2\ \rho(1 - \rho)}{2(1 - \rho)} + \frac{\rho^2 + \lambda^2\sigma^2}{2(1 - \rho)}$$

i ostatecznie:

$$\overline{n} = \rho + \frac{\rho^2 + \lambda^2\sigma^2}{2(1 - \rho)} \tag{3.8}$$

Wyrażenie to w literaturze nazywane jest wzorem Chinczyna-Pollaczka.

Sprawdzamy to wyrażenie na średnią liczbę zgłoszeń w systemie dla przypadku wykładniczego rozkładu czasu obsługi (system M/M/1), w którym:

$$\sigma^2 = \frac{1}{\mu^2}$$

więc

$$\overline{n} = \rho + \frac{\rho^2 + \rho^2}{2(1-\rho)} = \rho + \frac{\rho^2}{(1-\rho)} = \frac{\rho - \rho^2 + \rho^2}{1-\rho} = \frac{\rho}{1-\rho}$$

Inne parametry (charakterystyki) systemu można obliczyć w oparciu miedzy innymi, o wzory Little'a:

a) średnia liczba zgłoszeń na stanowisku obsługi (tak jak i dla systemu M/M/1):

$$\overline{l} = \rho \tag{3.9}$$

b) średnia liczba zgłoszeń w kolejce:

$$\overline{v} = \overline{n} - \overline{l} = \rho + \frac{\rho^2 + \lambda^2\sigma^2}{2(1-\rho)} - \rho = \frac{\rho^2}{2(1-\rho)}\left(1 + \frac{\sigma^2\lambda^2}{\frac{\lambda^2}{\mu^2}}\right) =$$

$$= \frac{\rho^2}{2(1-\rho)}\left(1 + \sigma^2\mu^2\right) = \frac{\rho^2}{2(1-\rho)}\left[1 + \left(\frac{\sigma}{\overline{\tau_2}}\right)^2\right] \tag{3.10}$$

gdzie, $\overline{\tau_2}$ - średni czas obsługi,

c) średni czas oczekiwania zgłoszeń w kolejce:

$$\overline{w} = \frac{\overline{v}}{\lambda} = \frac{\overline{n} - \overline{l}}{\lambda} = \frac{\rho}{\lambda} + \frac{\frac{\rho^2}{\lambda} + \lambda\,\sigma^2}{2(1-\rho)} - \frac{\rho}{\lambda} = \frac{\frac{\rho^2}{\lambda} + \lambda\,\sigma^2}{2(1-\rho)} \tag{3.11}$$

d) średni czas pobytu zgłoszenia w systemie:

$$\overline{q} = \overline{w} + \frac{1}{\mu}$$

Przeanalizujmy oddzielnie następującą składową ze wzoru (3.10):

$$\frac{1}{2}\cdot\left[1 + \left(\frac{\sigma}{\overline{\tau_2}}\right)^2\right] \tag{3.12}$$

Wyrażenie to występuje również i w innych parametrach opisujących podstawowe charakterystyki systemów kolejkowych. Można więc tutaj wydzielić dwa podstawowe przypadki:

1) rozkład deterministyczny (stałe czasy obsługi): system typu M/D/1, w którym:

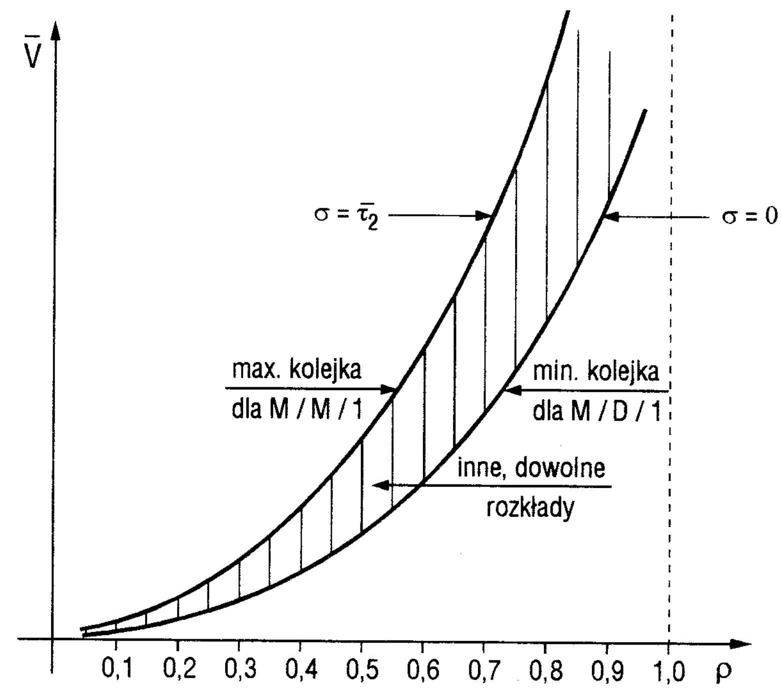
$$\sigma = 0$$

czyli wyrażenie (3.12) ma wartość równą $\frac{1}{2}$, a np. wyrażenie dla średniej liczby zgłoszeń w kolejce jest równe:

$$\overline{v} = \frac{\rho^2}{2(1-\rho)}$$

lub

$$\overline{n} = \rho + \frac{\rho^2}{2(1-\rho)}$$



Rys. 6. Długość kolejki w systemie z różnymi rozkładmi czasu obsługi (typ M/G/1)

2) rozkład wykładniczy, w którym:

$$\sigma = \overline{\tau_2} = \frac{1}{\mu}$$

i wyrażenie (3.12) równe jest 1, bo

$$\frac{1}{2}\cdot\left[1+\left(\frac{\overline{\tau_2}}{\overline{\tau_2}}\right)^2\right]=1$$

więc

$$\overline{v}=\frac{\rho^2}{1-\rho}$$

oraz

$$\overline{n}=\frac{\rho}{1-\rho}$$

Te dwa graniczne przypadki wyznaczają przedziały, w którym zawarte są wartości podstawowych parametrów systemów z dowolnym, innym rozkładem, ale identycznymi intensywnościami λ i μ. Jako przykład niech służy wykres rozmiarów kolejki, przy różnych rozkładach czasu obsługi, w zależności od wartości parametru obcążenia ρ (rys. 6).

3.2 System obsługi typu G/M/1 z nieograniczoną kolejką

System kolejkowy takiego typu charakteryzuje się tym, że:

a) odstępy czasu między kolejnymi zgłoszeniami mają rozkład dowolny, z jednakową dystrybuantą $A(t)$ i wartością średnią równą $\frac{1}{\lambda}$ i funkcją gęstości $a(t)$. Przekształcenie Laplace'a funkcji $a(t)$ oznaczmy jako $\widetilde{A}(t)$. Przypomnijmy, że przekształceniem Laplace'a nazywamy przyporządkowanie danej funkcji $a(t)$ zmiennej rzeczywistej t funkcji $\widetilde{A}(s)$ zmiennej zespolonej s wg następującej relacji:

$$\widetilde{A}(s)=\int_0^{\infty} a(t)\, e^{-st}dt \qquad (3.13)$$

przy założeniu istnienia powyższej całki w pewnej półpłaszczyźnie Re $s>\alpha_0$, gdzie $\alpha_0\in\Re$ (zbioru liczb rzeczywistych),

b) czas obsługi zgłoszeń ma rozkład wykładniczy z wartością średnią równą $\frac{1}{\mu}$.

Podstawowe rekurencyjne wyrażenie, podobnie jak dla systemu M/G/1, ma następującą postać:

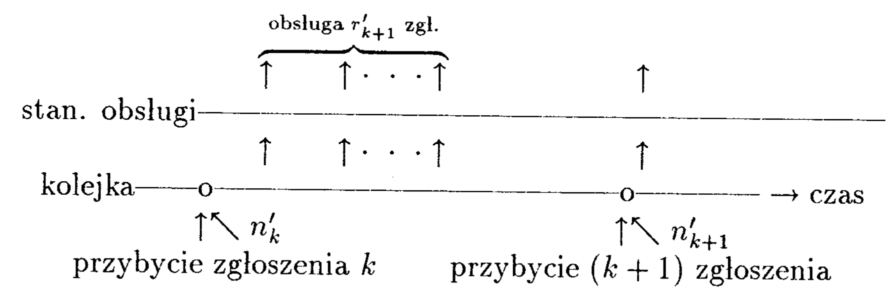
$$n'_{k+1}=n'_k+1-r'_{k+1} \qquad (3.14)$$

gdzie:

n'_k -liczba zgłoszeń w systemie, w chwili przybycia do systemu, zgłoszenia z numerem k,

r'_{k+1} - liczba zgłoszeń obsłużonych w czasie, między przybyciem do systemu zgłoszenia z numerem k i z numerem $k+1$.

Zatem ciag $\{n'_k\}$ stanowi, jak i w przypadku systemu M/G/1, łańcuch Markowa. Punkty regeneracyjne (momenty pojawienia się nowych zgłoszeń) włożonego łańcucha Markowa, dla systemu G/M/1 mogą być zilustrowane w następujący sposób:



Stosując metodę włożonych łańcuchów Markowa należy pamiętać o tym, że to wybrane punkty regeneracyjne zawierają informacje o czasie, jaki upłynął od przybycia ostatniego ze zgłoszeń (podobnie jak dla systemu M/G/1).

Prawdopodobieństwa przejść między stanami fazowymi systemu, które w tym przypadku związane są z globalną liczbą zgłoszeń w systemie, dla włożonego łańcucha Markowa, można obliczyć tak:

$$p_{ij}=P\left[n'_{k+1}=j\mid n'_k=i\right] \qquad (3.15)$$

Jasne jest, że p_{ij} równe jest prawdopodobieństwu tego, że w przedziale między przybyciem zgłoszeń obsłużono

$$(i+1-j)\text{ zgłoszeń}$$

i dlatego, że interesują nas tylko rezultaty otrzymane dla stanów ustalonych, to szukamy stacjonarnych rozkładów prawdopodobieństw, opisujących liczbę zgłoszeń znajdujących się w systemie, w momentach przybywania zgłoszeń. Więc należy też określić:

$$r_l = \lim_{k \to \infty} P\left[n'_k = l\right]$$

Dalsza, drobiazgowa analiza systemu G/M/1 jest dość zawiła, więc ograniczmy się tylko do podania podstawowych rezultatów. Wprowadźmy pewną wielkość, którą oznaczmy przez γ_l. Niech γ_l będzie liczbą przejść (dla włożonego łańcucha Markowa) do stanu fazowego $l+1$ między dwoma kolejnymi przejściami do stanu l. Okazuje się, że γ_l nie zależy od l i może być określona w sposób następujący:

$$\gamma = \lim_{t \to \infty} \frac{N_{l+1}(t)}{N_l(t)} \tag{3.16}$$

gdzie, $N_l(t)$- jest liczbą takich zdarzeń, w czasie od 0 do t, kiedy przybywające do systemu zgłoszenie zastaje ten system w stanie fazowym l, przy założeniu, że w momencie $t = 0$, w systemie było 0 zgłoszeń. Jeszcze inaczej rzecz ujmując, wyrażenie (3.16) jest granicą ilorazu stacjonarnego prawdopodobieństwa przebywania systemu w stanie fazowym $(l+1)$ i stacjonarnego prawdopodobieństwa przebywania systemu w stanie fazowym l.

Podajmy, bez dowodu, że podstawowe charakterystyki systemu wyrażone są przez pierwiastek γ, który (w obszarze $0 < \gamma < 1$) jest jedynym pierwiastkiem równania:

$$\gamma = \widetilde{A}(\mu - \mu\ \gamma) \tag{3.17}$$

bo

$$\gamma = \int_0^\infty e^{-(\mu - \mu\ \gamma)\ t} dA(t)$$

Jest to transformata (przekształcenie) Laplace'a gęstości rozkładu prawdopodobieństwa dla odstępów między przybywaniem kolejnych zgłoszeń, obliczona w określonym punkcie. Z powyższego wzoru wynika też, że w systemie G/M/1, w momentach przybywania nowych zgłoszeń, liczba zgłoszeń w systemie, ma rozkład geometryczny.

Znając γ możemy obliczyć, np. $\overline{w}$, czyli średni czas oczekiwania w kolejce:

$$\overline{w} = \frac{\gamma}{\mu(1-\gamma)} \tag{3.18}$$

Analiza pokazuje, że czas oczekiwania zgłoszeń w kolejce ma rozkład wykładniczy i nie zależy od rozkładu odstępów między przybyciem kolejnych zgłoszeń (tak jak dla systemu M/M/1).

Przykład 1: Zilustrujmy to dla systemu M/M/1, w którym:

$$A(t) = 1 - e^{-\lambda\ t} \qquad \text{dla } t > 0$$

$$\widetilde{A}(s) = \frac{\lambda}{s + \lambda}$$

bezpośrednio z wyrażenia (3.17) otrzymamy, że γ powinna spełniać równanie w postaci (za s wstawiamy $\mu - \mu\ \gamma$):

$$\gamma = \frac{\lambda}{\mu - \mu\ \gamma + \lambda}$$

lub

$$\mu\ \gamma^2 - (\mu + \lambda)\gamma + \lambda$$

czyli

$$(\gamma - 1)(\mu\ \gamma - \lambda) = 0$$

z tych dwóch rozwiązań, rozwiązanie $\gamma = 1$ odrzucamy, bo mamy warunek stabilności $(0 < \gamma < 1)$, więc jedynym rozwiązaniem jest:

$$\gamma = \frac{\lambda}{\mu} = \rho$$

czyli, ze wzoru (3.18) mamy:

$$\overline{w} = \frac{\rho}{\mu(1-\rho)}$$

dalej możemy obliczyć $\overline{v}$, potem:

$$\overline{q} = \overline{w} + \frac{1}{\mu}$$

następnie $\overline{n}$, itd.

Przykład 2: Dany jest system typu G/M/1, w którym gęstość rozkładu prawdopodobieństwa dla odstępów między zgłoszeniami ma transformatę Laplace'a równą:

$$\widetilde{A}(s) = \frac{2\,\mu^2}{(s+\mu)(s+2\,\mu)}$$

co odpowiada stystemowi $E_2/M/1$, w którym dwie fazy w erlangowskim modelu przybywania zgłoszeń mają różne intensywności, wybrane jako wielokrotność parametru intensywności obsługi μ.

Z wyrażenia (3.17), wstawiając za s : $\mu - \mu\,\gamma$ otrzymamy:

$$\gamma = \frac{2\,\mu^2}{(\mu - \mu\,\gamma + \mu)(\mu - \mu\,\gamma + 2\,\mu)}$$

a po odpowiednich przekształceniach mamy:

$$\gamma^3 - 5\,\gamma^2 + 6\,\gamma - 2 = 0$$

Wiadomo, że $\gamma = 1$ zawsze będzie rozwiązaniem tego równania i to pozwala rozłożyć równanie trzeciego stopnia na:

$$(\gamma - 1)(\gamma - 2 - \sqrt{2})(\gamma - 2 + \sqrt{2}) = 0$$

z powyższych pierwiastków, tylko jeden pierwiastek $2 - \sqrt{2}$ spełnia warunek $0 < \gamma < 1$, więc np. czas oczekiwania w kolejce równy jest:

$$\overline{w} = \frac{\gamma}{\mu(1-\gamma)} = \frac{2-\sqrt{2}}{\mu(1-2+\sqrt{2})} = \frac{2-\sqrt{2}}{\mu(\sqrt{2}-1)}$$

znając zaś $\overline{w}$ można obliczyć i inne charakterystyki.

3.3 System typu G/G/1 z nieograniczoną długością kolejki

Systemy takiego typu, jak i bardziej złożone, jak np. G/G/m, analizowane są w dziale tzw. ogólnej teorii kolejek. Wiadomo jest, że wpływ

na obsługę, w dowolnym systemie kolejkowym, mają następujące dwie zmienne losowe:

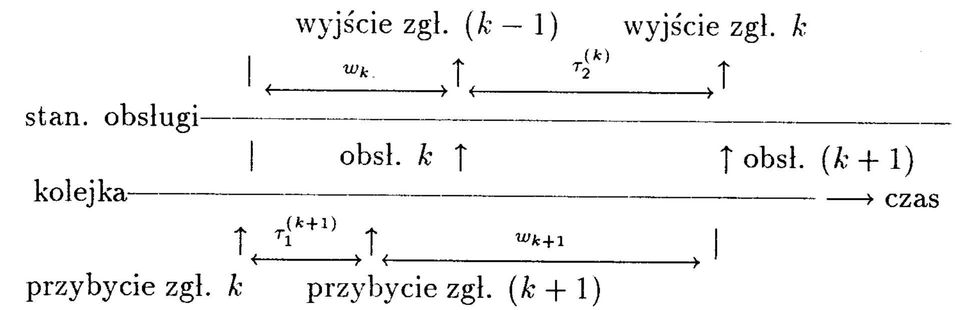
a) τ_1- odstępy czasu między przybyciem kolejnych zgłoszeń,

b) τ_2- czasy obsługi napływających do systemu zgłoszeń.

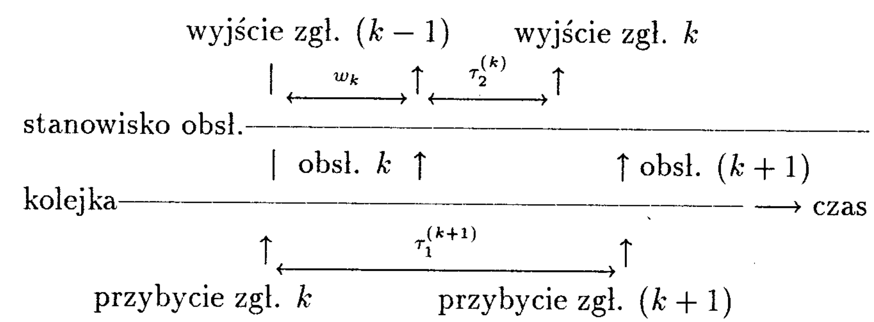
W przypadku systemu G/G/1 oba ciągi zmiennych losowych, czyli $\{\tau_1\}$ i $\{\tau_2\}$, są o dowolnych rozkładach i są niezależne od siebie.

Diagramy czasowe dla systemu G/G/1:

a) zgłoszenie z numerem $(k+1)$ zastaje system zajętym:



b) $(k+1)$ zgłoszenie zastaje system wolnym:



I w tym konkretnym przypadku, w celu uproszczenia rozważań, poszukać należy procesu Markowa. Okazuje się, że zmienne losowe τ_1 i τ_2 nie pojawiają się oddzielnie, a zawsze występują jako pewna różnica czasów. Należy więc wprowadzić nową zmienną losową związaną ze zgłoszeniem k:

$$u_k = \tau_2^{(k)} - \tau_1^{(k+1)} \qquad (3.19)$$

Zmienna ta, interpretowana jest jako różnica między czasem obsługi, który zgłoszenie z numerem k żąda od systemu i tzw. "czasem na odsapnięcie" $\tau_1^{(k+1)}$, tj. przedziałem czasu między przybyciem tego zgłoszenia a pojawieniem się zgłoszenia $(k+1)$. Może się zdarzyć, że różnica ta będzie liczbą ujemną, i tak najczęściej bywa, bo średni czas między kolejnymi zgłoszeniami jest większy od średniego czasu obsługi dla tzw. systemów stabilnych, w których:

$$\rho = \frac{\lambda}{\mu} < 1$$

I jest to prawda, albowiem gdy obliczymy wartość średnią zmiennej u_k, to otrzymamy:

$$E(u_k) = \frac{1}{\mu} - \frac{1}{\lambda} = \frac{1}{\lambda}\left(\frac{\lambda}{\mu} - 1\right) = \frac{1}{\lambda}(\rho - 1) \qquad \text{dla } \rho < 1 \tag{3.20}$$

Należy tutaj podkreślić, że wyrażenie to nie zależy od k i jest ujemne.

Powróćmy teraz do diagramów czasowych, prezentowanych powyżej. Z diagramu a) mamy:

$$w_{k+1} = w_k + \tau_2^{(k)} - \tau_1^{(k+1)} \tag{3.21}$$

i jeżeli

$$w_k + \tau_2^{(k)} - \tau_1^{(k+1)} \geq 0$$

to z tego wzoru widać, że podany warunek gwarantuje to, że zgłoszenie z numerem $(k+1)$ zastanie system zajęty.

Z diagramu b) wynika, że:

$$w_{k+1} = 0 \tag{3.22}$$

jeżeli

$$w_k + \tau_2^{(k)} - \tau_1^{(k+1)} < 0$$

i ten warunek gwarantuje, że przychodzące zgłoszenie zastaje system wolny.

łącząc wzory (3.19), (3.20), (3.21) i (3.22) otrzymamy tzw. wyrażenie Lindley'a:

$$w_{k+1} = \begin{cases} w_k + u_k & \text{jeżeli} \quad w_k + u_k \geq 0 \\ 0 & \text{jeżeli} \quad w_k + u_k < 0 \end{cases} \tag{3.23}$$

lub inaczej

$$w_{k+1} = \max\{0,\ w_k + u_k\}$$

Wyrażenie $w_k + u_k$, nazywane jest sumą "niedokończonej pracy" systemu dla w_k (od momentu przybycia zgłoszenia $(k+1)$ do rozpoczęcia obsługi zgłoszenia k), istniejącej w systemie, w momencie przybycia zgłoszenia $(k+1)$ i czasu obsługi $\tau_2^{(k)}$, dodanego do tej "niedokończonej pracy", bez brania pod uwagę przedziału czasu między przybyciem do systemu zgłoszenia k i zgłoszenia $(k+1)$, czyli $\tau_1^{(k+1)}$ (patrz diagram a)). Jeżeli wyrażenie to jest nieujemne, to jest to "niedokończona praca" systemu, którą zastaje zgłoszenie $(k+1)$ po przybyciu do systemu. Inaczej rzecz ujmując jest to czas oczekiwania na obsługę zgłoszenia $(k+1)$, tj. w_{k+1}.

Jeżeli wyrażenie $w_k + u_k$, jest ujemne, to "niedokończona praca" systemu, po przybyciu zgłoszenia z numerem k, zakończy się przed przybyciem zgłoszenia $(k+1)$ i oznacza to, że system jest wolny.

Jeśli pamięta się o tym, że zmienne losowe w ciągach $\left\{\tau_1^{(k)}\right\}$ i $\left\{\tau_2^{(k)}\right\}$ nie są ze sobą skorelowane, czyli są niezależne w każdym zbiorze swoich wartości, to ciąg zmiennych losowych:

$$\{w_0,\ w_1,\ w_2, \cdots\}$$

tworzy proces Markowa, ze stacjonarnymi prawdopodobieństwami przejść fazowych, dokładniej, jest to włożony proces Markowa. Widać to, także ze wzoru (3.23), gdzie ostatnia wartość w_{k+1}, powiązana jest z poprzedzającym ją ciągiem wielkości $\{w_i\}$, dla $i = 0,\ 1, \cdots, k$, tylko przez ostatnią wartość w_k, plus zmienna losowa u_k, która jest niezależna od zmiennych losowych w_i, dla wszystkich $i < k$.

Okazuje się, że dla systemów stabilnych, gdzie $\rho < 1$, istnieje graniczna wartość $\tilde{w}$, opisująca graniczny (stacjonarny) czas oczekiwania w kolejce. Znana jest również procedura określania dystrybuanty czasu oczekiwania w kolejce, ale trudno jest określić jej własności, a to oznacza, że ogólne wyrażenie dla średniej wartości czasu oczekiwania w kolejce, dla systemu G/G/1, nie istnieje.

Jednak parametr w można również wyrazić i przez inne wielkości. Wróćmy jeszcze raz do wyrażenia (3.23), w którym:

$$w_{k+1} = \max\{0,\ w_k + u_k\}$$

i określmy nową zmienną losową, która jest jakby "drugą częścią" czasu oczekiwania, tj.

$$y_k = -\min\{0,\ w_k + u_k\} \tag{3.24}$$

Z tych dwóch równań wynika, że jeżeli $y_k > 0$,to $w_{k+1} = 0$ i wtedy y_k jest interwałem czasu postoju systemu, który kończy się wraz z przybyciem zgloszenia $(k+1)$. Dlatego, że albo w_{k+1} lub y_k jest równe zero, to:

$$w_{k+1} \cdot y_k = 0$$

Uważa się, że czas postoju systemu istnieje, jeżeli nie jest równy zero, a w przypadku, gdy y_k oraz w_{k+1}, równocześnie równe są zeru, to przedłuża się okres zajętości systemu. Oznacza to, że w każdym przypadku (patrz diagramy czasowe):

$$w_{k+1} - y_k = w_k + u_k \tag{3.25}$$

Analizując powyższe wyrażenie, można otrzymać szereg ważnych rezultatów:

$$E[w_{k+1}] - E[y_k] = E[w_k] + E[u_k] \tag{3.26}$$

Zakładając, że $E[u_k] < 0$, co jest warunkiem równowagi statycznej $(\rho < 1)$, otrzymamy:

$$\lim_{k\to\infty} E[w_{k+1}] = \lim_{k\to\infty} E[w_k]$$

i stąd, gdy $y_k \to \tilde{y}$ i $u_k \to \tilde{u}$ otrzymamy:

$$E[\tilde{y}] = -E[\tilde{u}] \tag{3.27}$$

Należy tutaj podkreślić, że przedziały czasów przestoju są niezależne od siebie i mają jednakowy rozkład, ale konkretny czas przestoju zależy jednak od poprzedzającego go interwału zajętości.

Bezpośrednio ze wzoru (3.20) można zapisać, że:

$$E[\tilde{u}] = \frac{1}{\lambda}(\rho - 1)$$

więc

$$E[\tilde{y}] = \frac{1}{\lambda}(1 - \rho) \tag{3.28}$$

Wróćmy, jeszcze raz do wyrażenia (3.25), podnieśmy obie strony do kwadratu i potem obliczmy wartości średnie:

$$w_{k+1}^2 - \underbrace{2\ w_{k+1} \cdot y_k}_{0} + y_k^2 = w_k^2 + 2\ w_k \cdot u_k + u_k^2$$

a po przekształceniach, i gdy:

$$w_{k+1}^2 = w_k^2$$

więc

$$y_k^2 = 2\ w_k \cdot u_k + u_k^2$$

i pamiętając o tym, że w warunkach równowagi statycznej w_k nie zależy od k otrzymamy:

$$E\left[(\tilde{y})^2\right] = 2\ E[\tilde{w} \cdot \tilde{u}] + E\left[\ (\tilde{u})^2\right]$$

Oznaczając, $\overline{u^k} = E\left[(\tilde{u})^k\right]$ itd.,oraz wiedząc że w_k i u_k są od siebie niezależne można zapisać $E[\tilde{w}\ \tilde{u}] = \overline{w}\ \overline{u}$, więc:

$$\overline{y^2} = 2\ \overline{w}\ \overline{u} + \overline{u^2}$$

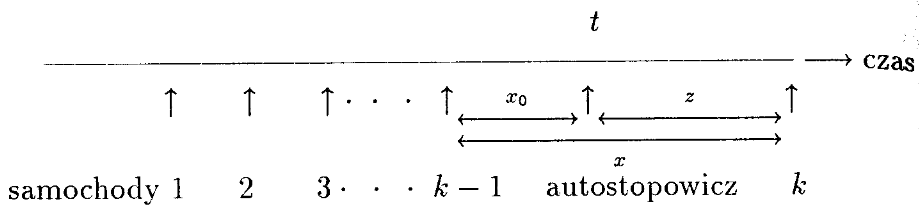
stąd

$$\overline{w} = \frac{\overline{y^2}}{2\ \overline{u}} - \frac{\overline{u^2}}{2\ \overline{u}}$$

i podstawiając za $\overline{u} = -\ \overline{y}$ otrzymamy:

$$\overline{w} = -\frac{\overline{u^2}}{2\ \overline{u}} - \frac{\overline{y^2}}{2\ \overline{y}} \tag{3.29}$$

W wyrażeniu tym występuje tzw. resztowa ilość pracy (resztowy czas pracy) pewnej zmiennej losowej. Jest to zagadnienie z teorii odnowy, które można zilustrować w następujący sposób: autostopowicz przychodzi do przydrożnej stacji benzynowej i czeka na tzw. "okazję":



gdzie z, to czas oczekiwania na przybycie kolejnego samochodu, i jest to tzw. resztowy czas pracy, pewnej zmiennej losowej x, w chwili oznaczonej jako t. Średnia wartość tego czasu równa jest:

$$\overline{z} = \frac{\overline{x^2}}{2\,\overline{x}} \tag{3.30}$$

Należy tutaj wspomnieć o tzw. paradoksie resztowego czasu, gdy w markowskim systemie kolejkowym, np. typu M/M/1, czas potrzebny na dokończenie obsługi (resztowy czas pracy), zgłoszenia znajdującego się na stanowisku obsługi, równy jest:

$$\overline{z} = \frac{\frac{2}{\mu^2}}{2\,\frac{1}{\mu}} = \frac{1}{\mu}$$

bo

$$\overline{x} = \frac{1}{\mu}, \qquad \text{oraz} \qquad \overline{x^2} = \frac{2}{\mu^2}$$

i czas ten, tzn. średni czas na dokończenie obsługi, równy jest średniemu czasowi pelnej obslugi (paradoks).

Pamiętając o tym, że we wzorze (3.29) $\overline{u}$ jest ujemna, to $\overline{w}$ równe jest średniemu resztowemu czasowi pracy dla parametru $\tilde{u}$ minus średni resztowy czas dla $\tilde{y}$.

Następnym krokiem jest obliczenie momentu drugiego rzędu dla $\tilde{u}$. Ze wzoru (3.19) mamy:

$$\overline{\tilde{u}} = \tilde{\tau}_2 - \tilde{\tau}_1$$

to

$$\overline{u^2} = \overline{(\tilde{\tau}_2 - \tilde{\tau}_1)^2} \tag{3.31}$$

Biorąc pod uwagę ten fakt, że wariancja jest różnicą momentu drugiego rzędu i kwadratu momemtu rzędu pierwszego (tj. $m_2 - m_1^2 = \sigma^2$) otrzymamy:

$$\overline{u^2} = \overline{\tau_2^2} - 2\,\overline{\tau_2}\,\overline{\tau_1} + \overline{\tau_1^2} =$$

$$= \underbrace{\overline{\tau_2^2} - \left(\overline{\tau_2}\right)^2}_{\sigma_b^2} + \left(\overline{\tau_2}\right)^2 + \underbrace{\overline{\tau_1^2} - \left(\overline{\tau_1}\right)^2}_{\sigma_a^2} + \left(\overline{\tau_1}\right)^2 - 2\,\overline{\tau_2}\,\overline{\tau_1} =$$

$$= \sigma_a^2 + \sigma_a^2 + \left(\overline{\tau_1} - \overline{\tau_2}\right)^2 = \sigma_a^2 + \sigma_b^2 + \left(\frac{1}{\lambda} - \frac{1}{\mu}\right)^2 =$$

$$= \sigma_a^2 + \sigma_b^2 + \frac{1}{\lambda^2}(1-\rho)^2 \tag{3.32}$$

gdzie, σ_a^2 i σ_b^2 - wariancje przedzialu czasu między kolejnymi zgłoszeniami i czasu obsługi, i po podstawieniach otrzymamy (podstawiając wzory (3.32), (3.28), do (3.29)):

$$\overline{w} = \frac{\sigma_a^2 + \sigma_b^2 + \frac{1}{\lambda^2}(1-\rho)^2}{\frac{2}{\lambda}\,(1-\rho)} - \frac{\overline{y^2}}{2\,\overline{y}} \tag{3.33}$$

Kolejnym etapem powinno być obliczenie momentów pierwszego i drugiego rzędu dla $\tilde{y}$, (wiadomo, że $y = \frac{1}{\lambda}\,(1-\rho)$)), ale należy zrobić to inaczej. Wprowadzmy tutaj warunkowe momenty, związane z pojawieniem się przedziału przestoju, określmy

$$r_0 = P\left\{\overline{\tilde{y}} > 0\right\} = P\left\{\text{zgłoszenie zastaje system wolnym}\right\} \tag{3.34}$$

Praca systemu kolejkowego jest stabilna, jeżeli $r_0 > 0$, a przestój pojawia się tylko wtedy, gdy ma on niezerową długość, to:

$$P\left\{\overline{\tilde{y}} \le y \mid \overline{\tilde{y}} > 0\right\} = P\left\{\text{przedział przestoju } \le y\right\} \tag{3.35}$$

Jest to dystrybuanta przedziału czasu przestoju; oznaczmy ją jako $F(y)$. Następnie oznaczmy przez I czas przestoju oraz obliczmy

$$\overline{y} = E\left[\overline{\tilde{y}} \mid \overline{\tilde{y}} = 0\right] \cdot P\left\{\overline{\tilde{y}} = 0\right\} + E\left[\overline{\tilde{y}} \mid \overline{\tilde{y}} > 0\right] \cdot P\left\{\overline{\tilde{y}} > 0\right\} =$$

$$= 0 + r_0 \, E[\overline{y} \mid \overline{y} > 0]$$

W tym wyrażeniu wartość średnia jest średnim znaczeniem dla I (czasu przestoju) i dlatego

$$\overline{y} = r_0 \, \overline{I} \tag{3.36}$$

analogicznie momemty rzędu k:

$$\overline{y^k} = r_0 \, \overline{I^k} \tag{3.37}$$

czyli

$$\frac{\overline{y^2}}{2\,\overline{y}} = \frac{r_0 \, \overline{I^2}}{2\, r_0 \, \overline{I}} = \frac{\overline{I^2}}{2\,\overline{I}}$$

i teraz podstawiając to do wzoru (3.33) otrzymamy:

$$\overline{w} = \frac{\sigma_a^2 + \sigma_b^2 + \frac{1}{\lambda^2}(1-\rho)^2}{\frac{2}{\lambda}(1-\rho)} - \frac{\overline{I^2}}{2\,\overline{I}} \tag{3.38}$$

Niestety, dalej obliczać $\overline{w}$ się nie udaje, chociaż tylko należy obliczyć dwa pierwsze momenty dla czasu przestoju, ale wielkość ta zależy od tego jaki był poprzedzający go przedział czasu zajętości systemu.

Można, oczywiście znaleźć jeszcze transformatę Laplace'a gęstości prawdopodobieństwa dla czasu oczekiwania na obsługę $\tilde{w}(s)$ (wzór ten zawiera również i rozkład czasu przestoju):

$$\tilde{w}(s) = \frac{r_0\left[1 - \tilde{I}(s)\right]}{1 - \tilde{C}(s)} \tag{3.39}$$

gdzie, $\tilde{C}(s)$- przekształcenie Laplace'a gęstości prawdopodobieństwa zmiennej losowej u. Wzór ten jest to ogólna postać równania Pollaczka-Chinczyna dla systemu G/G/1.

Znane są tylko przybliżone wzory na średni czas oczekiwania w kolejce i tak dla bardzo dużego obciążenia systemu, gdy $\rho \approx 1$ (ale $\rho < 1$) mamy:

$$\overline{w} \approx \frac{\sigma_a^2 + \sigma_b^2}{\frac{2}{\lambda}(1-\rho)} \tag{3.40}$$

oraz aproksymacja rozkładu czasu oczekiwania na obsługę (dystrybuanta) zadana jest wzorem:

$$W(y) = 1 - \exp\left(-\frac{\frac{2}{\lambda}(1-\rho)}{\sigma_a^2 + \sigma_b^2}\, y\right) \tag{3.41}$$

i są to podstawowe rezultaty otrzymane dla systemu G/G/1 przy dużych obciążeniach. Znane są również i dokładne wzory dla górnej i dolnej granicy średniej wartości czasu oczekiwania na obsługę.

Dla systemów typu G/G/c, przy dużych obciążeniach, znane są również wzory na średni czas oczekiwania, który równy jest

$$\overline{w} \approx \frac{\sigma_a^2 + \frac{\sigma_b^2}{c^2}}{\frac{2}{\lambda}(1-\rho)} \tag{3.42}$$

i znana jest aproksymacja rozkładu czasu oczekiwania na obsługę (dystrybuanta)

$$W(y) = 1 - \exp\left(-\frac{\frac{2}{\lambda}(1-\rho)}{\sigma_a^2 + \frac{\sigma_b^2}{c^2}}\, y\right) \tag{3.43}$$

oraz wzory na górną i dolną granicę dla tego średniego czasu.

3.4 Zadania

Zadanie 1. Niemarkowowskie systemy kolejkowe typu M/G/1.

Dla $\rho = 0.0,\ 0.1,\ 0.2, \cdots, 0.9,\ 0.95$ obliczyć i przedstawić graficznie następujące charakterystyki:

1. średnią liczbę zgłoszeń w systemie $\overline{n}$,
2. średnią liczbę zgłoszeń w kolejce $\overline{v}$,
3. średni czas pobytu zgłoszenia w systemie $\overline{q}$, dla $\lambda = 1.5$,
4. średni czas w kolejce $\overline{w}$, dla $\lambda = 2.5$,

dla:

a) rozkładu wykładniczego,
b) rozkładu deterministycznego,
c) $\sigma = 0.8\ \overline{\tau_2}$,
d) $\sigma = 0.6\ \overline{\tau_2}$,
e) $\sigma = 0.4\ \overline{\tau_2}$,
f) $\sigma = 0.2\ \overline{\tau_2}$

Rozwiązanie: Jak wiadomo dla rozkładu wykładniczego $\sigma = 1.0\,\overline{\tau_2}$, a dla deterministycznego $\sigma = 0.0$, więc od razu można przedstawić wyniki i wykresy porównawcze, albowiem rozkłady te wyznaczają przedziały, w których zawarte są wartości obliczanych charakterystyk, dla innych (dowolnych) rozkładów.

Średnia liczba zgłoszeń w systemie $\overline{n}$:

ρ	$\frac{\sigma}{\tau_2} = 0.0$	$\frac{\sigma}{\tau_2} = 0.2$	$\frac{\sigma}{\tau_2} = 0.4$	$\frac{\sigma}{\tau_2} = 0.6$	$\frac{\sigma}{\tau_2} = 0.8$	$\frac{\sigma}{\tau_2} = 1.0$
0.1	0.1056	0.1058	0.1064	0.1076	0.1091	0.1111
0.4	0.5333	0.5387	0.5547	0.5813	0.6187	0.6667
0.7	1.5167	1.5493	1.6473	1.8107	2.0393	2.3333
0.8	2.4000	2.4640	2.6560	2.9760	3.4240	4.0000
0.9	4.9500	5.1120	5.5980	6.4080	7.5420	9.0000
0.95	9.9750	10.3360	11.4190	13.2240	15.7510	19.0000

Średnia liczba zgłoszeń w kolejce $\overline{v}$:

ρ	$\frac{\sigma}{\tau_2} = 0.0$	$\frac{\sigma}{\tau_2} = 0.2$	$\frac{\sigma}{\tau_2} = 0.4$	$\frac{\sigma}{\tau_2} = 0.6$	$\frac{\sigma}{\tau_2} = 0.8$	$\frac{\sigma}{\tau_2} = 1.0$
0.1	0.0056	0.0058	0.0064	0.0076	0.0091	0.0111
0.4	0.1333	0.1387	0.1547	0.1813	0.2187	0.2667
0.7	0.8167	0.8493	0.9473	1.1107	1.3393	1.6333
0.8	1.6000	1.6640	1.8560	2.1760	2.6240	3.2000
0.9	4.0500	4.2120	4.6980	5.5080	6.6420	8.1000
0.95	9.0250	9.3860	10.4690	12.2740	14.8010	18.0500

Średni czas pobytu zgłoszenia w systemie $\overline{q}$ (dla $\lambda = 1.5$):

ρ	$\frac{\sigma}{\tau_2} = 0.0$	$\frac{\sigma}{\tau_2} = 0.2$	$\frac{\sigma}{\tau_2} = 0.4$	$\frac{\sigma}{\tau_2} = 0.6$	$\frac{\sigma}{\tau_2} = 0.8$	$\frac{\sigma}{\tau_2}$
0.1	0.0704	0.0705	0.0710	0.0717	0.0727	0.0741
0.4	0.3556	0.3591	0.3698	0.3876	0.4124	0.4444
0.7	1.0111	1.0329	1.0982	1.2071	1.3596	1.5556
0.8	1.6000	1.6427	1.7707	1.9840	2.2827	2.6667
0.9	3.3000	3.4080	3.7320	4.2720	5.0280	6.0000
0.95	6.6500	6.8907	7.6127	8.8160	10.5007	12.6667

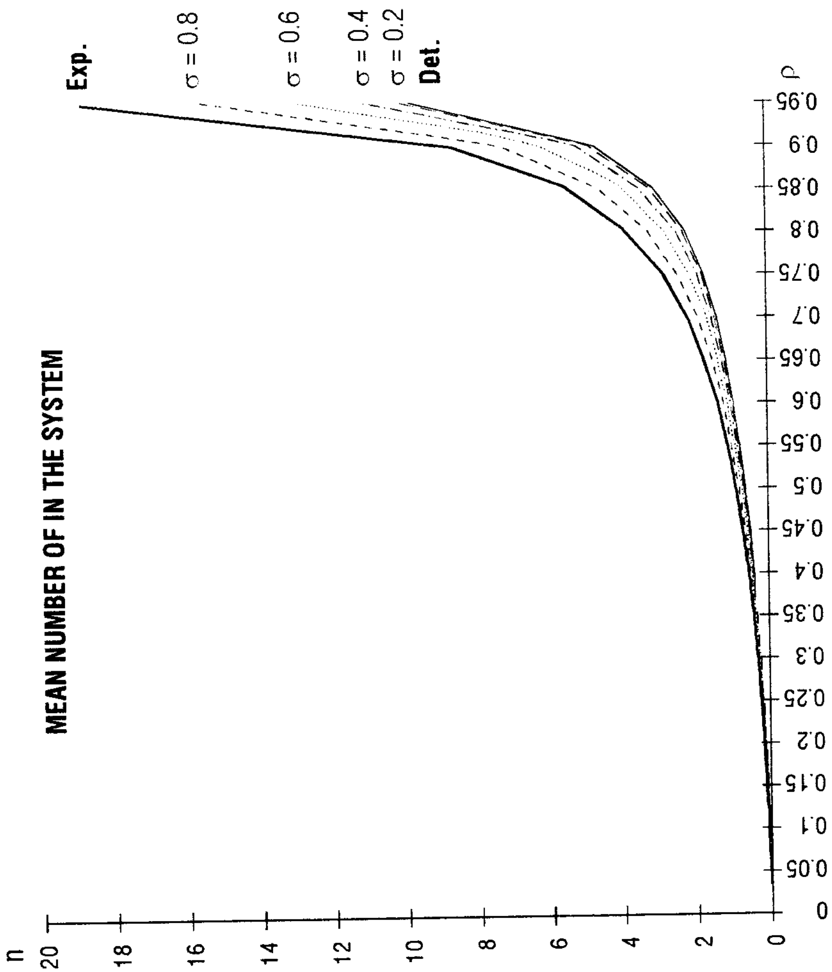
Średni czas w kolejce $\overline{w}$ (dla $\lambda = 2.5$):

ρ	$\frac{\sigma}{\tau_2} = 0.0$	$\frac{\sigma}{\tau_2} = 0.2$	$\frac{\sigma}{\tau_2} = 0.4$	$\frac{\sigma}{\tau_2} = 0.6$	$\frac{\sigma}{\tau_2} = 0.8$	$\frac{\sigma}{\tau_2} = 1.0$
0.1	0.0022	0.0023	0.0026	0.0030	0.0036	0.0044
0.4	0.0533	0.0555	0.0619	0.0725	0.0875	0.1067
0.7	0.3267	0.3397	0.3789	0.4443	0.5357	0.6533
0.8	0.6400	0.6656	0.7424	0.8704	1.0496	1.2800
0.9	1.6200	1.6848	1.8792	2.2032	2.6568	3.2400
0.95	3.6100	3.7544	4.1876	4.9096	5.9204	7.2200

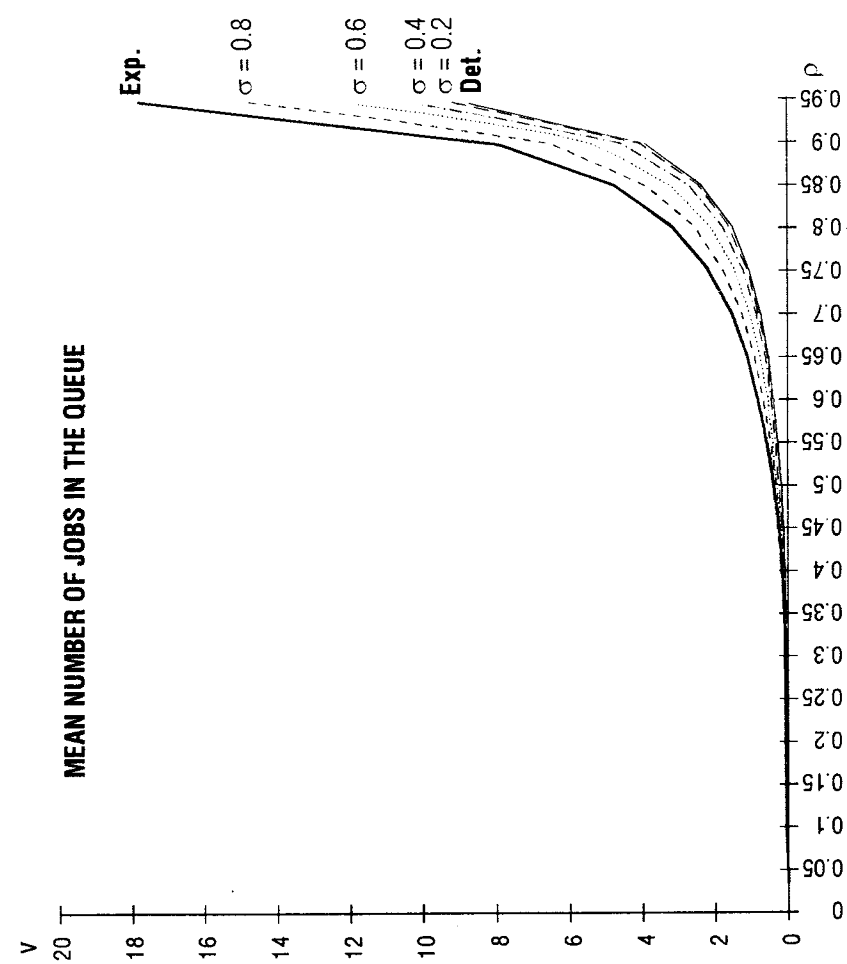
Wykresy do tych tabel przedstawione są na następnych stronach.

Analiza systemu obsługi typu M/G/1 pokazuje, że przy niewielkim obciążenu, gdy $\rho \leq 0.4$, różnica w wartościach parametrów dla różnych rozkładów jest znikoma. Dopiero gdy obciążenie systemu jest znaczne $\rho = 0.7 - 0.8$, różnice te są już dobrze widoczne, a dla systemów silnie obciążonych, gdy $\rho \geq 0.9$, są już bardzo duże. Wynika stąd praktyczny wniosek, że przy analizie systemów, założenie, że czasy obsługi mają rozkład wykładniczy, często może prowadzić do poważnych błędów w obliczeniach i, gdzie jest to możliwe, należy dokładniej badać procesy obsługi i wybierać odpowiedni rozkład dla czasów obsługi.

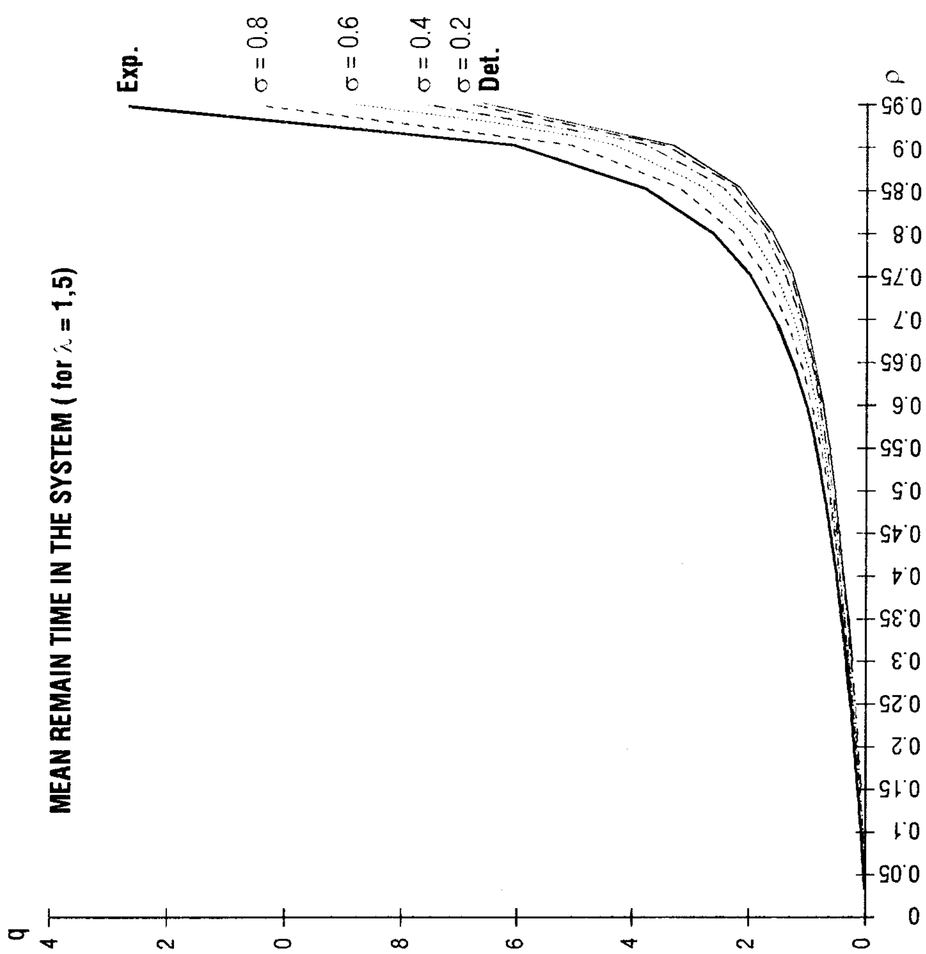
Wykresy dla średniej liczby zgłoszeń w systemie $\overline{n}$ w modelu M/G/1:



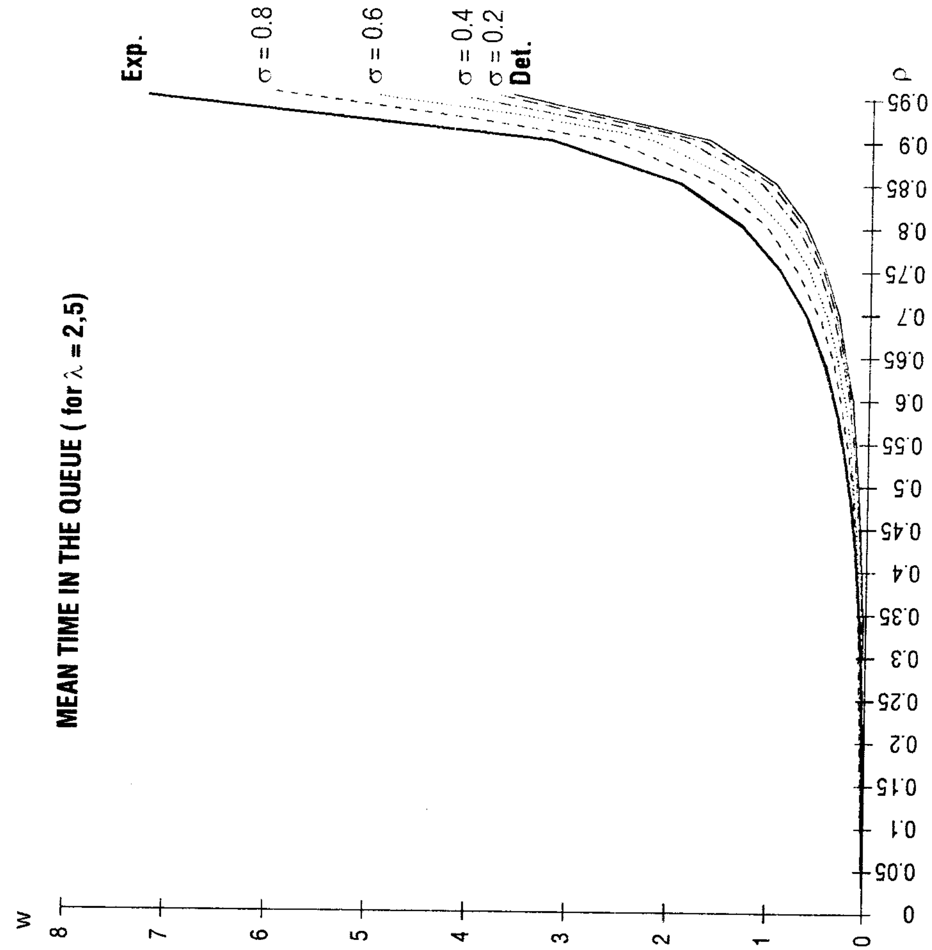
Wykresy dla średniej liczby zgłoszeń w kolejce $\overline{v}$:



Wykresy dla średniego czasu pobytu zgłoszenia w systemie $\overline{q}$ (dla $\lambda = 1.5$):



Wykresy dla średniego czasu w kolejce $\overline{w}$ (dla $\lambda = 2.5$):



Rozdział 4

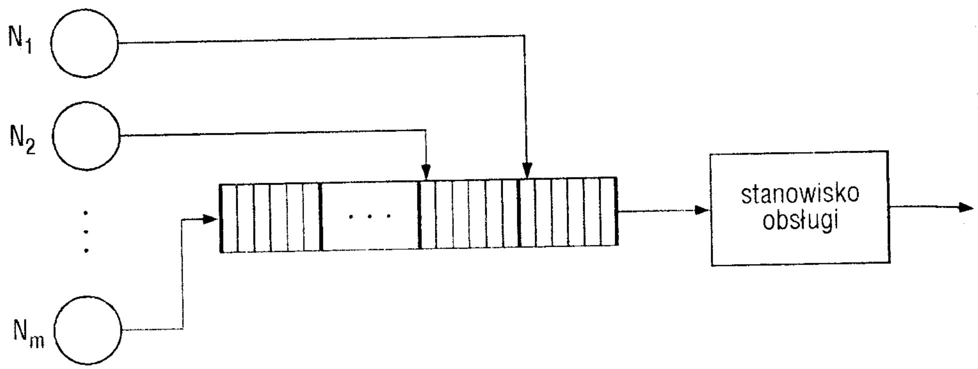
Modele masowej obsługi z priorytetowym regulaminem kolejki

Modele obsługi z priorytetowym regulaminem kolejki zajmują ważne miejsce w teorii kolejek. Jak wiadomo, stosowanie regulaminu naturalnego (FIFO) nie zawsze jest uzasadnione, a stosowanie priorytetowej obsługi, oczywiście tylko dla pewnych grup zgłoszeń, w znacznym stopniu różnicuje czasy oczekiwania w kolejce do stanowiska (stanowisk) obsługi.

Najbardziej znanym priorytetowym regulaminem obsługi jest obsługa z względnym lub absolutnym (bezwzględnym) regulaminem kolejki. O priorytecie względnym mówimy wtedy, gdy pojawienie się zgłoszenia o wyższym priorytecie niż priorytet obsługiwanego zgłoszenia nie powoduje przerwania obsługi i zgłoszenie to obsługiwane jest do końca, a po jego zakończeniu, następuje wybranie do obsługi pierwszego zgłoszenia z grupy zgłoszeń o najwyższym priorytecie.

W przypadku priorytetu absolutnego obsługa zgłoszenia o niższym priorytecie jest przerywana, czyli zgłoszenie jest rugowane i powraca na pierwsze miejsce w kolejce dla swojej klasy zgłoszeń, przy tym stan obsługi zostaje zapamiętany i to zgłoszenie jest doobsługiwane, gdy w systemie zabraknie zgłoszeń o wyższym priorytecie.

Priorytet względny

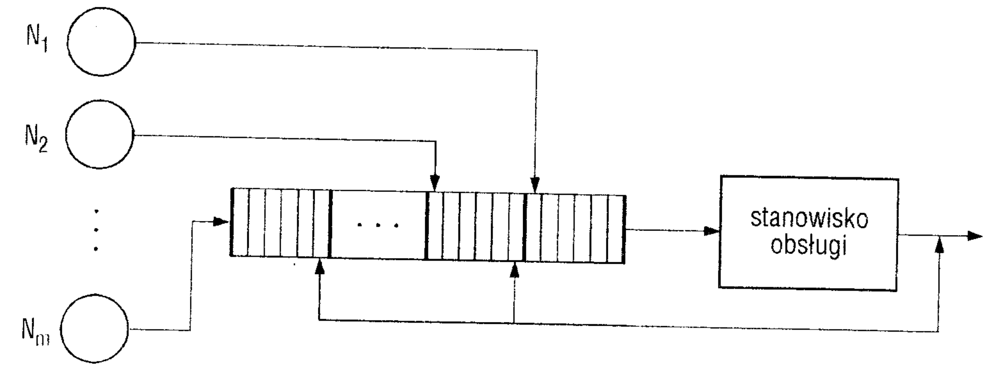


Rys. 7. Model systemu obsługi z priorytetem względnym

Istnieją również i inne regulaminy priorytetowego szeregowania, np. priorytet absolutny bez doobsługiwania, gdy zgłoszenie obsługiwane jest od nowa lub od nowa i z nowym czasem na obsługę. Mogą wystąpić też i dynamiczne priorytety, zależne od stanu kolejki itp.

Systemy z priorytetowym regulaminem kolejki pokazane są na rysunkach 7 i 8. W modelach tych zakładamy, że m niezależnych źródeł N_1, N_2, $\cdots$, N_m generuje poissonowskie strumienie zgłoszeń o parametrach λ_1, λ_2, $\cdots$, λ_m, przy tym każde ze źródeł jest niekończenie wymiarowe, czas zaś obsługi zgłoszeń ze źródła z numerem i ma dowolny rozkład o dystrybuancie $B_i(t)$ i funkcji gęstości $b_i(t)$, z wartością średnią i wariancją oznaczoną jako $\overline{s_i} = \frac{1}{\mu_i}$ i σ_i^2 (są to wielkości skończone). Inaczej to ujmując są to systemy priorytetowe typu M/G/1.

Priorytet absolutny



Rys. 8. Model systemu obsługi z priorytetem absolutnym

W takich systemach obciążenie systemu obsługą zgłoszeń ze źródła i równe jest:

$$\rho_i = \frac{\lambda_i}{\mu_i} \tag{4.1}$$

i nie zależy to, od konkretnego regulaminu szeregowania kolejki, a łączne obciążenie systemu (stanowiska lub stanowisk obsługi) wynosi:

$$\rho = \sum_{i=0}^{m} \rho_i \tag{4.2}$$

Warunkiem istnienia, w systemie jednostanowiskowym, stacjonarności jest warunek $\rho < 1$; w systemach wielostanowiskowych warunek ten dany jest jako $\frac{\rho}{c} < 1$.

Przyjmuje się, że numer źródła zgłoszeń odpowiada priorytetowi zgłoszeń, a zgłoszenia z najwyższym priorytetem przychodzą od źródła N_1, to zgłoszenia ze źródła N_m mają priorytet najniższy. W obrębie tego samego priorytetu obowiązuje regulamin naturalny (FIFO). Jeśli pamiętamy o tym, że strumienie zgłoszeń są poissonowskie, to intensywność sumarycznego, też poissonowskiego, strumienia zgłoszeń od wszystkich źródeł, równa jest:

$$\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_m = \sum_{i=1}^{m} \lambda_i \tag{4.3}$$

4.1 System obsługi typu M/G/1 z priorytetem względnym

W systemach takiego typu, przeanalizować należy w pierwszej kolejności czas oczekiwania w kolejce, dla zgłoszenia priorytetu i (dla $1 \leq i \leq m$), który składa się z:

a) czasu t_0, niezbędnego dla dokończenia obsługi wykonywanej (dla zgłoszenia dowolnego priorytetu, które jest na stanowisku obsługi) i to, że zgłoszenie jest na stanowisku obsługi, zadane jest przez prawdopodobieństwo zajętości stanowiska obsługi, równe ρ,

b) czasu t_i, potrzebnego na obsłużenie wszystkich zgłoszeń, ze źródeł od 1 do i, które znajdowały się w kolejce w chwili przybycia do systemu zgłoszenia priorytetu i,

c) czasu t'_i, niezbędnego dla obsługi zgłoszeń priorytetów od 1 do $i-1$, przybyłych do systemu w czasie gdy zgłoszenie priorytetu i czekało w kolejce, ale obsłużonych wcześniej, ze względu na swój wyższy priorytet, czyli

$$w_i = t_0 + t_i + t'_i \tag{4.4}$$

lub

$$\overline{w_i} = \overline{t_0} + \overline{t_i} + \overline{t'_i} \tag{4.5}$$

Przeanalizujmy, po kolei, wszystkie składowe powyższego równania.

1) Wartość średnią czasu na dokończenie obsługi zgłoszenia znajdującego się na stanowisku obsługi, można obliczyć od razu, ze wzoru na tzw. resztowy czas pracy, rozpatrywany dokładniej przy analizie systemu G/G/1, z uwzględnieniem tego, że system nie jest pusty. Średni resztowy czas pracy $\overline{z}$ dla zmiennej losowej x równy jest:

$$\overline{z} = \frac{\overline{x^2}}{2\,\overline{x}} \tag{4.6}$$

gdzie, $\overline{x^2}$ moment drugiego rzędu i oznacza to, że resztowy czas pracy, jest to średni czas na dokończenie obsługi zgloszenia znajdującego się na stanowisku obsługi:

$$\overline{t_0} = \frac{\overline{s^2}}{2\,\overline{s}} \cdot \rho = \frac{\frac{1}{\Lambda}\sum_{j=1}^{m} \lambda_j\,\overline{s_j^2}}{2\,\frac{1}{\Lambda}\sum_{j=1}^{m} \lambda_j \cdot \frac{1}{\mu_j}} \cdot \sum_{j=1}^{m} \rho_j = \sum_{j=1}^{m} \rho_j\, \frac{\overline{s_j^2}}{2\,\frac{1}{\mu_j}} =$$

$$= \frac{1}{2}\sum_{j=1}^{m} \frac{\overline{s_j{}^2}}{\frac{1}{\mu_j}} \cdot \frac{\lambda_j}{\mu_j} = \frac{1}{2}\sum_{j=1}^{m} \lambda_j \cdot \overline{s_j{}^2} \tag{4.7}$$

tutaj, przy obliczaniu momentu pierwszego i drugiego rzędu, wyrażenia $\frac{\lambda_j}{\Lambda}$ pełnią rolę wag, bo są to średnie ważone.

2) Oznaczmy przez k_j długość kolejki zgłoszeń priorytetu j ($j \leq i$) i bezpośrednio z określenia parametru t_i otrzymamy:

$$\overline{t_i} = \sum_{j=1}^{i} \overline{k_j} \cdot \frac{1}{\mu_j}$$

i dalej, z formuły Little'a (odnosząc to do kolejki zgłoszeń), mamy:

$$\overline{k_j} = \lambda_j \cdot \overline{w_j}$$

więc

$$\overline{t_i} = \sum_{j=1}^{i} \lambda_j\, \overline{w_j}\, \frac{1}{\mu_j} = \sum_{j=1}^{i} \rho_j\, \overline{w_j} \tag{4.8}$$

3) Dalej oznaczmy przez k'_j liczbę zgłoszeń wyższego priorytetu niż i, które napłynęły do systemu, gdy zgloszenie priorytetu i oczekiwało w kolejce, czyli które zostały obsłużone w czasie oczekiwania w_i:

$$\overline{k'_j} = \lambda_j\, \overline{w_i}$$

więc

$$\overline{t'_i} = \sum_{j=1}^{i-1} \overline{k'_j} \cdot \frac{1}{\mu_j} = \overline{w_i} \sum_{j=1}^{i-1} \rho_j \tag{4.9}$$

Podstawiając do wyrażenia (4.5) wyrażenia (4.7),(4.8) i (4.9) otrzymamy:

$$\overline{w_i} = \frac{1}{2}\sum_{j=1}^{m} \lambda_j\, \overline{s_j{}^2} + \sum_{j=1}^{i} \overline{w_j}\, \rho_j + \overline{w_i} \sum_{j=1}^{i-1} \rho_i$$

i po przekształceniach:

$$\overline{w_i} = \frac{\frac{1}{2}\sum_{j=1}^{m} \lambda_j\, \overline{s_j{}^2} + \sum_{j=1}^{i} \overline{w_j}\, \rho_j}{1 - \sum_{j=1}^{i-1} \rho_j} = \frac{\overline{t_0} + \sum_{j=1}^{i} \overline{w_j}\, \rho_j}{1 - \sum_{j=1}^{i-1} \rho_j} \tag{4.10}$$

Jest to rekurencyjny układ równań z niewiadomymi $\overline{w_i}$, który można rozwiązać zaczynając od $i = 1$, znajdując $\overline{w_1}$, dalej $\overline{w_2}$ itd.

Ogólne rozwiązanie dla średniego czasu oczekiwania w kolejce zgłoszeń priorytetu i przyjmuje następującą postać:

$$\overline{w_i} = \frac{\sum_{j=1}^{m} \lambda_j \overline{s_j^2}}{\left(1 - \sum_{j=1}^{i-1} \rho_j\right)\left(1 - \sum_{j=1}^{i} \rho_j\right)} \tag{4.11}$$

Średni czas pobytu zgłoszenia priorytetu i (dla $1 \leq i \leq m$) w systemie wynosi:

$$\overline{q_i} = \overline{w_i} + \frac{1}{\mu_i} \tag{4.12}$$

Znając powstawowe parametry czasowe, można bez trudu obliczyć średnią liczbę zgłoszeń danego priorytetu w systenie, czy średnią liczbę zgłoszeń w kolejce itp.

4.2 System typu M/G/1 z priorytetem absolutnym

Zakładamy tutaj, że jest to system z doobsługiwaniem przerwanej obsługi. Oznaczając przez $\overline{q_i}$ średni czas pobytu zgłoszenia priorytetu i w systemie można ten czas rozdzielić na następujące składowe:

1) Średni czas obsługi zgłoszenia priorytetu i, czyli $\overline{s_i}$ równy $\frac{1}{\mu_i}$.

2) Średni czas potrzebny na obsługę zgłoszeń priorytetu od 1 do i, które już były w kolejce, w momencie przybycia zgłoszenia priorytetu i.

Czas ten jest równy średniemu czasowi pobytu w kolejce, dla systemu M/G/1 zgłoszeń priorytetu od 1 do i, jeżeli zakłada się, że do systemu przychodzą wyłącznie zgłoszenia ze źródel od 1 do i, albowiem w systemie takiego typu zgłoszenia priorytetu niższego od i są jakby niezauważalne, czyli można tu wykorzystać wyrażenie na średni czas w kolejce dla systemu M/G/1 bez priorytetów (to tylko dla zgłoszeń

priorytetu od 1 do i), w którym:

$$\overline{w} = \frac{\frac{\rho^2}{\lambda} + \lambda\,\sigma^2}{2\,(1-\rho)} = \frac{\frac{\lambda^2}{\mu^2\,\lambda} + \lambda\,\left[\overline{s^2} - (\overline{s})^2\right]}{2\,(1-\rho)} =$$

$$= \frac{\frac{1}{2}\left[\frac{\lambda}{\mu^2} + \lambda\,\overline{s^2} - \frac{\lambda}{\mu^2}\right]}{1-\rho} = \frac{\frac{\lambda}{2}\cdot\overline{s^2}}{1-\rho} \tag{4.13}$$

gdzie, $\overline{s}$ i $\overline{s^2}$ momenty pierwszego i drugiego rzędu czasu obsługi; wyrażenie to można od razu wykorzystać do prowadzonych rozważań, czyli druga składowa dla średniego czasu pobytu zgłoszenia priorytetu i w systemie wynosi:

$$\frac{\frac{1}{2}\sum_{j=1}^{i} \lambda_j \overline{s_j^2}}{1 - \sum_{j=1}^{i} \rho_j}$$

3) Średni czas oczekiwania związany z obsługą zgłoszeń wyższych priorytetów, przychodzących do systemu w tym czasie, kiedy zgłoszenie priorytetu i jest w kolejce do obsługi. Średnia liczba tego typu zgłoszeń, dla priorytetu k ($k < i$) równa jest:

$$\lambda_k \cdot \overline{q_i}$$

i każde takie zgłoszenie wydłuża czas oczekiwania na obsługę dla zgłoszeń priorytetu i średnio o $\frac{1}{\mu_k}$ jednostek czasu. Czyli

$$\overline{q_i} = \frac{1}{\mu_i} + \frac{\frac{1}{2}\sum_{j=1}^{i} \lambda_j \overline{s_j^2}}{1 - \sum_{j=1}^{i} \rho_j} + \overline{q_i}\sum_{j=1}^{i-1} \frac{\lambda_j}{\mu_j}$$

i po odpowiednich przekształceniach otrzymamy:

$$\overline{q_i} = \frac{\frac{1}{\mu_i}\left(1 - \sum_{j=1}^{i} \rho_j\right) + \frac{1}{2}\sum_{j=1}^{i} \lambda_j \overline{s_j^2}}{\left(1 - \sum_{j=1}^{i} \rho_j\right)\left(1 - \sum_{j=1}^{i-1} \rho_j\right)} \tag{4.14}$$

gdzie, $\overline{s_j^2}$ - moment drugiego rzędu czasu obsługi zgłoszeń priorytetu j.

Inne parametry, jak średni czas oczekiwania w kolejce, łatwo dają się obliczyć, gdy znany jest czas $\overline{q_i}$ (dla $1 \leq i \leq m$):

$$\overline{w_i} = \overline{q_i} - \frac{1}{\mu_i} \tag{4.15}$$

Średnie długości kolejek wg priorytetów, czy średnią liczbę zgłoszeń w kolejce, można obliczyć, znając parametry czasowe i korzystając ze wzorów Little'a.

4.3 System z absolutnym priorytetem i ze skończenie wymiarowymi źródłami zgłoszeń typu M/G/1/N

W modelowaniu analitycznym systemów komputerowych ważne miejsce zajmuje pewna klasa systemów kolejkowych, w których zgłoszenia (zadania) przychodzą ze źródeł o ograniczonym wymiarze. Mogą być to np. wielodostępne systemy komputerowe z kilkoma klasami użytkowników, czy lokalne sieci komputerowe, dające się przedstawić jako systemy kolejkowe z priorytetowym szeregowaniem zadań i skończenie wymiarowymi źródłami zgłoszeń (systemy zamknięte, rys. 9).

W systemach tego typu zgłoszenia napływają od m różnych źródeł, o wymiarze N_l każde (dla $1 \leq l \leq m$) i generujących zgłoszenia o priorytetach różniących się między sobą. Zgłoszenie wyższego priorytetu przychodząc do systemu, w którym np. obsługiwane jest zgłoszenie niższego priorytetu, wywłaszcza to zgłoszenie i wywłaszczone zgłoszenie wraca na pierwsze miejsce kolejki swojego priorytetu. Zakłada się, że:

- wszystkie źródła zgłoszeń generują je niezależnie od siebie,
- czas między kolejnymi zgłoszeniami w pojedynczym źródle ma rozkład wykładniczy z wartością średnią $\overline{a_l}$ równą $\frac{1}{\lambda_l}$,
- czas obsługi zgłoszenia priorytetu l na stanowisku obsługi s_l podlega rozkładowi dowolnemu, ze skończonymi momentami pierwszego i drugiego rzędu: $\overline{s_l}$ i $\overline{s_l^2}$.

Średni czas pobytu zgłoszenia priorytetu l w systemie równy jest:

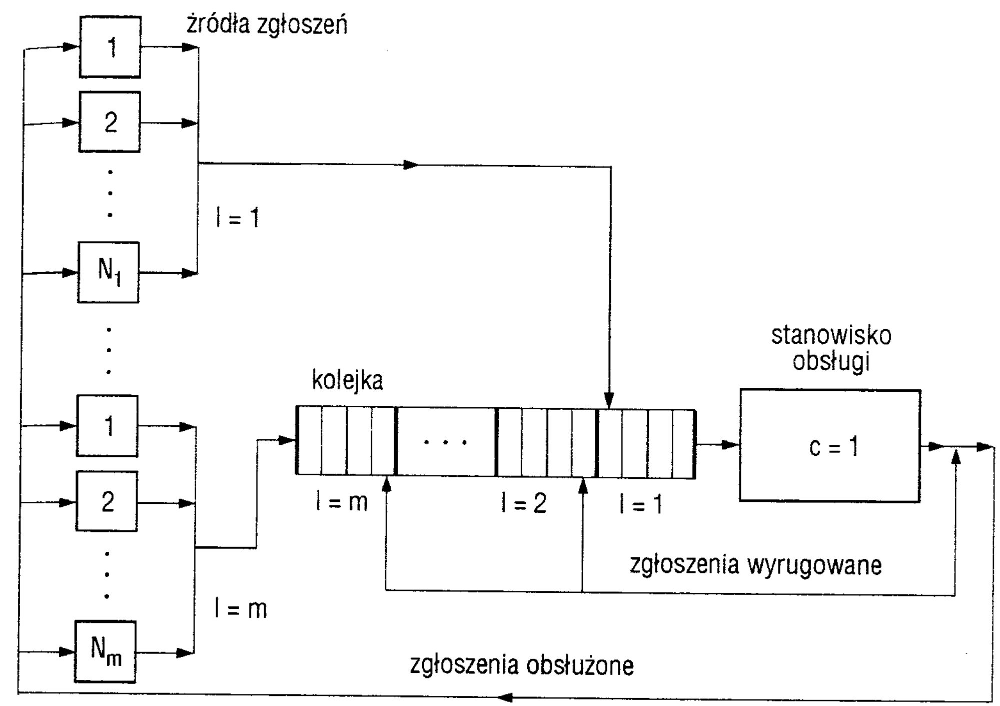
$$\overline{q_l} = \overline{w_l} + \overline{s_l} + \overline{u_l} \tag{4.16}$$

gdzie:

$\overline{w_l}$ - średni czas oczekiwania do rozpoczęcia obsługi (obsługa może być przerwana i zgłoszenie wraca na pierwsze miejsce w kolejce),

$\overline{s_l}$ - średni czas "czystej" obsługi (bez uwzględniania przerwania i nowego oczekiwania na obsługę),

$\overline{u_l}$ - średni czas przerwań, tj. ponownego oczekiwania w kolejce, po wyrugowaniu ze stanowiska obsługi.



Rys. 9. Model systemu zamkniętego z absolutnym priorytetem (M/G/1/N)

Biorąc pod uwagę to, że w systemach ze skończenie wymiarowymi źródłami zgłoszeń intensywność napływu nowych zgłoszeń zależy od stanu systemu, tj. liczby zgłoszeń w kolejkach i na stanowisku obsługi, konieczne jest wprowadzenie dla zgłoszeń priorytetu l ($1 \leq l \leq m$), tzw. okresu zajętości l - priorytetu, w którym w systemie znajdują się konkretne zgłoszenie tego priorytetu i okresu niezajętości. Wtedy, z punktu widzenia tego konkretnego zgłoszenia priorytetu l, intensywność napływu nowych zgłoszeń priorytetu l, w okresie zajętości równa jest:

$$\widehat{\Lambda_l} = \lambda_l \left(N_l - 1 - \widehat{n_l}\right) \tag{4.17}$$

gdzie:
N_l - wymiar źródła priorytetu l,
λ_l - intensywność napływu zgłoszeń od pojedynczego źródła,
$\widehat{n_l}$ - średnia liczba zgłoszeń priorytetu l, które napłynęły do systemu, w okresie zajętości l–priorytetu.

W priorytetowych systemach, zgłoszenia priorytetu l mogą zająć stanowisko obsługi, tylko wtedy, gdy będą obsłużone wszystkie zgłoszenia wyższych priorytetów, a to oczywiście wpływa na czas oczekiwania w kolejce. W okresie zajętości l - priorytetu, intensywność napływu zgłoszeń wyższych priorytetów, charakteryzujących się swoim parametrem obciążenia systemu ρ_k $(1 \leq k \leq l-1)$, równa jest:

$$\widetilde{\Lambda_k} = \lambda_k \left(N_k - \rho_k - \overline{n_k}\right) \qquad \text{dla} \qquad 1 \leq k \leq l-1 \tag{4.18}$$

gdzie:
ρ_k - współczynnik obciążenia systemu zgłoszeniami priorytetu k,
$\overline{n_k}$ - średnia liczba zgłoszeń priorytetu k w systemie.

Średnia więc intensywność napływu zgłoszeń priorytetu k (dla $1 \leq k \leq l$) wynosi

$$\Lambda_k = \begin{cases} \lambda_k \left(N_k - \rho_k - \overline{n_k}\right) & \text{dla} \quad 1 \leq k \leq l-1 \\ \lambda_k \left(N_k - 1 - \widehat{n_k}\right) & \text{dla} \qquad k = l \end{cases}$$

Powróćmy do wyrażenia (4.16) i obliczmy $\overline{u_l}$. Średnio w czasie obsługi zgłoszenia priorytetu l $(2 \leq l \leq m)$ wystąpi:

$$\Lambda_{l-1} \cdot \overline{s_l}$$

przerwań, gdzie sumaryczna intensywność zgłoszeń wyższego priorytetu równa jest:

$$\Lambda_{l-1} = \sum_{k=1}^{l-1} \Lambda_k \tag{4.19}$$

a sumaryczny współczynnik obciążenia systemu:

$$R_{l-1} = \sum_{k=1}^{l-1} \rho_k \tag{4.20}$$

oraz

$$\rho_k = \Lambda_k \, \overline{s_k} \tag{4.21}$$

Jeśli założymy, że do systemu wpływają tylko zgłoszenia priorytetu od 1 do $l-1$, to na dość długim przedziale (interwale) czasowym T system będzie zajęty $T \; R_{l-1}$ jednostek czasowych i wolny (przestój) przez $T \; (1 - R_{l-1})$. W wolnym systemie (bez zgłoszeń) sumaryczny strumień zgłoszeń równy jest:

$$M_{l-1} = \sum_{k=1}^{l-1} N_k \, \lambda_k \tag{4.22}$$

a czas do przybycia pierwszego zgłoszenia można obliczyć w następujący sposób:

$$T_{l-1} = \frac{1}{M_{l-1}}$$

oznacza to, że w przedziale T pojawiło się H zmian okresu zajętości i przestoju, czyli

$$H = \frac{T \; (1 - R_{l-1})}{T_{l-1}} \tag{4.23}$$

a długość okresu zajętości równa jest:

$$\Pi_{l-1} = \frac{T \; R_{l-1}}{H} = \frac{R_{l-1}}{(1 - R_{l-1}) \; M_{l-1}} \tag{4.24}$$

Teraz już można od razu zapisać wyrażenie na średni czas wszystkich przerwań dla zgłoszeń priorytetu l:

$$\overline{u_l} = \Lambda_{l-1} \; \overline{s_l} \; \Pi_{l-1} = \frac{R_{l-1} \; \Lambda_{l-1}}{(1 - R_{l-1}) \; M_{l-1}} \cdot \overline{s_l} \tag{4.25}$$

Dalej już można obliczyć średni czas pełnej obsługi, dla zgłoszeń priorytetu l:

$$\overline{r_l} = \overline{s_l} + \overline{u_l} = \overline{s_l} \left[1 + \frac{R_{l-1} \; \Lambda_{l-1}}{(1 - R_{l-1}) \; M_{l-1}}\right] \tag{4.26}$$

i aby obliczyć średni czas pobytu zgłoszenia priorytetu l w kolejce, należy przystąpić do obliczenia $\overline{n_l}$, czyli średniej liczby zgłoszeń tego priorytetu w systemie. W przedziale zajętości priorytetu l średnia liczba zgloszeń przybyłych do systemu w czasie oczekiwania i obsługi pojedynczego zgłoszenia priorytetu l równa jest (ze wzoru Little'a):

$$\widehat{n_l} = \widehat{\Lambda_l} \cdot \overline{q_l} = (N_l - 1 - \widehat{n_l})\ \lambda_l \cdot (\overline{w_l} + \overline{r_l}) \tag{4.27}$$

a po przekształceniach otrzymamy:

$$\widehat{n_l} = \frac{(N_l - 1)\ \lambda_l \cdot (\overline{w_l} + \overline{r_l})}{1 + \lambda_l\ (\overline{w_l} + \overline{r_l})}$$

Średnia liczba zgloszeń priorytetu k $(1 \leq k \leq l-1)$ w przedziałach zajętości i niezajętości, równa jest:

$$\widetilde{n_k} = \frac{N_k\ \lambda_k\ (\overline{w_k} + \overline{r_k})}{1 + \lambda_k\ (\overline{w_k} + \overline{r_k})}$$

Łącząc te dwa wyrażenia otrzymamy:

$$\overline{n_k} = \begin{cases} \frac{(N_k-1)\ \lambda_k\ (\overline{w_k}+\overline{r_k})}{1+\lambda_k\ (\overline{w_k}+\overline{r_k})} & \text{dla} \quad k = l \\ \\ \frac{N_k\ \lambda_k\ (\overline{w_k}+\overline{r_k})}{1+\lambda_k\ (\overline{w_k}+\overline{r_k})} & \text{dla} \quad 1 \leq k \leq l-1 \end{cases} \tag{4.28}$$

W wyrażeniu tym jedynym nieznanym składnikiem jest $\overline{w_k}$.

Te warunki powinny być spełnione, by zgłoszenie priorytetu l trafilo na stanowisko obsługi:

- obsłużone do końca lub wyrugowane zgłoszenie aktualnie obsługiwane,
- dokończona obsługa zgłoszeń priorytetu l i wyższych, których obsluga zostala przerwana,
- obsłużone zgłoszenia priorytetu l i wyższych, które są już w kolejce,
- obsłużone nowe zgłoszenia wyższych priorytetów, które przybyły do systemu w czasie, gdy zgłoszenie l priorytetu czekalo w kolejce.

Obliczmy te komponenty.

1) Średni czas na dokończenie obsługi zgłoszenia ze stanowiska obsługi, równy:

$$\overline{t_{1_l}} = \sum_{k=1}^{l} \rho_k\ \Delta_k = \sum_{k=1}^{l} \Lambda_k\ \overline{s_k}\ \Delta_k \tag{4.29}$$

gdzie Δ_k - średni czas doobsługiwania zgłoszenia priorytetu k $(1 \leq k \leq l)$, równy

$$\Delta_k = \frac{\overline{s_k^2}}{2\ \overline{s_k}}$$

więc

$$\overline{t_{1_l}} = \sum_{k=1}^{l} \Lambda_k \cdot \overline{s_k} \cdot \frac{\overline{s_k^2}}{2\ \overline{s_k}} \tag{4.30}$$

2) Średni czas dokończenia obsługi wyrugowanych zgłoszeń priorytetu od 2 do l (nie więcej niż jedno z każdego priorytetu). Czas ten daje się obliczyć, gdy przemnożymy i zsumujemy średnie czasy doobsługiwania przez prawdopodobieństwa, że zgłoszenie priorytetu k przerwie obsługę zgłoszenia priorytetu m $(k < m)$:

$$\overline{t_{2_l}} = \sum_{k=2}^{l} (\overline{r_k} - \overline{s_k})\ \Lambda_k\ \Delta_k \tag{4.31}$$

3) Średni czas obsługi zgłoszeń priorytetu od 1 do l znajdujących się w kolejce. Wiadomo, że średnia liczba zgłoszeń k priorytetu w kolejce równa jest

$$\overline{w_k} \cdot \Lambda_k$$

to

$$\overline{t_{3_l}} = \sum_{k=1}^{l} \overline{s_k} \cdot \overline{w_k} \cdot \Lambda_k \tag{4.32}$$

4) Średni czas obsługi nowych zgłoszeń wyższych priorytetów. Dla priorytetu k $(k < l)$ średnia liczba nowych zgłoszeń, przybyłych do systemu w czasie $\overline{w_l}$, równa jest:

$$\overline{w_l} \cdot \Lambda_k$$

a średni czas ich obsługi:

$$\overline{w_l} \cdot \Lambda_k \cdot \overline{s_k}$$

to

$$\overline{t_{4_l}} = \overline{w_l} \sum_{k=1}^{l-1} \Lambda_k \cdot \overline{s_k} \tag{4.33}$$

Czyli średni czas oczekiwania na rozpoczęcie obsługi, dla zdarzenia priorytetu l jest równy:

$$\overline{w_l} = \overline{t_{1_l}} + \overline{t_{2_l}} + \overline{t_{3_l}} + \overline{t_{4_l}} \tag{4.34}$$

a po podstawieniu:

$$\overline{w_l} = \sum_{k=1}^{l} \Lambda_k \, \overline{s_k} \, \frac{\overline{s_k^2}}{2\,\overline{s_k}} + \sum_{k=2}^{l} (\overline{r_k} - \overline{s_k}) \; \Lambda_k \, \frac{\overline{s_k^2}}{2\,\overline{s_k}} +$$

$$+ \sum_{k=1}^{l} \overline{s_k} \, \overline{w_k} \, \Lambda_k + \overline{w_l} \sum_{k=1}^{l-1} \Lambda_k \, \overline{s_k} \tag{4.35}$$

Podstawiając w to wyrażenie wartości odpowiednich parametrów i grupując wyrazy otrzymamy równanie kwadratowe z niewiadomą $\overline{w_l}$:

$$a_0 \cdot (\overline{w_l})^2 + a_1 \cdot \overline{w_l} + a_2 = 0 \tag{4.36}$$

gdzie:

$$a_0 = 2 \; \lambda_l \; (1 - R_l)$$

$$a_1 = 2 + 2 \; \lambda_l \left[\overline{r_l} - \sum_{k=1}^{l-1} \rho_k \left(\frac{\overline{s_k^2}}{2\,\overline{s_k}} + \overline{w_k} \right) - \sum_{k=2}^{l-1} (\overline{r_k} - \overline{s_k}) \; \Lambda_k \, \frac{\overline{s_k^2}}{2\,\overline{s_k}} \right] -$$

$$-2 \; [(N_l - 1) \; \lambda_l \; \overline{s_l} + R_{l-1} \; (1 + \lambda_l \; \overline{r_l})]$$

$$a_2 = -2 \left[\sum_{k=1}^{l-1} \rho_k \left(\frac{\overline{s_k^2}}{2\,\overline{s_k}} + \overline{w_k} \right) + \sum_{k=2}^{l-1} (\overline{r_k} - \overline{s_k}) \; \Lambda_k \, \frac{\overline{s_k^2}}{2\,\overline{s_k}} \right] (1 + \lambda_l \; \overline{r_l}) -$$

$$-(N_l - 1) \; \lambda_l \; \overline{s_l^2} \left[1 + \underbrace{\left\{ \frac{\overline{r_l} - \overline{s_l}}{\overline{s_l}} \right\}} \right]$$

przy tym wyrażenie w $\underbrace{\{\}}$ tylko dla $l \geq 2$.

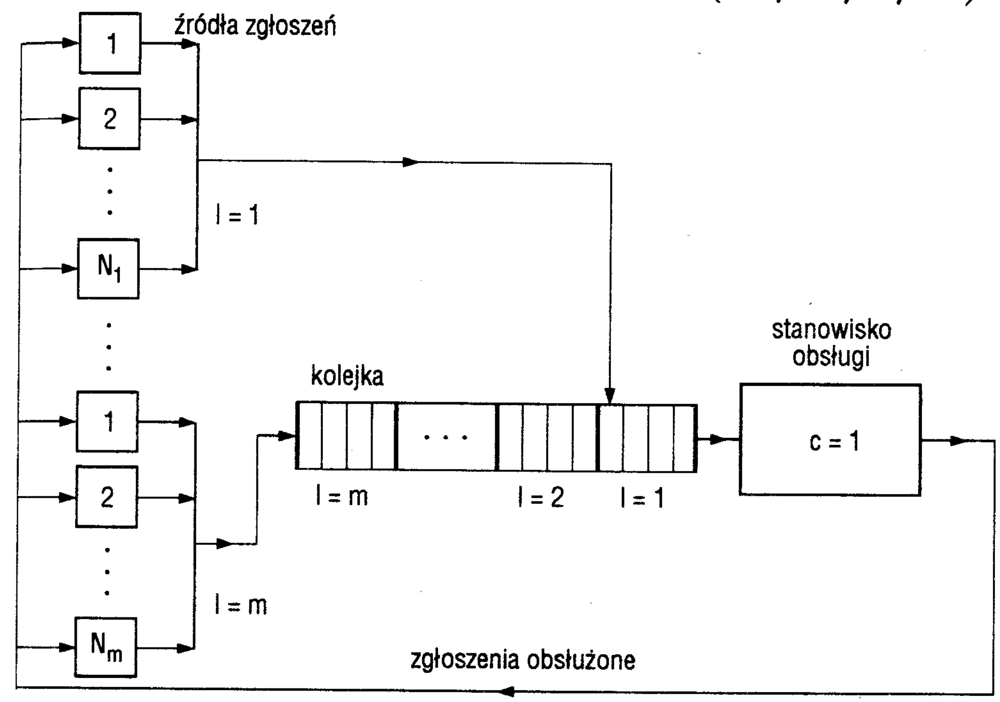
Analiza tego trójmianu kwadratowego pokazuje, że współczynnik a_0 w stanie ustalonym jest zawsze dodatni, albowiem $R_{l-1} < 1$. Współczynnik a_2 jest ujemny, więc niezależnie od znaku współczynnika a_1, równanie to ma tylko jeden dodatni pierwiastek. Oznacza to, że otrzymane zostało jednoznaczne rozwiązanie, względem $\overline{w_l}$.

Podstawowe parametry, charakteryzujące proces przejścia przez system zgłoszeń różnych priorytetów, obliczane są na podstawie danych

wejściowych i parametrach odnoszących się do zgłoszeń wyższych priorytetów. Pozwala ta na analizę systemu dla każdego priorytetu oddzielnie, zaczynając od $l = 1$, gdy system kolejkowy może być rozpatrywany jako system bez priorytetów.

Znając $\overline{w_l}$ (dla $1 \leq l \leq m$), można obliczyć czasy reakcji systemu dla zgłoszeń każdego priorytetu i inne charakterystyki.

4.4 System obsługi typu zamkniętego z priorytetem względnym (M/G/1/N)



Rys. 10. Model systemu zamkniętego z priorytetem względnym

Jest to jednostanowiskowy system ze skończenie wymiarowemi źródłami zgłoszeń, pokazany na rys. 10. Założenia wejściowe dla systemów tego typu, podobne są do tych z priorytetem absolutnym, gdzie do systemu napływają zgłoszenia z m źródeł, o wymiarach N_l każde, dla $1 \leq l \leq m$, tylko tutaj przychodzące zgłoszenie, czeka na zakończenie obsługi zadań (zgłoszeń) niższych priorytetów. Przypaminamy, że czas między kolejnymi zgłoszeniami od pojedynczego źródła ma rozkład wykładniczy, a czas obsługi zgłoszenia priorytetu l, na stanowisku obsługi, rozkład dowolny, z momentami pierwszego i drugiego rzędu $\overline{s_l}$ i $\overline{s_l^2}$.

Średni czas pobytu zgłoszenia priorytetu l (dla $1 \leq l \leq m$) w systemie równy jest:

$$\overline{q_l} = \overline{w_l} + \overline{s_l} \tag{4.37}$$

gdzie:

w_l - średni czas pobytu w kolejce,

s_l - średni czas obsługi.

Podobnie jak w systemach z priorytetem absolutnym, wprowadzimy pojęcie przedziału zajętości dla zgłoszeń priorytetu l (dla $1 \leq l \leq m$) i opierając się na przeprowadzonych tam rozważaniach od razu napiszemy wyrażenie na średnią intensywność napływu zgłoszeń priorytetu k (dla $1 \leq k \leq m$):

$$\Lambda_k = \begin{cases} \lambda_k (N_k - \rho_k - \overline{n_k}) & \text{dla} \quad k \neq l \\ \lambda_k (N_k - 1 - \widehat{n_k}) & \text{dla} \quad k = l \end{cases} \tag{4.38}$$

gdzie:

N_k - wymiar źródła priorytetu k,

λ_k - intensywność napływu zgłoszeń od pojedynczego źródła,

$\overline{n_k}$ - średnia liczba zgłoszeń priorytetu k w systemie,

$\widehat{n_k}$ - średnia liczba zgłoszeń priorytetu k, które napłynęły do systemu w przedziale zajętości k - priorytetu,

ρ_k - współczynnik obciążenia systemu zgłoszeniami k - priorytetu.

W systemach typy zamkniętego obciążenia systemu zgłoszeniami priorytetu k, wynosi

$$\rho_k = \Lambda_k \overline{s_k} \qquad \text{dla} \qquad 1 \leq k \leq m \tag{4.39}$$

a sumaryczny współczynnik obciążenia systemu:

$$R_k = \sum_{i=1}^{k} \rho_i \tag{4.40}$$

Obliczmy średnią liczbę zgłoszeń priorytetu l w systemie:

$$\widehat{n_l} = \widehat{\Lambda_l}\, \overline{q_l} = (N_l - 1 - \widehat{n_l})(\overline{w_l} + \overline{s_l})$$

stąd

$$\widehat{n_l} = \frac{\lambda_l (N_l - 1)(\overline{w_l} + \overline{s_l})}{1 + \lambda_l (\overline{w_l} + \overline{s_l})} \tag{4.41}$$

Podobnie postępujemy i z innymi priorytetami ($k \neq l$):

$$\widetilde{n_k} = \frac{\lambda_k N_k (\overline{w_k} + \overline{s_k})}{1 + \lambda_k (\overline{w_k} + \overline{s_k})} \tag{4.42}$$

Dalej łącząc te dwa wyrażenia, otrzymamy:

$$\overline{n_k} = \begin{cases} \frac{\lambda_k (N_k - 1)(\overline{w_k} + \overline{s_k})}{1 + \lambda_k (\overline{w_k} + \overline{s_k})} & \text{dla} \quad k = l \\ \frac{\lambda_k N_k (\overline{w_k} + \overline{s_k})}{1 + \lambda_k (\overline{w_k} + \overline{s_k})} & \text{dla} \quad k \neq l \end{cases} \tag{4.43}$$

We wzorze tym jedyną niewiadomą jest średni czas oczekiwania w kolejce zgłoszeń priorytetu l ($1 \leq l \leq m$), czyli w_l. Czas ten jest sumą trzech komponentów. Obliczmy kolejno każdy z nich:

1) Średni czas na dokończenie obsługi zgłoszenia aktualnie obsługiwanego:

$$\overline{t_{1_l}} = \sum_{k=1}^{m} \rho_k \Lambda_k = \sum_{k=1}^{m} \Lambda_k \overline{s_k} \Delta_k \tag{4.44}$$

gdzie

$$\Delta_k = \frac{\overline{s_k^2}}{2\, \overline{s_k}}$$

2) Średni czas obsługi zgłoszeń priorytetu od 1 do l, które już wcześniej trafiły do kolejki. Korzystając ze wzorów Little'a można obliczyć średnią długość kolejki dla zadań priorytetu k ($k < l$):

$$\overline{w_k} \Lambda_k$$

i średni czas jej obsługi:

$$\overline{w_k} \Lambda_k \overline{s_k}$$

czyli

$$\overline{t_{2_l}} = \sum_{k=1}^{l} \overline{w_k} \Lambda_k \overline{s_k} \tag{4.45}$$

3) Średni czas obsługi nowych zgłoszeń priorytetu od 1 do $l-1$, które przybędą do systemu w czasie, gdy zgłoszenie priorytetu l czeka w kolejce. Ten średni czas obsługi dla zgłoszeń priorytetu k równy jest:

$$\overline{w_l}\ \Lambda_k\ s_k$$

$$t_{3_l} = \overline{w_l} \sum_{k=1}^{l-1} \Lambda_k\ \overline{s_k} \tag{4.46}$$

Po połączeniu tych wyrażeń, dla zgłoszeń priorytetu l, gdzie $1 \leq l \leq m$, otrzymamy:

$$\overline{w_l} = \sum_{k=1}^{m} \Lambda_k\ \overline{s_k}\ \frac{\overline{s_k^2}}{2\ \overline{s_k}} + \sum_{k=1}^{l} \overline{w_k}\ \Lambda_k\ \overline{s_k} + \overline{w_l} \sum_{k=1}^{l-1} \Lambda_k\ \overline{s_k} \tag{4.47}$$

Rozwijając to wyrażenie i odpowiednio grupując wyrazy, otrzymamy równanie kwadratowe w postaci:

$$a_0 \cdot (\overline{w_l})^2 + a_1 \cdot \overline{w_l} + a_2 = 0 \tag{4.48}$$

gdzie:

$$a_0 = \lambda_l\ (1 - R_{l-1})$$

$$a_1 = (1 - R_{l-1})(1 + \lambda_l\ \overline{s_l}) - \lambda_l\ (N_l - 1)\,\overline{s_l} -$$

$$-\lambda_l\ \left[\sum_{k=1}^{l-1} \rho_k \left(\frac{\overline{s_k^2}}{2\ \overline{s_k}} + \overline{w_k}\right) + \sum_{k=l+1}^{m} \rho_k\ \frac{\overline{s_k^2}}{2\ \overline{s_k}}\right]$$

$$a_2 = -(1 + \lambda_l\ \overline{s_l}) \left[\sum_{k=1}^{l-1} \rho_k \left(\frac{\overline{s_k^2}}{2\ \overline{s_k}} + \overline{w_k}\right) + \sum_{k=l+1}^{m} \rho_k\ \frac{\overline{s_k^2}}{2\ \overline{s_k}}\right] - \frac{(N_l - 1)\ \lambda_l\ \overline{s_l^2}}{2}$$

W równaniu tym współczynnik a_0 jest zawsze dodatni, a_2 - zawsze ujemny, więc niezależnie od znaku współczynnika a_1, równanie to ma jeden dodatni pierwiastek i jest to wartość parametru w_l. Analiza równania (4.48) pokazuje, że współczynniki a_1 i a_2 dają się obliczyć w oparciu o początkowe dane i parametry charakteryzujące proces obsługi zgłoszeń wyższych priorytetów, oprócz jednego wyrażenia:

$$C_l = \sum_{k=l+1}^{m} \rho_k\ \frac{\overline{s_k^2}}{2\ \overline{s_k}} \tag{4.49}$$

Oznacza to, że miary wydajności (parametry) dla jednostanowiskowego systemu obsługi z priorytetem względnym, można wyliczyć tylko metodą kolejnych iteracji, rozpoczynając proces obliczeń od najwyższego priorytetu. W pierwszym kroku iteracji zakładamy, że $C_l = 0$ i obliczamy $\overline{w_l}$, $\overline{n_l}$, ρ_l, Λ_l, R_l, dla $1 \leq l \leq m$, a potem ze wzoru (4.49) wyliczymy C_l i wtedy można zacząć nowy krok iteracji. Oczywiście liczba potrzebnych iteracji zależy od założonej dokładności ϵ i proces ten można zakończyć po:

$$\left|\frac{\overline{w_l^{(p)}} - \overline{w_l^{(p-1)}}}{\overline{w_l^{(p)}}}\right| < \epsilon$$

gdzie:

p - numer iteracji.

Rozdział 5

Sieci stanowisk obsługi (sieci kolejkowe)

5.1 Wiadomości wstępne

Sieci kolejkowe (angielska nazwa queueing networks lub networks of queues) jako pewien wyodrębniony rozdział w teorii masowej obsługi z powodzeniem mogą być zastosowane do modelowania i oceny parametrów bardziej złożonych systemów i sieci komputerowych. Modele takie wymogają często dość złożonych obliczeń i dlatego ważne są nie tylko analityczne rezultaty, ale również odpowiednia numeryczna implementacja takich modeli. Dlatego też tam gdzie brak jest dokładnego rozwiązania lub jest ono bardzo pracochłonne, wykorzystuje się przybliżone metody obliczeń.

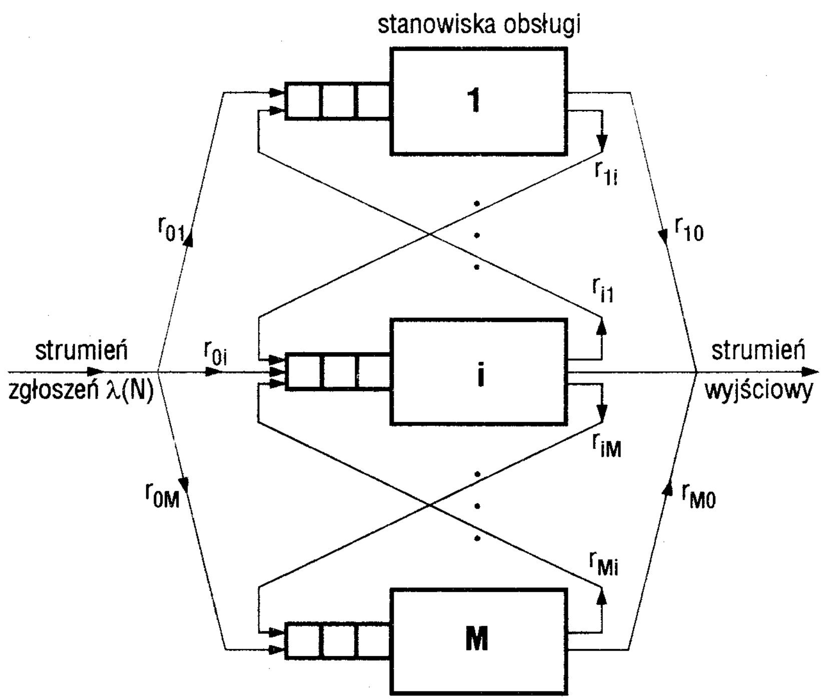
Ogólnie stosowane dziś modele takich złożonych systemów to tzw. sieci stanowisk obsługi, złożone, z dowolnej liczby stanowisk z dowolną jednak, racjonalną, topologią (rys. 11).

W zadanym modelu stanowiska reprezentują zasoby systemu, a zgłoszenia obsługiwane w nich reprezentują programy, zadania (lub ich fragmenty) użytkowników.

Opis każdego stanowiska obsługi obejmuje rozkład czasu obsługi zgłoszeń oraz regulamin szregowania kolejki. Opis ten uzupełniają prawdopodobieństwa przejść między każdą dowolną, parą stanowisk.

Sieć może być zamkniętą, gdy zawiera stałą liczbę przebywających w niej zgłoszeń lub otwartą, gdy zgłoszenia wchodzą do sieci w dowolnym punkcie i opuszczają ją, też w dowolnym punkcie.

Zgłoszenia są nierozróżnialne lub zgrupowane w pewne klasy, gdy zgłoszenia różnych klas mają inne czasy obsługi i różną trasę do przebycia w sieci. Zgłoszenia mogą też różnić się priorytetami.



Rys. 11 Ogólna postać sieci stanowisk obsługi

Wprowadźmy tutaj następujące, nowe, charakterystyczne tylko dla sieci, parametry:

M - liczba stanowisk (węzłów) w sieci,

N - liczba zgłoszeń (klientów, zadań) wszystkich klas w sieci,

$\mathbf{N} = (N^1, \cdot\cdot\cdot, N^K)$ - wektor zgłoszeń,

K - liczba klas zgłoszeń,

λ- intensywność napływu zgłoszeń.

a dla stanowiska o numerze i:

$A_i(x)$, $a_i(x)$ - dystrybuanta i funkcja gęstości rozkładu czasu między zgłoszeniami, z wartością średnią równą $\frac{1}{\lambda_i}$,

$\sigma^2_{a_i}$- wariancja odstępów czasu między zgłoszeniami,

$C^2_{a_i} = \sigma^2_{a_i} \cdot \lambda^2_i$- współczynnik zmienności (znormalizowana wariancja) rozkładu $A_i(x)$,

$B_i(x)$, $b_i(x)$ - dystrybuanta i funkcja gęstości rozkładu czasu obsługi $s_i = \frac{1}{\mu_i}$ - średni czas obsługi jest albo stały albo zależny od liczby zgłoszeń na stanowisku obsługi, wtedy $s_i = s_i(n_i)$,

$\sigma^2_{b_i}$, $C^2_{b_i} = \sigma^2_{b_i} \cdot \mu^2_i$ - wariancja i współczynnik zmienności rozkładu $B_i(x)$,

n_i - liczba zgłoszeń na stanowisku i,

λ_i - intensywność napływu strumienia zgłoszeń do stanowiska,

e_i - średnia liczba pobytów zgłoszenia na stanowisku i, $e_i = \frac{\lambda_i}{\lambda}$,

μ_i - intensywność obsługi,

ρ_i - obciążenie stanowiska obsługi,

r_{ij} - prawdopodobieństwo, że zgłoszenie po zakończeniu obsługi na stanowisku i przejdzie na stanowisko j,

r_{i0} - prawdopodobieństwo, że po zakończeniu obsługi na stanowisku i zgłoszenie opuści sieć,

r_{0i} - prawdopodobińswo, że zgłoszenie rozpocznie obsługę w sieci od stanowiska i, w sieci zamkniętej $r_{0i} = r_{i0} = 0$,

$R = [r_{ij}]$ - macierz tranzycji.

W przypadku gdy mamy wiele różnych klas zgłoszeń, parametry odnoszące się np. do klasy k mają dodatkowy górny indeks:

N^k - liczba zgłoszeń klasy k,

n^k_i - liczba zgłoszeń klasy k na stanowisku i,

s^k_i - średni czas obsługi zgłoszenia klasy k na stanowisku i,

λ^k_i - intensywność napływu zgłoszeń klasy k do stanowiska i,

t^k_i - średni czas reakcji (pobytu) zgłoszenia klasy k na stanowisku i.

Wartości parametrów e_i, oraz λ_i można otrzymać z tzw. układu równań ruchu zgłoszeń. I tak dla sieci otwartej mamy:

$$\lambda_i = \lambda_{0i} + \sum_{j=1}^{M} \lambda_j \, r_{ji} \qquad \text{dla } i = 1, \cdot\cdot\cdot, M \tag{5.1}$$

w przypadku jednej klasy zgłoszeń, oraz:

$$\lambda_i^k = \lambda_{0i}^k + \sum_{j=1}^{M} \sum_{l=1}^{k} \lambda_j^k \, r_{ji}^{lk} \quad \text{dla } i = 1, \cdots, M, \quad k = 1, \cdots, K \qquad (5.2)$$

Poprzez rozwiązanie modelu sieciowego, rozumie się z reguły określenie stacjonarnego rozkładu prawdopodobieństw wszystkich rozróżnialnych stanów sieci i jeżeli np. przyjmiemy za stan sieci liczbę zgłoszeń na poszczególnych stanowiskach, to rozwiązaniem jest:

$$p(n_1, n_2, \cdots, n_M) = p(\mathbf{n}) \qquad (5.3)$$

gdzie, $(n_1, n_2, \cdots, n_M)$ jest to wektor opisujący stan sieci, a $p(\mathbf{n})$ jest prawdopodobieństwem takiego stanu, w którym, np. na stanowisku k jest n_k zgłoszeń, na stanowisku $k+1$ jest n_{k+1} itd.

W praktyce modelowania bardziej interesujące jest określenie kilku, wynikających z rozkładu $p(\mathbf{n})$, wielkości zwanych miarami pracy sieci (parametrami), jak:

a) rozkład prawdopodobieństwa $p_i(n_i)$, tj. tego, że liczba zgłoszeń na stanowisku obsługi i równa jest n_i lub średnia liczba zgłoszeń na tym stanowisku,

b) średni czas oczekiwania w kolejkach poszczególnych stanowisk,

c) przepustowość, tj. średnia liczba zgłoszeń przechodzacych przez zadane stanowisko w jednostce czasu,

d) obciążenie (wykorzystanie) stanowiska, np. i równego ρ_i,

e) czas reakcji t , tj. średni czas przejścia zgłoszenia przez sieć lub między zadanymi węzłami.

Należy tutaj podkreślić, że nie ma jeszcze ogólnej metody rozwiązania takiego problemu i wszystko zależy od szczególnych założeń odnoszących się do rozkładu czasu obsługi, regulaminu kolejki, rozmiaru sieci, itp.

5.2 Markowowskie modele sieci kolejkowych

W poprzednich rozdziałach zostały omówione takie modele masowej obsługi, w których zgłoszenia przechodziły tylko jedną operację obsługi na pojedynczym stanowisku obsługi, a w sieciach kolejkowych zgłoszenia obsługiwane są na wielu stanowiskach, przy tym przez niektóre stanowiska

zgłoszenia przechodzą wielokrotnie. Interesuje nas tutaj opis procesów stochastycznych (strumieni zgłoszeń) w takich wielostanowiskowych sieciach.

Badania natury takich strumieni rozpoczęły się od prac P.J. Burke (dla stanowisk połączonych szeregowo), który pokazał, że dla stacjonarnych systemów kolejkowych wejściowy strumień Poissona, po przejściu przez stanowisko obsługi z wykładniczym rozkładem czasem obsługi, tworzy strumień wyjściowy, który jest także strumieniem Poissona z parametrem (intensywność) takim samym jak i wejściowy.

To twierdzenie Burke'ego daje się rozszerzyć i na systemy typu M/M/c z wejściowym strumieniem Poissona, z parametrem λ, w których rozkład czasu obsługi zgłoszenia na dowolnym stanowisku (kanale) obsługi jest wykładniczy z parametrem (intensywnością) μ i wtedy też strumień wyjściowy jest strumieniem Poissona z parametrem λ. Pozwala to na wydzielenie poszczególnych stanowisk (węzłów) sieci i na analizę ich niezależnie od pozostałych stanowisk sieci.

Sieć o dowolnej konfiguracji badał w swoich pracach J. R. Jackson i była to sieć składająca się z M węzłów i każdy węzeł np. i, składał się z c_i równoległych kanałów obsługi, z wykładniczym rozkładem czasu obsługi z parametrem μ_i, a gdy $M = 1$, był to klasyczny system typu M/M/c i należało tylko obliczyć sumaryczną intensywność strumienia zgłoszeń dochodzących od innych stanowisk i z zewnątrz, do wybranego stanowiska sieci, przy założeniu, że:

$$\frac{\lambda_i}{c_i \, \mu_i} < 1 \qquad (5.4)$$

Jackson wykazał, że każdy węzeł sieci zachowuje się tak jakby był on niezależnym systemem kolejkowym typu M/M/c z wejściowym strumieniem Poissona z intensywnością λ_i. Stan takiej sieci, w której jest M węzłów, określony jest przez wektor $(n_1, n_2, \cdots, n_M)$, gdzie n_i jest liczbą zgłoszeń na stanowisku i (łącznie z obsługiwanym). Oznaczając przez $p(n_1, n_2, \cdots, n_M)$ stacjonarne prawdopodobieństwo tego stanu, wtedy $p_i(n_i)$ będzie marginalnym prawdopodobieństwem tego, że w stanie równowagi statycznej, w węźle i znajduje się n_i zgłoszeń.

Jackson udowodnił, że łączne stacjonarne prawdopodobieństwo stanu dla wszystkich węzłów rozkłada się na iloczyn marginalnych prawdopodobieństw dla każdego węzła (twierdzenie Jacksona):

$$p(n_1,\ n_2,\ \cdots,\ n_M) = p_1(n_1)\cdot p_2(n_2)\ \cdots\ p_M(n_M) \tag{5.5}$$

gdzie, $p_i(n_i)$ $(i = 1, \cdots M)$- stacjonarne prawdopodobieństwo w klasycznym modelu M/M/c (dla $i = 1, \cdots M$).

Modyfikacją sieci Jacksona jest sieć zamknięta badana przez W. J. Gordona i J. F. Newella, w której krąży stała i skończona liczba zgłoszeń, bez możliwości wejścia i wyjścia, czyli:

$$\sum_{i=1}^{M} n_i = N \tag{5.6}$$

Stała liczba zgłoszeń w sieci pozwala nam na określenie jedynego stacjonarnego rozkładu prawdopodobieństw $p(n_1,\ n_2,\ \cdots,\ n_M)$, gdyż zbiór stanów sieci jest zbiorem ograniczonym.

Oczywiste jest, że liczba różniących się od siebie stanów sieci równa jest liczbie sposobów rozłożenia N zgłoszeń po M węzłach sieci, czyli:

$$\begin{pmatrix} M+N-1 \\ N-1 \end{pmatrix} = C^{N-1}_{M+N-1} \tag{5.7}$$

i jest to liczba $N-1$ elementowych kombinacji bez powtórzeń z $M+N-1$ elementów.

Dla pojedynczego stanowiska obsługi (patrz system typu $M/M/1/\infty$), w którym liczba zgłoszeń przyjmuje wartości całkowite $n = 0,\ 1, \cdots$, stan procesu stochastycznego jest w pełni określony przez wartość parametru n, a graf stanów systemu i równania wiążące stacjonarne prawdopodobieńastwa tych stanów, mają następującą postać:

$$H_{n-1}\ \underset{\longleftrightarrow}{\mu_n \qquad \lambda_{n-1}}\ H_n\ \underset{\longleftrightarrow}{\mu_{n+1} \qquad \lambda_n}\ H_{n+1}$$

$$0 = -\underbrace{p(n)\ [\lambda(n)+\mu(n)]}_{\text{ze stanu } n \text{ do } n-1 \text{ lub } n+1} + \underbrace{p(n-1)\ \lambda(n-1)}_{\text{ze stanu } n-1 \text{ do } n} + \underbrace{p(n+1)\ \mu(n+1)}_{\text{ze stanu } n+1 \text{ do } n} \tag{5.8}$$

$$\mu(0) = 0$$

gdzie, $n = 0, 1, \cdots$, oraz λ i μ - mogą być zależne od liczby zgłoszeń na konkretnym stanowisku obsługi i jest to wyrażenie z procesu narodzin i śmierci, w którym użyto nowych oznaczeń, jak:

$$\begin{aligned} p_n &\to p(n) \\ \lambda_n &\to \lambda(n) \\ \mu_n &\to \mu(n) \end{aligned}$$

a rozwiązaniem jest:

$$p(n) = p(0)\ \Lambda(n)\ \beta(n)\ s^n \tag{5.9}$$

gdzie:

$$\Lambda(n) = \prod_{i=0}^{n-1}\lambda(i), \qquad \beta(n) = \prod_{i=1}^{n}\frac{1}{C(i)}, \qquad C(i) = \mu(i)\ s \tag{5.10}$$

a s jest średnim czasem obsługi.

Pod warunkiem, że:

$$p(0) = \left[1 + \sum_{n=1}^{\infty}\Lambda(n)\ \beta(n)\ s^n\right]^{-1} > 0 \tag{5.11}$$

co jest warunkiem istnienia stanu ustalonego i jest to równoważne warunkowi $\rho < 1$.

Jest to ogólna postać rozwiązania, przydatna do dalszych rozważań, w którym:

$\lambda(i)$ - intensywność przejścia ze stanu systemu (stanowiska) i, w wydzielonym stanowisku obsługi, do stanu $i+1$,

$\mu(i)$ - intensywność przejścia ze stanu i do stanu $i-1$ (czyli intensywność obsługi).

Należy tutaj jeszcze raz zaznaczyć, że indeks i nie odnosi się do numeru stanowiska obsługi, jest tylko stanem stanowiska obsługi, na wydzielonym stanowisku obsługi. Jeżeli założymy, że:

$$\mu(1) = \mu(2) = \mu(3) = \cdots = \mu$$

to $s = \frac{1}{\mu}$ i wtedy:

$$p(n) = p(0)\ \Lambda(n)\ \frac{1}{\underbrace{\mu^n}_{s^n}}$$

i można to wyrażenie porównać do sposobu obliczania prawdopodobieństw stanów w systemach opisywanych procesem narodzin i śmierci (patrz rozdział 1 i 2), gdzie:

$$p_i = p_0 \frac{\lambda_0 \cdot \lambda_1 \cdot \cdot \cdot \lambda_{i-1}}{\mu_1 \cdot \mu_2 \cdot \cdot \cdot \mu_i} = p_0 \frac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k}$$

Dodatkowym problemem jest tutaj określenie wartości s, gdy mamy K klas zgłoszeń, a zgłoszenia klasy $i = 1, \cdot \cdot \cdot, K$, występują z prawdopodobieństwem a^i, gdzie i to indeks klasy, więc:

$$s = \sum_{i=1}^{K} a^i \, s^i \qquad \text{przy} \; \sum_{i=1}^{K} a^i = 1$$

Dla sieci otwartej z M stanowiskami (węzłami) tego typu bilans równowagi stanu dla $\mathbf{n} = (n_1, n_2, \cdot \cdot \cdot, n_M)$ wygląda następująco:

$$0 = \underbrace{- \, p(\mathbf{n}) \left[\lambda(N) + \sum_{i=1}^{M} \mu_i(n_i) \, (1 - r_{ii}) \right]}_{\text{przejście ze stanu } \mathbf{n} \text{ do innych stanów}} +$$

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

$$+ \sum_{i=1}^{M} p \left[\mathbf{n}(i^-) \right] \underbrace{\lambda(N-1) \; r_{0i}}_{\text{dopływ zgł.}} + \underbrace{\sum_{i=1}^{M} p \left[\mathbf{n}(i^+) \right] \; \mu_i(n_i + 1) \; r_{i0}}_{\text{przejście od } \mathbf{n}(i^+) \text{ do } \mathbf{n}, \text{ czyli od } n_i+1 \text{ do } n_i} \tag{5.12}$$

gdzie:

$\lambda(N)$ - intensywność dopływu zgłoszeń w sieci otwartej, a z prawdopodobieństwem r_{0i} kierowane są one do stanowiska i, więc intensywność dopływu zgłoszeń do tego stanowiska równa jest: $\lambda_{0i} = \lambda(N) \; r_{0i}$,

$\mathbf{n}(i^-) = (n_1, \cdot \cdot \cdot, n_i - 1, \cdot \cdot \cdot, n_M)$, gdzie $n_i - 1$ jest stanem sąsiednim do n_i, czyli zgłoszenie przybywa (wpływa) do sieci, więc system przechodzi ze stanu $n_i - 1$ do stanu n_i,

$\mathbf{n}(i^+) = (n_1, \cdot \cdot \cdot, n_i + 1, \cdot \cdot \cdot, n_M)$ - opuszczenie sieci (ubywa jedno zgłoszenie, bo system przechodzi ze stanu $n_i + 1$ do stanu n_i),

$\mathbf{n}(i^+, j^-) = (n_1, \cdot \cdot \cdot, n_i + 1, \cdot \cdot \cdot, n_j - 1, \cdot \cdot \cdot, n_M)$ - przejście między węzłami, zgłoszenie opuszcza węzeł i $(n_i + 1)$, przechodząc do węzła j $(n_j - 1)$ z prawdopodobieństwem r_{ij}.

Są to stany sąsiednie do $\mathbf{n}$, zaś rozwiązanie równania (5.12) jest:

$$p(\mathbf{n}) = \frac{1}{G} \prod_{i=1}^{M} \lambda(N) \; \beta_i(n_i) \; (e_i \; s_i)^{n_i} \tag{5.13}$$

a dla sieci zamkniętej, gdy $r_{0i} = r_{i0} = 0$, dla $i = 1, 2, \cdot \cdot \cdot, M$:

$$p(\mathbf{n}) = \prod_{i=1}^{M} \beta_i(n_i) \; (e_i \; s_i)^{n_i} \tag{5.14}$$

gdzie:

G - jest stałą normalizacyjną, gwarantującą, że suma prawdopodobieństw jest równa jedności:

$$G = \sum_{n} p(\mathbf{n})$$

e_i - jest rozwiązaniem układu równań zwanych równaniami ruchu:

$$e_i = r_{0i} + \sum_{j=1}^{M} e_j \; r_{ji} \qquad \text{dla } i = 1, \cdot \cdot \cdot, M \tag{5.15}$$

a z definicji:

$$e_i = \frac{\lambda_i}{\lambda} \tag{5.16}$$

i określa średnią liczbę wizyt zgłoszenia przechodzącego przez sieć, na stanowisku obsługi i,

$e_i \; s_i$ - jest łącznym średnim zapotrzebowaniem zgłoszenia na obsługę przez stanowisko i.

Dla sieci otwartej rozwiązanie układu równań (5.15) jest jednoznaczne, a $e_i \; \lambda(N)$ określa strumień zgłoszeń przechodzących przez stanowisko i. Jeżeli λ, oraz μ_i, $i = 1, \cdot \cdot \cdot, M$ są stałe, tzn. nie zależą od obciążenia systemu, to:

$$p(\mathbf{n}) = \prod_{i=1}^{M} \rho_i^{n_i} (1 - \rho_i) \qquad \text{gdzie } \rho_i = \frac{\lambda_i}{\mu_i} = \frac{\lambda \; e_i}{\mu_i} \tag{5.17}$$

Poważnym ograniczeniem modeli przedstawionych powyżej jest założenie o wykładniczych rozkładach $a(t)$ i $b(t)$. Dla rozkładów bardziej ogólnych, gdy chcemy pozostać w sferze modeli markowskich, można wprowadzić rozkłady będące kombinacją rozkładów wykladniczych, jak np. rozkład prawie ogólny Coxa, którym można aproksymować dowolne rozkłady. I tak dla sieci, w której $a(t)$ podlega rozkładowi wykładniczemu a obsługa rozkładowi Coxa z F fazami, można też zestawić równania bilansu prawdopodobieństw stanu dla aktualnie wykonywanej fazy.

Pojedyncze stanowisko (węzeł) oraz cała sieć (w stanie równowagi statycznej) charakteryzują się tym, że intensywność wejścia do danego stanu, związana z przybyciem zgłoszenia do stanowiska, równa jest intensywności wyjścia z tego stanu, gdy zgłoszenie opuszcza stanowisko, więc równanie (5.8) można rozbić na dwa równania, jak:

$$0 = -p(n)\ \lambda(n) + p(n+1)\ \mu(n+1) \qquad \text{jak } H_n \overset{\mu_{n+1} \qquad \lambda_n}{\longleftrightarrow} H_{n+1}$$

$$0 = -p(n)\ \mu(n) + p(n-1)\ \lambda(n-1) \qquad \text{jak } H_{n-1} \overset{\mu_n \qquad \lambda_{n-1}}{\longleftrightarrow} H_n \tag{5.18}$$

a równanie (5.12), też można rozbić na równania:

$$0 = -p(\mathbf{n})\ \mu_i(n_i) + e_i\ \lambda(N-1)\ p\left[\mathbf{n}(i^-)\right] \qquad \text{dla } i = 1, \cdots, M \tag{5.19}$$

gdzie, $e_i\ \lambda(N-1)$ - strumień zgłoszeń przechodzący przez stanowisko i.

Podobnej operacji dla rozkładu Coxa zrobić nie można.

Równania (5.8) i (5.12) uwzględniające wszystkie dopuszczalne przejścia między stanami, nazywane są równaniami równowagi globalnej. Równania (5.18) i (5.19), związane z ruchem jednego zgłoszenia - równaniami równowagi lokalnej.

Wsród modeli Markowa jest więc klasa modeli posiadających analityczne rozwiązanie równań równowagi lokalnej w postaci:

$$p(\mathbf{n}) = \frac{1}{G} \prod_{i=1}^{M} h_i(n_i) \tag{5.20}$$

w której postać $h_i(n_i)$ zależy tylko od własności stanowiska i (jest to funkcja ρ_i).

Modele posiadające rozwiązanie w postaci (5.20) nazywane są sieciami dekomponowalnymi (separowalnymi) lub sieciami o rozwiązaniu iloczynowym Dla pozostałych modeli Markowa trzeba szukać rozwiązania równań równowagi globalnej drogą numeryczną.

Stanowiska z Coxowskim rozkładem czasu obsługi pozostają w równowadze globalnej tylko w przypadku, gdy obsługa zgłoszenia rozpoczyna się natychmiast po nadejściu. Są to stanowiska z dostatecznie wieloma identycznymi, równoległymi kanałami obsługi, aby żadne zgłoszenie nie czekało w kolejce (jest to infinite server model).

Do modelu sieci można wprowadzić klasy zgłoszeń, regulaminy priorytetowego szeregowania kolejki. Taka sieć może być dla jednych klas zgłoszeń otwarta, dla innych zamknięta. Taka sieć nosi nazwę sieci typu BCMP (skrót od nazwisk : Baskett F., Chandy M., Muntz R., Palacios J.), a marginalne prawdopodobieństwo tego, że na stanowisku i znajduje się n_i zgłoszeń, zadane jest następującym wzorem:

$$p(n_i) = \frac{1}{G} \prod_{i=1}^{M} \Lambda(N)\ \beta_i(n_i)\ s_i^{n_i} \qquad \text{przy } s_i = \sum_{k=1}^{K} e_i^k\ s_i^k \tag{5.21}$$

gdzie: k indeks klasy.

Oczywiście dla jednej klasy zgłoszeń mamy:

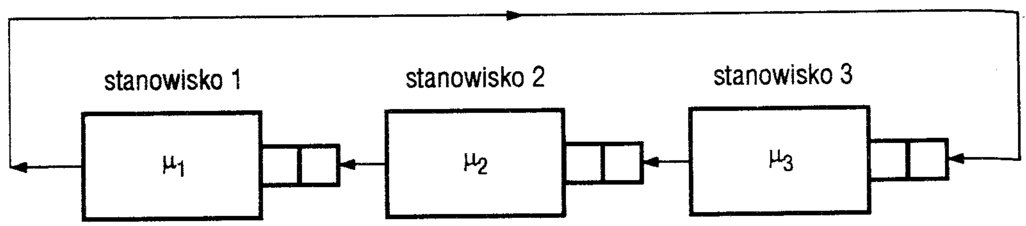
$$p_i(n_i) = (1-\rho_i)\ \rho^{n_i} \tag{5.22}$$

Pomimo istnienia rozwiązania w jawnej formie, jego praktyczne wykorzystanie nie jest proste, bo jeszcze należy wyznaczyć stałą normalizacyjną G. I tak dla sieci zamkniętej z regulàminem naturalnym i jedną klasą zgłoszeń, gdy liczba stanowisk jest równa 10, a liczba zgłoszeń krążących w sieci też 10, to w takiej sieci liczba stanów przekracza 92000 i tyle będzie wyrażeń (5.14). Wyrażenia te, opierając się na definicji $G = \sum_n p(\mathbf{n})$ trzeba obliczyć, zsumować, przeskalować, a dopiero potem przystąpić do obliczania miar wydajności (parametrów) sieci. Dla dużych sieci zadanie to praktycznie staje się niewykonalne.

Zwykle w praktyce, aby ominąć te trudności, stosuje się jedną z dwóch metod obliczeniowych:

- metodę splotową,
- metodę wartości średnich.

Przykład 1. Dla zadanej cyklicznej sieci z liczbą stanowisk $M = 3$ i stałą liczbą zgłoszeń w sieci $N = 2$, obliczyć prawdopodobieństwa wszystkich możliwych, rozróżnialnych stanów sieci:



Parametry: $r_{13} = r_{32} = r_{21} = 1.0$, reszta przejść $r_{ij} = 0.0$.

Stan takiej sieci określony jest wektorem liczby zgłoszeń na stanowiskach (węzłach): $(n_1,\ n_2,\ n_3)$, gdzie n_i jest liczbą zgłoszeń na stanowisku i:

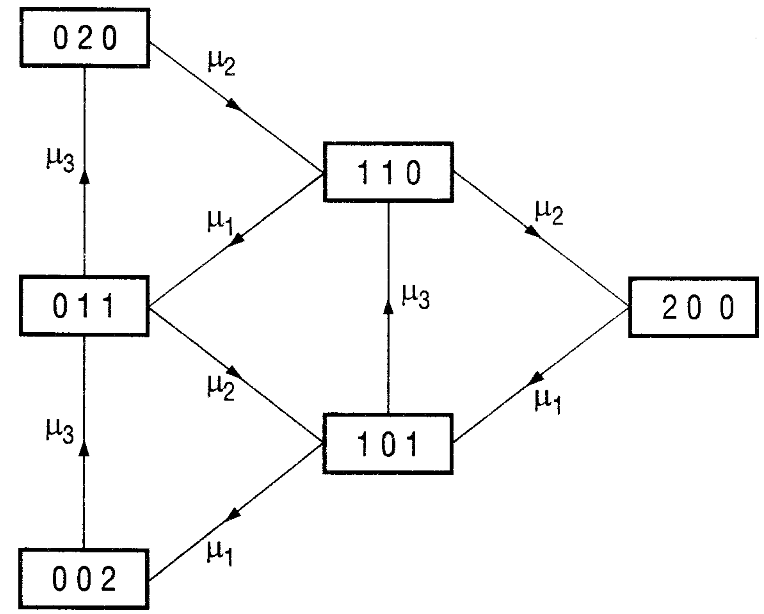
$$n_1 + n_2 + n_3 = 2$$

Rozróżnialnych stanów sieci mamy dokładnie:

$$\binom{M+N-1}{N-1} = 6$$

Intensywność przejść między stanami pokazana jest na diagramie przejść.

Międzystanowy diagram przejść:



Układ równań równowagi globalnej, dla zadanej sieci, zawiera sześć równań (jedno jest zbędne, bo dodatkowe równanie wypływa z warunku normalizacyjnego prawdopodobieństw) o następującej postaci:

$$\mu_1\, p(2,\ 0,\ 0) = \mu_2\, p(1,\ 1,\ 0)$$

$$\mu_2\, p(0,\ 2,\ 0) = \mu_3\, p(0,\ 1,\ 1)$$

$$\mu_3\, p(0,\ 0,\ 2) = \mu_1\, p(1,\ 0,\ 1)$$

$$\mu_1\, p(1,\ 1,\ 0) + \mu_2\, p(1,\ 1,\ 0) = \mu_2\, p(0,\ 2,\ 0) + \mu_3\, p(1,\ 0,\ 1)$$

$$\mu_2\, p(0,\ 1,\ 1) + \mu_3\, p(0,\ 1,\ 1) = \mu_3\, p(0,\ 0,\ 2) + \mu_1\, p(1,\ 1,\ 0)$$

$$\mu_1\, p(1,\ 0,\ 1) + \mu_3\, p(1,\ 0,\ 1) = \mu_2\, p(0,\ 1,\ 1) + \mu_1\, p(2,\ 0,\ 0)$$

Każde z tych równań równowagi globalnej ma taką postać, że lewa strona odpowiada strumieniowi wychodzącemu z ukazanego stanu, a prawa - strumieniowi wchodzącemu w zadany stan.

Pierwsze trzy równania są, równaniami równowagi lokalnej. Pozostałe trzy równania zbudowane są tak, że pierwszy składnik lewej strony jest zbalansowany pierwszym składnikiem prawej strony, to samo dotyczy i składników drugich (jest to, tzw. "chodzenie po trójkącie" - patrz strzałki na rysunku).

Przekształćmy czwarte równanie w następujący układ równowagi lokalnej:

$$\mu_1\, p(1,\ 1,\ 0) = \mu_2\, p(0,\ 2,\ 0)$$

$$\mu_2\, p(1,\ 1,\ 0) = \mu_3\, p(1,\ 0,\ 1)$$

W pierwszym równaniu intensywność strumienia wychodzącego ze stanu (1, 1, 0), gdy zgłoszenie opuszcza stanowisko 1, przyrównuje się do intensywności strumienia wchodzącego do tego stanu, kiedy zgłoszenie przybywa do stanowiska. To samo jest z drugim równaniem opisującym stanowisko 2 i to jest zasada równowagi lokalnej. W taki oto sposób powstało 9 równań równowagi lokalnej, z których 4 są zbyteczne, plus równanie normalizacyjne prawdopodobieństw:

$$\begin{aligned}
\mu_1\, p(2,\ 0,\ 0) &= \mu_2\, p(1,\ 1,\ 0)\\
\mu_2\, p(0,\ 2,\ 0) &= \mu_3\, p(0,\ 1,\ 1)\\
\mu_3\, p(0,\ 0,\ 2) &= \mu_1\, p(1,\ 0,\ 1)\\
\mu_1\, p(1,\ 1,\ 0) &= \mu_2\, p(0,\ 2,\ 0)\\
\mu_2\, p(1,\ 1,\ 0) &= \mu_3\, p(1,\ 0,\ 1)\\
\mu_2\, p(0,\ 1,\ 1) &= \mu_3\, p(0,\ 0,\ 2)\\
\mu_3\, p(0,\ 1,\ 1) &= \mu_1\, p(1,\ 1,\ 0)\\
\mu_1\, p(1,\ 0,\ 1) &= \mu_2\, p(0,\ 1,\ 1)\\
\mu_3\, p(1,\ 0,\ 1) &= \mu_1\, p(2,\ 0,\ 0)\\
\sum_{n\in A} p(n_1,\ n_2,\ n_3) &= 1
\end{aligned}$$

gdzie, A jest zbiorem wektorów n.

Jeżeli taki układ ma rozwiązanie, to rozwiązanie to spełnia również układ równań równowagi globalnej, czyli posiada jednoznaczne rozwiązanie,

a rozwiązanie to jest takiego oto typu:

$$p(1,\ 0,\ 1) = \frac{\mu_1}{\mu_3}\, p(2,\ 0,\ 0)$$

$$p(1,\ 1,\ 0) = \frac{\mu_1}{\mu_2}\, p(2,\ 0,\ 0)$$

$$p(0,\ 1,\ 1) = \frac{(\mu_1)^2}{\mu_2\, \mu_3}\, p(2,\ 0,\ 0)$$

$$p(0,\ 0,\ 2) = \left(\frac{\mu_1}{\mu_3}\right)^2 p(2,\ 0,\ 0)$$

$$p(0,\ 2,\ 0) = \left(\frac{\mu_1}{\mu_2}\right)^2 p(2,\ 0,\ 0)$$

$$p(2,\ 0,\ 0) = \left[1 + \frac{\mu_1}{\mu_3} + \frac{\mu_1}{\mu_2} + \frac{(\mu_1)^2}{\mu_2\, \mu_3} + \left(\frac{\mu_1}{\mu_3}\right)^2 + \left(\frac{\mu_1}{\mu_2}\right)^2\right]^{-1}$$

5.2.1 Metoda splotowa

W metodzie tej, obliczanie stałej normalizacyjnej opiera się na prostym spostrzeżeniu, że dla sieci o rozwiązaniu iloczynowym, wartość stałej G, miedzy innymi, jest funkcją rozmiaru sieci, jak i liczby zgłoszeń krążących w sieci, czyli $G = G(M,N)$, więc można tę stałą wyrazić poprzez stałe dla sieci zmniejszonej o jedno stanowisko i zawierającej mniejszą liczbę zgłoszeń. I tak: dla sieci zamkniętej z jedną klasą zgłoszeń otrzymamy:

$$G(M,N) = \sum_{n=0}^{N} G_{i^-}(M-1, M-n)\ h_i(n) \tag{5.23}$$

gdzie, indeks i^- wskazuje, że stała normalizacyjna odnosi się do sieci, z której usunięto stanowisko o numerze i, a $h_i(n)$ zależy tylko od własności stanowiska i (jest funkcją ρ_i). Obliczenie stałej $G(M,N)$ sprowadza się więc do rekurencyjnego liczenia splotu $(N+1)$ - wymiarowych wektorów (od $n = 0$ do N, czyli $N+1$).

W przypadku wielu klas zgłoszeń i gdy zgłoszenia nie mogą zmieniać klasy ($N^k = const$) w miejscu splotu wektorowego, pojawiają się sploty macierzy K - wymiarowych, w których indeks k przyjmuje wartości od 0 do N^k.

Jeżeli zgłoszenia mogą zmieniać klasy, to zbiór K klas można podzielić na $C \leq K$ podzbiorów, zwanych łańcuchami, pomiędzy którymi nie ma wymiany zgłoszeń i tak określoną sieć można transformować w sieć z C klasami zgłoszeń, bez możliwości zmiany klas.

5.2.2 Metoda wartości średnich

W metodzie tej, liczenie stałej normalizacyjnej jest zbędne, wykorzystywany jest tutaj ten fakt, że rozkład kolejki, obserwowany przez nadchodzące zgłoszenie, jest stacjonarnym rozkładem tej kolejki po usunięciu z niej ukazanego zgłoszenia:

$$p_i\left(\mathbf{n_i},\ \mathbf{N}\ \middle|\ \begin{array}{c}\text{tuż przed przybyciem}\\ \text{zgłoszenia klasy } k\end{array}\right) = p_i\left[\mathbf{n_i},\ \mathbf{N}(k^{-1})\right] \qquad (5.24)$$

gdzie:

$$\mathbf{N}(k^-) = (N^1,\ \cdots,\ N^k - 1,\ \cdots,\ N^K)$$

a więc średni czas przejścia t_i^k zgłoszenia przez stanowisko i (przy niezależnym od liczby zgłoszeń czasie obsługi), dla stanowiska typu infinite server (w przypadku, gdy liczba kanałów obsługi jest większa od $\mathbf{N}$) wynosi:

$$t_i^k(\mathbf{N}) = s_i^k$$

oraz, dla pozostałych typów stanowisk:

$$t_i^k(\mathbf{N}) = s_i^k\ \left[1 + Q_i(\mathbf{N}(k^{-1}))\right] \qquad (5.25)$$

gdzie, Q_i - średnia liczba zgłoszeń na stanowisku i (wcześniej używaliśmy oznaczenia $\overline{n_i}$).

W metodzie tej zostało wykorzystane twierdzenie Little'a wiążące średnią liczbe zgłoszeń, intensywność napływu zgłoszeń oraz średni czas przejścia zgłoszenia przez sieć ($Q = \lambda\ t$) i tak dla całej sieci mamy:

$$\lambda_i^k(\mathbf{N}) = \frac{N^k}{T_i^k(\mathbf{N})} \qquad (5.26)$$

a dla stanowiska i:

$$Q_i^k(\mathbf{N}) = \lambda_i^k(\mathbf{N})\ t_i^k(\mathbf{N}) \qquad (5.27)$$

gdzie, T_i^k jest średnim czasem obiegu sieci, poczynając od stanowiska i, dla zgłoszeń klasy k.

Ponieważ między pobytami na stanowisku i zgłoszenie przechodzi przez stanowisko j średnio:

$$\frac{e_j^k}{e_i^k}\ \text{razy}$$

gdzie, np. e_i^k - średnia liczba pobytów zgłoszenia klasy k na stanowisku i, to:

$$T_i^k(\mathbf{N}) = \sum_{j=1}^{M} t_j^k(\mathbf{N})\ \frac{e_j^k}{e_i^k} \qquad (5.28)$$

Zakładając, że $t_i^k(\mathbf{0}) = 0$ i zaczynając od populacji $\mathbf{0}$, a kończąc na $\mathbf{N}$ iteracyjnie oblicza się ze wzorów (5.25) (5.26) (5.27) $Q_i^k(\mathbf{N})$ i $t_i^k(\mathbf{N})$, potem pozostałe miary wydajności (parametry) sieci.

Znane są i inne odmiany algorytmu wartości średnich, wśród nich i łączące metodę wartości średnich z metodą splotową, oraz rozszerzające metodę wartości średnich na sieć otwartą i na sieci mieszane, czy na czasy obsługi zależne od obciążenia stanowiska, itp.

Metody splotowa i metoda wartości średnich, są dokładne tylko dla sieci separowalnych, a z kolei z nich powstało szereg metod przybliżonych, np. dla różnych czasów obsługi dla różnych klas zgłoszeń, dla rozkładów dowolnych, dla obsługi priorytetowej z różnymi klasami zgłoszeń, dla sieci z blokowaniem stanowisk, itp.

5.3 Niemarkowowskie modele sieci stanowisk

Omawianie tej grupy modeli należy zacząć od znalezienia przybliżonych rozwiązań, przez rozszerzenie rozwiązania iloczynowego na przypadki, gdzie ono nie obowiązuje w sposób dokładny.

5.3.1 Model rozszerzonej formy iloczynowej

Rozpatrzmy sieć stanowisk obsługi, z regulaminem naturalnym FIFO i dowolnym rozkładem czasu obsługi.

Do sieci zamkniętej z jedną klasą zgłoszeń, rozwiązanie typu $p(\mathbf{n})$ jest w przybliżeniu iloczynem rozwiązań $p_i(n_i)$ dla niezależnych, pojedynczych, stanowisk obsługi typu M/G/1/N, gdzie $a(t)$ ma rozkład wykładniczy, a $b(t)$ - rozkład dowolny, a liczba zgłoszeń na stanowisku (węźle) ograniczona jest do N. Jest to model rozszerzonej formy iloczynowej.

Za podstawę przyjmuje się tutaj wyniki otrzymane dla systemu typu M/G/1 z nieograniczoną kolejką, a dla systemu typu M/G/1/N postać $p(n)$ można określić na podstawie jej związku z rozkładem $p'(n)$ liczby zgłoszeń w systemie M/G/1, więc:

$$p(n) = \frac{\frac{p'(n)}{(1-\rho)}}{1 + \frac{\rho}{(1-\rho)} \sum_{v=0}^{N-1} p'(v)} \quad \text{dla } n = 0, \cdots, N-1 \tag{5.29}$$

oraz

$$p(N) = 1 - \sum_{n=0}^{N-1} p(n) \tag{5.30}$$

Wielkość $p'(n)$ znana jest w jawnej formie tylko dla niektórych rozkładów $b(t)$; dla innych może być obliczona ze znanej zależności Pollaczka - Chinczyna dla funkcji tworzącej $p'(n)$:

$$Q(z) = B^*(\lambda - \lambda\, z)\frac{(1-\rho)(1-z)}{B^*(\lambda - \lambda\, z) - z} \tag{5.31}$$

gdzie:

$$Q(z) = \sum_{k=0}^{\infty} p'(k)\, z^k \tag{5.32}$$

i jest to funkcja tworząca $p'(n)$, a B^* jest transformatą Laplace'a funkcji gęstości rozkładu czasu obsługi.

Udaje się również wprowadzić do tego rozwiązania stanowiska typu M/G/c/N.

5.3.2 Iteracyjna metoda równoważnego stanowiska

Tę metodę stosuje się z reguły do sieci zamkniętych, a proces obliczenia $p_i(n_i)$ opiera się na redukcji sieci do dwóch wyróżnianych stanowisk: wyróżnionego stanowiska i i stanowiska zastępczego, które reprezentuje resztę sieci.

Stanowisko zastępcze charakteryzuje się wykładniczym rozkładem $b_{z(i)}(t)$, którego wartość średnia $s_{z(i)}(n)$ obliczana jest na podstawie przepustowości sieci z n zgłoszeniami plus "zwarte" stanowisko i (gdy $s_i = 0$). Redukcja sieci, poza stanowiskiem i do jednego stanowiska (lub źródła zgłoszeń dla sieci otwartej) nazywana jest twierdzeniem Nortona, przez analogię z obwodami elektrycznymi.

Rozwiązanie takiej sieci, złożonej z wykładniczego stanowiska zastępczego i stanowiska z dowolnym rozkładem czasu obsługi, dostarcza wartości $p_i(n_i)$, dla $i = 1, \cdots, M$.

Ogólniejsza wersja tej metody dotyczy wielu klas zgłoszeń lub dowolnego fragmentu sieci. Redukcja ta jest dokładna w przypadku sieci dekomponowalnych i dlatego niektóre algorytmy wykorzystują zastępczą sieć dekomponowalną o takiej samej topologii, jaką ma sieć badana i o średnich czasach obsługi s_i^* korygowanych iteracyjnie, po wyznaczeniu $p_i(n_i)$, $i = 1, \cdots, M$, tak aby wyznaczone na podstawie $p(\mathbf{n})$ przepustowości stanowisk spełniały równania ruchu, a łączna średnia długość kolejki zgadzała się z rzeczywistą, czyli $\sum_{i=1}^{M} Q_i \approx N$.

Inne zaś algorytmy, w celu zwiększenia dokładności metody, obliczają po wyznaczeniu strumienia wyjściowego $\lambda_i(N-n_i)$ ze stanowiska zastępczego, $p_i(n_i)$ jako rozkład w izolowanym stanowisku z Cox'owskim rozkładem czasu obsługi i ograniczoną do N kolejką, a wtedy:

$$s_i^*(n_i) = \frac{p_i(n_i)}{\lambda_i(n_i - 1)\; p_i(n_i - 1)} \tag{5.33}$$

gdzie, $(n_i - 1)$ oznacza liczbę zgłoszeń.

W takim algorytmie, jeżeli testy poprawności rozwiązania nie są spełnione, inaczej dobiera się czasy obsługi dla sieci pomocniczej i rozwiązaniem nie jest iloczyn $p_i(n_i)$, lecz rozwiązanie dekomponowalnej sieci pomocniczej. Poszukuje się w ten sposób równoważnej sieci dekomponowalnej o tych samych rozkładach kolejek, co sieć badana.

Można też tutaj wprowadzić obsługę priorytetową lub zwiększać dokładność metody przez ocenę drugiego momentu rozkładu czasu między zgłoszeniami w strumieniu wyjściowym stanowiska zastępczego, tj. w strumieniu wejściowym do stanowiska i.

5.3.3 Metody izolacji stanowisk

Metody te zajmują się problemem własności strumieni w sieci, wyznaczając w sposób przybliżony wartości drugiego momentu rozkładu $a_i(t)$ dla każdego stanowiska. Dla uproszczenia zakłada się, że strumienie wyjściowe z poszczególnych stanowisk są niezależne i są procesami odnowy, tzn. odstępy czasu pomiędzy kolejnymi zgłoszeniami opuszczającymi stanowisko obsługi i są niezależne i mają ten sam rozkład $d_i(t)$.

Parametry tego rozkładu można oszacować korzystając z dokładnej zależności dla stanowiska typu M/G/1, a przybliżonej dla stanowiska z ogólnym strumieniem wejściowym typu G:

$$d_i(t) = \rho_i\ b_i(t) + (1-\rho_i)\ a_i(t) * b_i(t) \tag{5.34}$$

Przypomnijmy w tym miejscu, że splot funkcji $f(t)$ i funkcji $g(t)$ równy jest:

$$f(t) * g(t) = \int_0^t f(\tau)\ g(t-\tau)\ d\tau \tag{5.35}$$

i transformata Laplace'a splotu:

$$\mathcal{L}\left\{f(t) * g(t)\right\} = F(s)\ G(s) \tag{5.36}$$

Teraz ze wzoru (5.34) można obliczyć $E\left[d_i^2\right]$, a następnie współczynniki zmienności rozkładu $d_i(t)$:

$$C_{di} = \frac{E\left[d_i^2\right] - E\left[d_i\right]^2}{E\left[d_i\right]^2} = C_{ai}\ (1-\rho_i) + \rho_i^2\ C_{bi} + \rho_i\ (1-\rho_i) \tag{5.37}$$

Wartość średnia $E\left[d_i\right]$, to oczywiście $\frac{1}{\lambda_i}$, a równania ruchu są te same dla dowolnych rozkładów $b_i(t)$.

Strumień wyjściowy rozkłada się za każdym stanowiskiem, zgodnie z prawdopodobieństwami przejść r_{ij}, a strumień pomiędzy stanowiskami i oraz j charakteryzuje następujący rozkład:

$$d_{ij}(t) = r_{ij}\ d_i(t) + r_{ij}\ (1-r_{ij})\ d_i(t) * d_i(t) + \\ + r_{ij}\ (1-r_{ij})^2\ d_i(t) * d_i(t) * d_i(t) + \cdots \tag{5.38}$$

stąd

$$\lambda_{ij} = r_{ij}\ \lambda_i, \qquad C_{dij} = (C_{di} - 1)\ r_{ij} \tag{5.39}$$

Strumień wejściowy do stanowiska j jest superpozycją niezależnych procesów $d_{ij}(t)$ i jako suma procesów odnowy ma współczynnik zmienności równy:

$$C_{aj} = \frac{1}{\lambda_j}\sum_{i=0}^{M} r_{ij}\ \left[(C_{di} - 1)\ r_{ij} + 1\right] \tag{5.40}$$

a wyrażenia z indeksem $i = 0$ odnoszą się do strumienia zewnętrznego dla sieci otwartej.

Wyrażenia (5.37) i (5.40) tworzą układ równań liniowych, z których można wyznaczyć C_{ai}, dla $i = 1, \cdots, M$. Wartości λ_i określają równania ruchu, a znając powyższe parametry strumienia wejściowego, można każde ze stanowisk analizować oddzielnie, jako stanowisko typu GI/G/1 z ogólnym strumieniem wejściowym i ogólnym rozkładem czasów obsługi dla $a_i(t)$ i $b_i(t)$, znanych z dokładnością do dwóch pierwszych momentów.

5.3.4 Aproksymacja dyfuzyjna

Powróćmy jeszcze raz do systemu obsługi typu G/G/1. Moment przybycia do systemu zgłoszenia z numerem k składa się z sumy k odstępów między przybywaniem kolejnych zgłoszeń, tj.:

$$\tau_1^{(1)} + \tau_1^{(2)} + \tau_1^{(3)} + \cdots + \tau_1^{(k)} \tag{5.41}$$

przy czym uważa się, że $\tau_1^{(0)}$ jest równy zero.

Dla systemu G/G/1 zakłada się, że ciąg $\{\tau_1\}$ jest ciągiem niezależnych, o jednakowym rozkładzie, zmiennych losowych z dystrybuantą $A(t)$.

Jeżeli czas t i oczywiście liczba k przyjmą wartość dostatecznie dużą, to moment przybycia zgłoszenia k jest sumą dużej liczby niezależnych o jednakowym rozkładzie zmiennych losowych. Można oczekiwać, że tutaj da się zastosować centralne graniczne twierdzenie, pozwalające opisać zmienną losową - moment przybycia zgłoszenia k, więc i stochastyczny proces wejściowy, jako podlegający rozkładowi normalnemu. To założenie o normalnym rozkładzie procesu wejściowego, a zatem i procesu wyjściowego, jest podstawą aproksymacji dyfuzyjnej.

Przy aproksymacji dyfuzyjnej zakłada się, że proces przybywania i wyjścia (obsługi) zgłoszeń aproksymuje się ciągłymi probabilistycznymi procesami o przyrostach niezależnych, które w momencie t podlegają rozkładowi normalnemu, ze znaną wartością średnią i wariancją.

Znając te cztery parametry (dwie wartości średnie i dwie wariancje) otrzymujemy pełne charakterystyki tych dwóch probabilistycznych procesów, albowiem proces o rozkładzie normalnym, o przyrostach niezależnych, jest procesem dwuparametrowym. Wiadomo też, że wariancje wprowadza się do opisu losowych fluktuacji procesów, w stosunku do ich wartości średnich.

Z praktycznego punktu widzenia aproksymacja dyfuzyjna przybliża więc wartość $p(n)$, zastępując liczbę zgłoszeń n w stanowisku obsługi wartością x procesu dyfuzji, bo zmiany n spowodowane traktowanymi niezależnie procesami: wejściowym i obsługi (wyjściowym) mają, po dostatecznie długim czasie t, rozkład normalny o średniej:

$$(\lambda - \mu)\ t = \beta\ t \tag{5.42}$$

i wariancji:

$$(\lambda\ C_a + \mu\ C_b) = \alpha\ t \tag{5.43}$$

niezależnie od postaci $a(t)$ i $b(t)$. Jest to więc proces dyfuzji, którego zmiany mają rozkład normalny o wartości średniej $\beta\ dt$ i wariancji $\alpha\ dt$.

Funkcja gęstości prawdopodobieństwa takiego procesu spełnia w stanie ustalonym następujące równanie:

$$0 = \frac{\alpha}{2}\ \frac{d^2 f(x)}{dx^2} - \beta\ \frac{df(x)}{dx} \tag{5.44}$$

którego rozwiązanie, przy warunku brzegowym ograniczającym proces do $x \geq 0$ przybliża $p(n)$:

$$f(n) \approx p(n)$$

Odpowiednio dobierając współczynniki β i α można uwzględnić:

a) klasy zgłoszeń, w ramach regulaminu naturalnego FIFO:

$$\beta = (\lambda - \mu) \tag{5.45}$$

$$\alpha = \sum_{k=1}^{K} \lambda^k\ C_a^k + \mu^3 \left[\sum_{k=1}^{K} \frac{a^k\ (C_b^k + 1)}{(\mu^k)^2} - \frac{1}{\mu^2} \right] \tag{5.46}$$

gdzie, k - oznacza klasę zgłoszeń, oraz:

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

a obliczenia strumieni poszczególnych klas, przebiegają analogicznie do przedstawionego,

b) c - kanałowe stanowisko obsługi:

$$\left.\begin{aligned} \beta(x) &= \lambda - l\ \mu \\ \alpha(x) &= \lambda\ C_a + l\ \mu\ C_b \end{aligned}\right\} \text{ dla } \ l < x \leq l+1, \qquad l = 0, \cdots, c-1$$

$$\left.\begin{aligned} \beta(x) &= \lambda - c\ \mu \\ \alpha(x) &= \lambda\ C_a + c\ \mu\ C_b \end{aligned}\right\} \text{ dla } \ x \geq c \tag{5.47}$$

c) zgłoszenia napływające grupami:

$$\beta = \lambda\ v_g - \mu \tag{5.48}$$

$$\alpha = \lambda\ v_g\,[C_a + Cg] + \mu\ C_b \tag{5.49}$$

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

d) regulamin obsługi z priorytetem absolutnym (z wywłaszczeniem):

$$\beta = \sum_{k=1}^{K} \lambda^k - \sum_{k=1}^{K} \frac{\rho^k}{R^k}\ \mu^k \tag{5.50}$$

$$\alpha = \sum_{k=1}^{K} \lambda^k\ C_a^k + \sum_{k=1}^{K} \frac{\rho^k}{R^k}\ \mu^k\ C_b^k \tag{5.51}$$

gdzie, k - indeks priorytetu oraz:

$$R^k = \sum_{i=1}^{k} \rho^i$$

Wracając do wyrażenia (5.44) rozwiązanie $f^K(n)$ określa rozkład zgłoszeń wszystkich klas łącznie, a z zależności:

$$f^k(n) = \sum_{v=0}^{n} f^{k-1}(n-v)\, p^k(v), \quad k = 2, \cdots, K, \quad f^1(n) = p^1(n) \quad (5.52)$$

można wyznaczyć rozkład $p^k(n)$ dla poszczególnych klas.

Reasumując rozważania z tego paragrafu można np. w metodzie izolacji stanowisk, zamiast wyrażenia (5.37), dla współczynnika zmienności rozkładu, posłużyć się dokładną zależnością, wiążącą C_{di} ze średnim czasem czekania $\overline{w_i}$ na stanowisku GI/G/1:

$$C_{di} = C_{ai} + 2\,\rho_i^2\, C_{bi} - 2\,\rho_i\,(1-\rho_i)\,\frac{\overline{w_i}}{s_i} \qquad (5.53)$$

gdzie, $\overline{w_i}$ jest szacowane jedną z zaprezentowanych metod przybliżonych; jeżeli będzie to np. aproksymacja dyfuzyjna, to:

$$\overline{w_i} \approx s_i\,\frac{\rho_i}{2\,(1-\rho_i)}\,(C_{ai} + C_{bi}) \qquad (5.54)$$

i wtedy otrzymamy:

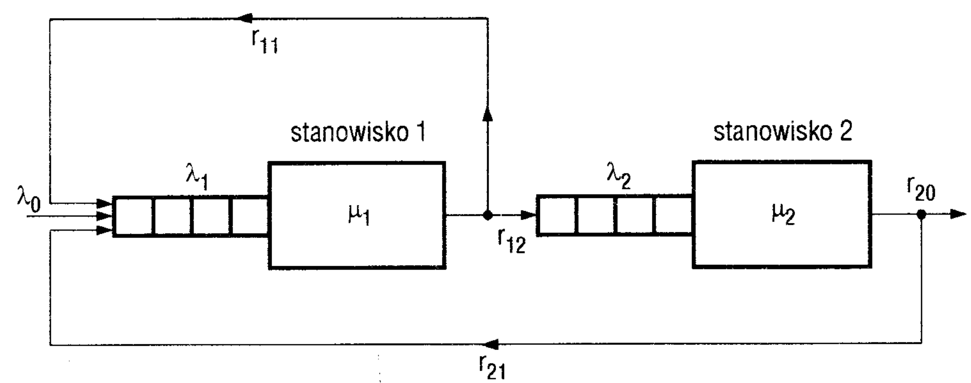
$$C_{di} \approx C_{ai} + \rho_i^2 (C_{ai} - C_{bi}) \qquad (5.55)$$

Należy tutaj zaznaczyć, że w rozwiązaniach przybliżonych pominięto wpływ trzecich momentów rozkładów $a_i(t)$ i $b_i(t)$, mających znaczny wpływ na sposób zachowania się stanowisk obsługi.

5.4 Zadania

Modele z teorii kolejek mogą być łatwo zaadaptowane jako statyczne modele różnorodnych systamów komputerowych. Pozwala to, np. w trakcie projektowania lub kompletowania takich systemów, na sprawdzenie doboru wszystkich komponentów systemu, np. pod względem przepustowości, a także tego czy nie ma urządzeń działających zbyt wolno, przez to nadmiernie przeciążonych pracą i hamujących cały proces obliczeniowy.

Zadanie 1. Dany jest system komputerowy, składający się z jednostki centralnej i zewnętrznej pamięci dyskowej:



Parametry wejściowe tego systemu, liczone w liczbie zgłoszeń na jednostkę czasu wynoszą:

$$\lambda_0 = 0.4, \qquad \mu_1 = 4.0, \qquad \mu_2 = 1.0$$

prawdopodobieństwa zaś przejść między stanowiskami równe są:

$$r_{11} = 0.4, \qquad r_{12} = 0.6, \qquad r_{20} = 0.5, \qquad r_{21} = 0.5$$

Traktując jednostkę centralną i dyski jako systemy masowej obsługi typu M/M/1 należy obliczyć średnią liczbę zgłoszeń na każdym z tych stanowisk, średnie liczby zgłoszeń w każdej z kolejek oraz średnią liczbę zgłoszeń na każdym stanowisku, plus parametry czasowe.

Wychodząc z tego, że zgłoszenia wielokrotnie przechodzą przez każde stanowisko, należy obliczyć łączne strumienie zgłoszeń do każdego ze stanowisk obsługi:

$$\begin{cases} \lambda_1 = \lambda_0 + \lambda_1 \cdot r_{11} + \lambda_2 \cdot r_{21} \\ \lambda_2 = \lambda_1 \cdot r_{12} \end{cases}$$

stąd

$$\lambda_1 = \lambda_0 + \lambda_1 \cdot r_{11} + \lambda_1 \cdot r_{12} \cdot r_{21}$$

$$\lambda_1 = \frac{\lambda_0}{1 - r_{11} - r_{12} \cdot r_{21}} = \frac{0.4}{1 - 0.4 - 0.6 \cdot 0.5} = \frac{0.4}{0.3} = 1\frac{1}{3}$$

$$\lambda_2 = \lambda_1 \cdot r_{12} = \frac{4}{3} \cdot 0.6 = \frac{4}{5} = 0.8$$

to z kolei pozwala na obliczenie obciążeń stanowisk obsługi:

$$\rho_1 = \frac{\lambda_1}{\mu_1} = \frac{4}{3 \cdot 4} = \frac{1}{3}$$

$$\rho_2 = \frac{\lambda_2}{\mu_2} = \frac{0.8}{1} = 0.8$$

oraz średniej liczby zgłoszeń na stanowisku pierwszym i drugim:

$$\overline{n_1} = \frac{\rho_1}{1 - \rho_1} = \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{1}{2} = 0.5$$

$$\overline{n_2} = \frac{\rho_2}{1 - \rho_2} = \frac{0.8}{1 - 0.8} = 4$$

Średnia liczba zgłoszeń w kolejkach:

$$\overline{v_1} = \frac{\rho_1^2}{1 - \rho_1} = \frac{\left(\frac{1}{3}\right)^2}{1 - \frac{1}{3}} = \frac{1}{6}$$

$$\overline{v_1} = \frac{\rho_2^2}{1 - \rho_2} = \frac{(0.8)^2}{1 - 0.8} = 3.2$$

Średnia liczba zgłoszeń na stanowiskach obsługi równa jest współczynnikowi obciążenia systemu, więc:

$$\overline{l_1} = \rho_1 = \frac{1}{3}$$

$$\overline{l_2} = \rho_2 = 0.8$$

Czasy oczekiwania w kolejkach i czasy pobytu zgłoszeń na stanowiskach obsługi, równe są:

$$\overline{w_1} = \frac{\rho_1}{\mu_1\ (1 - \rho_1)} = \frac{\frac{1}{3}}{4\ \left(1 - \frac{1}{3}\right)} = \frac{1}{8}$$

$$\overline{w_2} = \frac{\rho_2}{\mu_2\ (1 - \rho_2)} = \frac{0.8}{1 \cdot (1 - 0.8)} = 4$$

$$\overline{q_1} = \frac{1}{\mu_1 (1 - \rho_1)} = \frac{1}{4 \cdot \left(1 - \frac{1}{3}\right)} = \frac{3}{8}$$

lub

$$\overline{q_1} = \overline{w_1} + \frac{1}{\mu_1} = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}.$$

$$\overline{q_2} = \frac{1}{\mu_2 (1 - \rho_2)} = \frac{1}{1 \cdot (1 - 0.8)} = 5$$

lub

$$\overline{q_2} = \overline{w_1} + \frac{1}{\mu_2} = 4 + \frac{1}{1} = 5$$

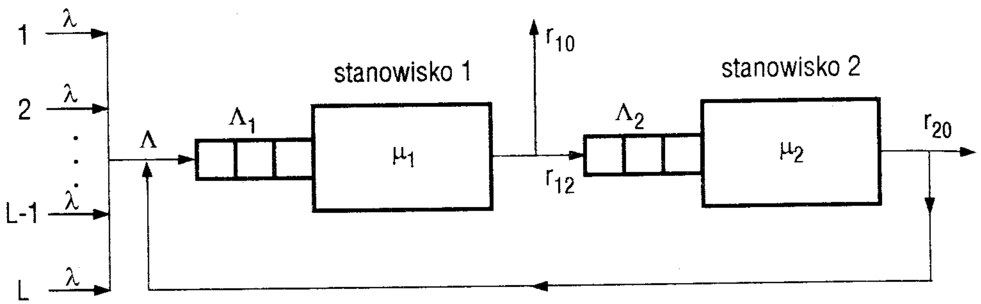
Zadanie 2. Wielodostępny system komputerowy można przedstawić tak jak na zadanym rysunku.

Każdy z L terminali jest źródłem generującym poissonowski strumień zgłoszeń, czas zaś obsługi zgłoszeń na stanowiskach obsługi podlega rozkładowi wykładniczemu. Obliczyć:

a) średnią liczbę zgłoszeń na każdym ze stanowisk (w kolejce plus na obsłudze) oraz średnią liczbę zgłoszeń w całym systemie,

b) średni czas przejścia zgłoszenia przez każde ze stanowisk (niejednokrotny pobyt na stanowisku),

c) średni czas pobytu zgłoszenia w całym systemie.



Dane: $L = 20$, $\lambda = 0.02\ s^{-1}$, $\mu_1 = 1.0\ s^{-1}$, $\mu_2 = 0.5\ s^{-1}$, $r_{10} = 0.4$, $r_{20} = 0.3$

Rozwiązanie: Zaczynamy od obliczenia intensywności strumieni wpływającą do stanowisk z numerem 1 i 2:

$$\begin{aligned}\Lambda &= L \cdot \lambda \\ \Lambda_1 &= \Lambda + \Lambda_2\,(1 - r_{20}) \\ \Lambda_2 &= \Lambda_1\,(1 - r_{10})\end{aligned}$$

czyli

$$\Lambda_1 = \Lambda + \Lambda_1\,(1 - r_{10})(1 - r_{20})$$

$$\Lambda_1 = \frac{\Lambda}{1 - (1 - r_{10})(1 - r_{20})} = \frac{20 \cdot 0.02}{1 - 0.6 \cdot 0.7} = \frac{20}{29}$$

więc, intensywność strumieni, równa jest:

$$\Lambda = 20{\cdot}0.02 = 0.4\ s^{-1}, \qquad \Lambda_1 = \frac{20}{29}\ s^{-1}, \qquad \Lambda_2 = \frac{20}{29}{\cdot}(1{-}0.4) = \frac{12}{29}\ s^{-1}$$

Obciążenia stanowisk obsługi:

$$\rho_1 = \frac{\Lambda_1}{\mu_1} = \frac{20}{1 \cdot 29} = \frac{20}{29} \approx 0.62$$

$$\rho_2 = \frac{\Lambda_2}{\mu_2} = \frac{12}{0.5 \cdot 29} = \frac{24}{29} \approx 0.83$$

Średnia liczba zgłoszeń w każdym ze stanowisk (w kolejce i na obsłudze):

$$\overline{n_1} = \frac{\rho_1}{1 - \rho_1} = \frac{\frac{20}{29}}{1 - \frac{20}{29}} = \frac{20}{9} = 2\frac{2}{9}$$

$$\overline{n_2} = \frac{\rho_2}{1 - \rho_2} = \frac{\frac{24}{29}}{1 - \frac{24}{29}} = \frac{24}{5} = 4.8$$

Średnia liczba zgłoszeń w systemie:

$$\overline{n} = \overline{n_1} + \overline{n_2} = 2\frac{2}{9} + 4.8 = \frac{316}{45} = 7\frac{1}{45} \approx 7.02$$

Średni czas pobytu zgłoszenia na stanowisku (wzory Little'a):

$$\overline{q_{p_1}} = \frac{1}{\Lambda} \cdot \overline{n_1} = \frac{1}{0.4} \cdot \frac{20}{9} = \frac{50}{9} = 5\frac{5}{9}\ s$$

$$\overline{q_{p_2}} = \frac{1}{\Lambda} \cdot \overline{n_2} = \frac{1}{0.4} \cdot \frac{24}{5} = 12\ s$$

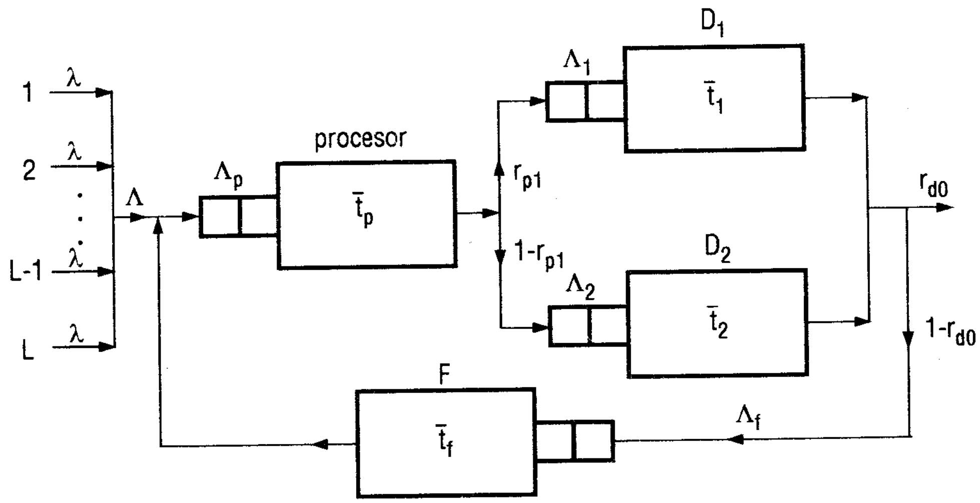
Średni czas reakcji całego systemu:

$$\overline{q} = \frac{1}{\Lambda} \cdot \overline{n} = \frac{1}{0.4} \cdot \frac{316}{45} = \frac{158}{9} = 17\frac{5}{9}\ s$$

Zadanie 3. Załóżmy, że model sieci komputerowej składa się z następujących elementów (patrz rysunek).

Zgłoszenia napływają od L niezależnych źródeł (terminali), z identycznymi charakterystykami (strumienie Poissona). Każde zgłoszenie jest obsługiwane przez procesor, następnie przez jedno z urządzeń wejścia-wyjścia D_1 lub D_2 z prawdopodobieństwami r_{p1} i $1 - r_{p1}$. Na tym obsługa może być zakończona i zgłoszenie opuszcza system z prawdopodobieństwem r_{d0} lub obsługiwane jest przez file server F i powtarza cykl obsługi.

Prawdopodobieństwa r_{p1} i r_{d0} są zadane a czasy obsługi są wykładnicze z wartościami: $\overline{t_p}$, $\overline{t_1}$, $\overline{t_2}$, $\overline{t_f}$, więc można tutaj, dla obliczenia parametrów takiej sieci, zastosować model Jacksona.



Dane: $L = 40$, $\lambda = 0.01$ s^{-1}, $\overline{t_p} = 0.8$ s, $\overline{t_1} = 0.3$ s, $\overline{t_2} = 0.6$ s, $\overline{t_f} = 1.0$ s, $r_{p1} = 0.6$, $r_{d0} = 0.4$.

Rozwiązanie: Obliczanie intensywności strumieni wpływających do stanowisk obsługi:

$$\begin{aligned}
\Lambda &= L \cdot \lambda \\
\Lambda_p &= \Lambda + \Lambda_f \\
\Lambda_1 &= r_{p1} \cdot \Lambda_p \\
\Lambda_2 &= (1 - r_{p1}) \cdot \Lambda_p \\
\Lambda_f &= (\Lambda_1 + \Lambda_2) \cdot (1 - r_{d0}) = \Lambda_p \cdot (1 - r_{d0})
\end{aligned}$$

i po odpowiednich podstawieniach otrzymamy:

$$\Lambda_p = \Lambda + \Lambda_p - \Lambda_p \cdot r_{d0}$$

stąd

$$\Lambda = L \cdot \lambda = 0.4$$
$$\Lambda_p = \frac{\Lambda}{r_{d0}} = 1.0$$
$$\Lambda_1 = \frac{r_{p1} \cdot \Lambda}{r_{d0}} = 0.6$$
$$\Lambda_2 = \frac{(1 - r_{p1}) \cdot \Lambda}{r_{d0}} = 0.4$$
$$\Lambda_f = \frac{(1 - r_{d0}) \cdot \Lambda}{r_{d0}} = 0.6$$

Następnym krokiem powinno być obliczenie obciążeń stanowisk obsługi:

$$\begin{aligned}
\rho_p &= \Lambda_p \cdot \overline{t_p} = 0.8 \\
\rho_1 &= \Lambda_1 \cdot \overline{t_1} = 0.18 \\
\rho_2 &= \Lambda_2 \cdot \overline{t_2} = 0.24 \\
\rho_f &= \Lambda_f \cdot \overline{t_f} = 0.6
\end{aligned}$$

Traktując każde ze stanowisk jako niezależny system typu M/M/1, możemy teraz obliczyć średnią liczbę zgłoszeń, znajdujących się w kolejce i na stanowisku obsługi:

$$\overline{n_i} = \frac{\rho_i}{1 - \rho_i}$$

więc:

$$\begin{aligned}
\overline{n_p} &= 4.00 & \overline{n_1} &= 0.22 \\
\overline{n_2} &= 0.32 & \overline{n_f} &= 1.50
\end{aligned}$$

oraz średnią liczbę zgłoszeń w całym systemie:

$$\overline{n} = \overline{n_p} + \overline{n_1} + \overline{n_2} + \overline{n_f} = 6.04$$

Inną ważna charakterystyką sieci stanowisk obsługi jest średnia liczba cykli e na każdym stanowisku obsługi, które musi przejść zgłoszenia zanim obsłużone opuści system:

$$e_p = \frac{\Lambda_p}{\Lambda} = 2.5 \qquad e_1 = \frac{\Lambda_1}{\Lambda} = 1.5$$

$$e_2 = \frac{\Lambda_2}{\Lambda} = 1.0 \qquad e_f = \frac{\Lambda_p}{\Lambda} = 1.5$$

parametry te mogą być użyte do obliczenia łącznego czasu pobytu zgłoszenia na stanowisku obsługi, albowiem, np. dla procesora parametr:

$$\overline{q'_p} = \frac{\overline{n_p}}{\Lambda_p} = \frac{4.0}{1.0} = 4.0 \text{ s}$$

charakteryzuje średni czas jednokrotnego pobytu zgłoszenia, a łączny czas pobytu równy jest:

$$\overline{q_p} = \overline{q'_p} \cdot e_p = 4.0 \cdot 2.5 = 10 \text{ s}$$

a to pozwala obliczyć średni czas pobytu złoszenia w całym systemie.

Dodatkowo, opierając się na już obliczonych parametrach, można wyliczyć maksymalną liczbę terminali, które mogą spowodować pełne nasycenie (zablokowanie) systemu. Zjawisko to wystąpi, gdy obciążenie przynajmniej jednego ze stanowisk, będzie równe 1.

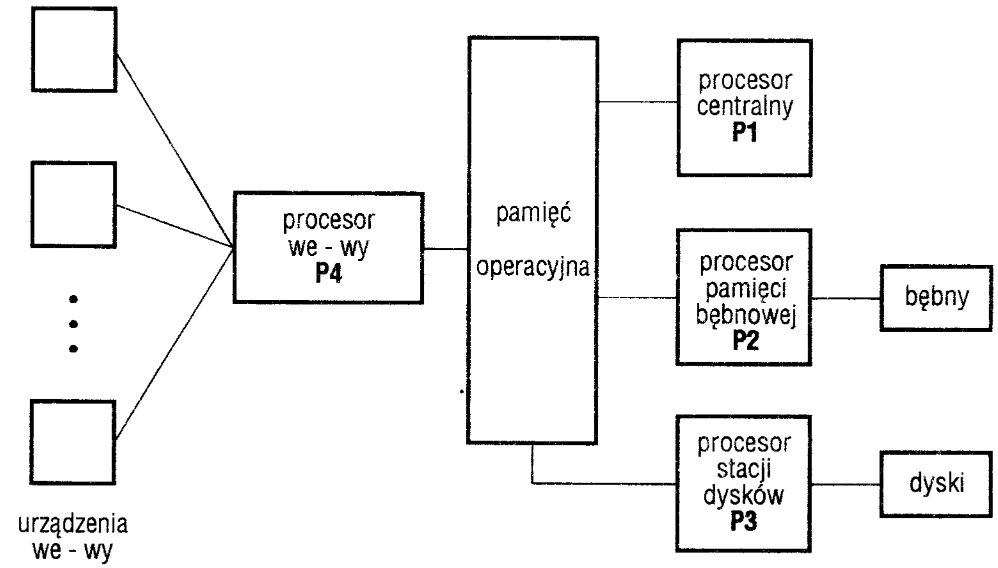
Najbardziej obciążonym stanowiskiem w zadanym systemie jest procesor z $\rho_p = 0.8$, oraz:

$$\rho_p = \overline{t_p} \cdot \Lambda_p = \overline{t_p} \cdot \frac{L \cdot \lambda}{r_{d0}} = 0.02 \cdot L$$

a $L_{\max}$, otrzymamy, gdy $\rho_p = 1,0$, więc:

$$L_{\max} = \frac{1.0}{0.02} = 50$$

Zadanie 4. Analiza sieci otwartej metodą Jacksona. Zadany jest system komputerowy o konfiguracji pokazanej na rysunku:

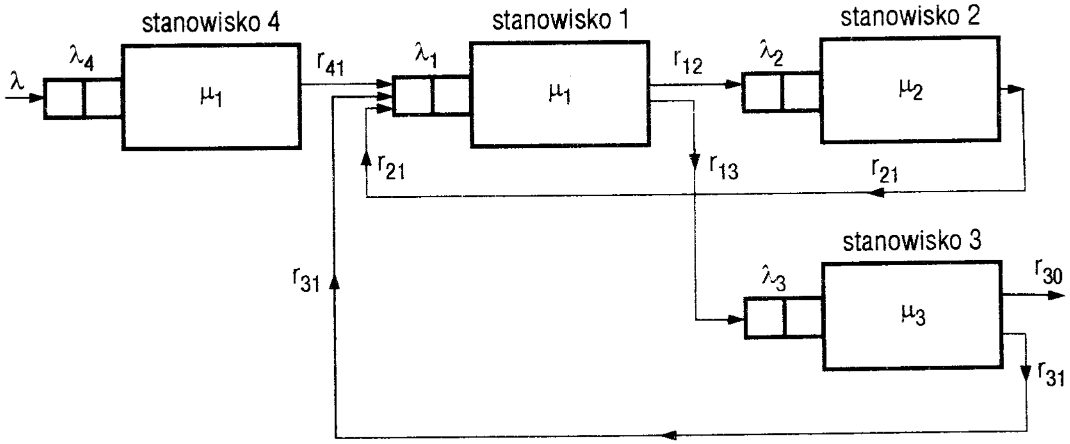


Tworząc model systemu każdy z czterech procesorów P_1, P_2, P_3, P_4 tego systemu przedstawimy jako jednokanałowe stanowisko obsługi, z nieograniczoną kolejką i z regulaminem szeregowania zadań typu FIFO. Zakładamy również, że czas obsługi na stanowiskach obsługi jest niezależny i ma rozkład wykładniczy. Intensywność obsługi, np. na stanowisku i oznaczymy jako μ_i.

Użytkownicy systemu przedstawieni są jako nieskończenie wymiarowe źródło, generujące zadania, które tworzą strumień Poissona z intensywnością λ.

Wiadomo jest, że metoda Jacksona pozwala również na analizę sieci, w których parametr λ zależny jest od liczby zgłoszeń N, oraz na analizę sieci, w których intensywność obsługi na stanowiskach obsługi zależy od liczby zgłoszeń na tych stanowiskach. W tym przykładzie, upraszczając model, zakładamy, że $\lambda = const$ oraz że $\mu_i = const$, chociaż wiadomo, że w bardzo obciążonych systemach parametry te zależą od N.

Model systemu oraz powiązania między stanowiskami obsługi, pokazujące drogi poruszania się zadań wewnątrz systemu, przedstawione są na rysunku poniżej:



Jak to widać na rysunku, zadania wchodzą do systemu przez procesor wejścia-wyjścia (stanowisko 4) i zadanie, które zostało obsłużone na tym stanowisku obsługi, uważane jest za załadowane do pamięci operacyjnej i gotowe jest do dalszego przetwarzania, gdy przyjdzie jego kolejka, w centralnym procesorze (stanowisko 1). Następnie zadanie opuszcza centralny procesor (jego miejsce zajmuje kolejne zadanie z pamięci operacyjnej) i przechodzi, z prawdopodobieństwem r_{12} do stanowiska 2 (model procesora bębna) lub z prawdopodobiństwem $r_{13} = 1 - r_{12}$ do stanowiska 3 (model procesora dysków).

Kiedy stanowisko odpowiadające pamięci bębnowej lub stacji dysków kończy obsługę zadania, to może ono wyjść z sieci lub wrócić do centralnego procesora dla dalszej obróbki. Prawdopodobieństwo powrotu z pamięci bębnowej zastało tutaj oznaczone jako r_{21}, a ze stacji dysków jako r_{31}. Wszystkie te prawdopodobieństwa są stałe i jednakowe dla wszystkich zadań.

Należy w tym miejscu zaznaczyć, że zadania opuszczające sieć powinny przejść przez procesor wejścia-wyjścia, ale operacja ta nie została uwzględniona w tym modelu.

Ważnym zasobem systemu komputerowego jest również pamięć operacyjna. Trudno ją jednak przedstawić jako model kolejkowy, gdyż jest ona rozdzielana między zadaniami i często wykorzystywana jest łącznie z innymi zasobami. Także założenie, że długość kolejki jest nieskończona, oderwane jest od rzeczywistości, albowiem zadania rezydują w pamięci operacyjnej, która jest ograniczona.

Parametry sieci otwartej:

a) średnia intensywność napływu zadań: $\lambda = 0.7$ zad/s,

b) średni czas obsługi dla centralnego procesora: $\frac{1}{\mu_1} = 30$ ms,

c) średni czas obsługi w pamięci bębnowej: $\frac{1}{\mu_2} = 20$ ms,

d) średni czas obsługi w stacji dysków: $\frac{1}{\mu_3} = 80$ ms,

e) średni czas obsługi operacji wejścia: $\frac{1}{\mu_4} = 500$ ms,

f) prawdopodobieństwo przetwarzania w pamięci bębnowej: $r_{12} = 0.75$,

g) prawdopodobieństwo powrotu do procesora centralnego, po obsłudze w pamięci bębnowej: $r_{21} = 1.0$,

h) prawdopodobieństwo przetwarzania w procesorze centralnym, po obsłudze w stacji dysków: $r_{31} = 0.9$.

Stan zadanej wyżej sieci określony jest przez rozkład wektora $\mathbf{n} = (n_1,\ n_2,\ n_3,\ n_4)$, a ogólna liczba zadań w systemie równa jest:

$$N = \sum_{i=1}^{4} n_i$$

Rozwiązaniem w powyższym modelu są stacjonarne prawdopodobieństwa rozróżnialnych stanów sieci, czyli:

$$p(n_1,\ n_2,\ n_3,\ n_4) = p_1(n_1) \cdot p_2(n_2) \cdot p_3(n_3) \cdot p_4(n_4)$$

a to oznacza, że liczba zadań na stanowisku i nie zależy od stanu innych stanowisk i można każde stanowisko traktować jako wydzielony system typu M/M/1 i wtedy:

$$p_i(n_i) = (1 - \rho_i) \cdot \rho_i^{n_i} \qquad \text{dla } i = 1,\ 2,\ 3,\ 4$$

gdzie

$$\rho_i = \frac{\lambda_i}{\mu_i}$$

oraz

$$\begin{aligned}
\lambda_1 &= r_{21} \cdot \lambda_2 + r_{31} \cdot \lambda_3 + r_{41} \cdot \lambda_4 \\
\lambda_2 &= r_{12} \cdot \lambda_1 \\
\lambda_3 &= r_{13} \cdot \lambda_1 \\
\lambda_4 &= \lambda
\end{aligned}$$

pamiętając o tym, że jest to sieć typu Jacksona, możemy od razu obliczyć średnią liczbę zadań na stanowiskach obsługi i tak dla stanowiska i mamy:

$$\overline{n_i} = \frac{\rho_i}{1-\rho_i}$$

a to z kolei pozwala na obliczenie, ze wzoru Little'a, średniego czasu odpowiedzi (reakcji) systemu, rozumianego tutaj jako średni czas pomiędzy wejściem danego zadania do systemu komputerowego a jego opuszczeniem po zakończeniu obsługi:

$$\overline{t} = \frac{1}{\lambda} \cdot N = \frac{1}{\lambda} \sum_{i=1}^{4} \frac{\rho_i}{1-\rho_i}$$

Oznaczając przez $\overline{e_i}$ liczbę przejść przez zadane stanowisko obsługi i bazując na zadanych parametrach otrzymamy:

$$\begin{array}{lllllllll}
\lambda_1 & = & 28\ \text{zad/s} & \rho_1 & = & 0.84 & e_1 & = & 40 \\
\lambda_2 & = & 21\ \text{zad/s} & \rho_2 & = & 0.42 & e_2 & = & 30 \\
\lambda_3 & = & 7\ \text{zad/s} & \rho_3 & = & 0.56 & e_3 & = & 10 \\
\lambda_4 & = & 0.7\ \text{zad/s} & \rho_4 & = & 0.35 & e_4 & = & 1
\end{array}$$

stąd:
- średnia liczba zadań w systemie: $\overline{N} = 7.78$,
- średni czas reakcji: $\overline{t} = 11.1$ s,
- dla procesora centralnego: $\overline{n_1} = 5.25$, oraz $\overline{t_1} = 7.5$ s,
- dla pamięci bębnowej: $\overline{n_2} = 0.72$, oraz $\overline{t_2} = 1.034$ s,
- dla stacji dysków: $\overline{n_3} = 1.27$, oraz $\overline{t_3} = 1.818$ s,
- dla procesora wejścia: $\overline{n_4} = 0.54$, oraz $\overline{t_4} = 0.769$ s.

Analizując otrzymane wyniki, widzimy że tzw. "wąskim gardłem" w systemie jest centralny procesor i od jego pracy zależy wydajność całego

systemu komputerowego, a maksymalna intensywność wejściowego strumienia zadań może być określona z wykorzystaniem tzw. całkowitego nasycenia najbardziej obciążonego stanowiska obsługi, czyli procesora centralnego, gdy $\rho_1 = 1.0$ i wtedy:

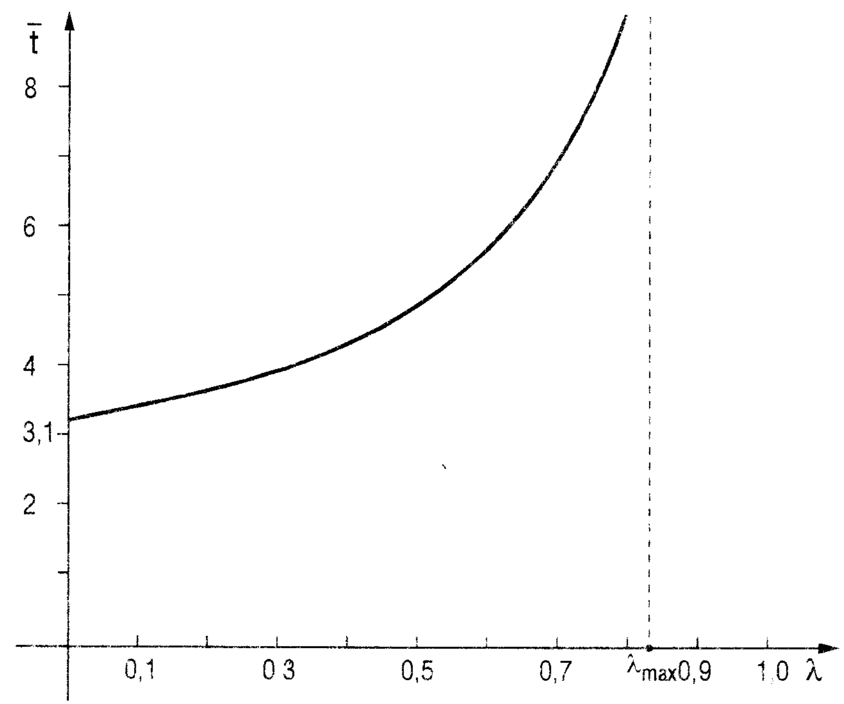
$$\rho_1 = \frac{\lambda_1}{\mu_1} = \frac{1}{\mu_1} \cdot \frac{1}{1 - r_{21}\, r_{12} - r_{31}\, r_{13}} \cdot \lambda = 1.2\ \lambda$$

stąd obliczymy $\lambda_{\max}$, przy założeniu, że $\rho_1 = 1.0$:

$$1.2\ \lambda_{\max} = 1.0$$

$$\lambda_{\max} = 0.833\ \text{zad/s}$$

Ogólnie rzecz ujmując, w zależności od intensywności napływu zadań, wykres średniego czasu reakcji systemu wygląda następująco:



Dla wartości λ bliskiej zero $\overline{t} = 3.1$ s, bo jeśli przyjąć, że czas pobytu w kolejce jest równy zero, zadanie potrzebuje średnio

$0.03 \cdot 40 = 1.2$ s na obsługę przez procesor centralny, $0.02 \cdot 30 = 0.6$ s na obsługę przez pamięć bębnową, $0.08 \cdot 10 = 0.8$ s na obsługę w stacji dysków i $0.5 \cdot 1 = 0.5$ s na obsługę operacji wejścia i jest to średni czas jednoprogramowego wypełnienia zadania.

Dla $\lambda = 0.7$ zad/s ogólny czas reakcji systemu większy jest około 3.6 raza od minimalnej wartości $\overline{t}$ i opóźnienie to wynika ze współzawodnictwa w dostępie do zasobów systemu, zwiększające:

- średni czas wejścia z 0.5 s do 0.769 s,
- średni czas centralnego procesora z 1.2 s do 7.5 s,
- średni czas w pamięci bębnowej z 0.6 s do 1.034 s,
- średni czas w pamięci dyskowej z 0.8 s do 1.818 s.

Rozdział 6

Modele deterministyczne

Dział modeli deterministycznych zajmuje ważne miejsce, podobnie jak dział modeli probabilistycznych, w przedmiocie zwanym modelowanie. Modele z tej grupy są bardzo przydatne w wielu zastosowaniach matematyki numerycznej i informatyki (np. maszyny Turinga, czy sieci Petriego). Modele deterministyczne przydatne są do właściwego zrozumienia działania systemów komputerowych, do opracowywania modeli symulacyjnych. Zastosowanie ich do badania wydajności i obliczania parametrów takich systemów jest jednak ograniczone.

Podstawową trudnością w procesie obliczania wydajności systemów komputerowych czy sieci, w przypadku stosowania modeli deterministycznych, jest brak możliwości ujmowania częstych zmian obciażenia takich systemów, co jest regułą działania tych systemów. W takich modelach obciążenie systemu może być opisane za pomocą pewnego zbioru zdeterminowanych parametrów, przy dodatkowych założeniach, że np. zadania przychodzą do systemu jednocześnie lub pojawiają się ze stałą częstotliwością, lub na obciążenie systemu składa się kilka zadań, pojawiających się periodycznie. Wtedy stany, w których mogą znaleźć się te zadania, mogą być przedstawione bardzo dokładnie. Często modele deterministyczne stosuje się do opisania pracy pewnych systemów specjalizowanych, które obciążane są periodycznie jednakową klasą zadań, ale nawet w takich modelach, z jedną klasą zadań, modele probabilistyczne są równie przydatne.

Deterministyczne modele stosuje się również jako początkowe przybliżenie, do oceny prostych miar wydajności systemów komputerowych i są to tzw. modele wartości średnich, które mogą być traktowane jako otrzymane z modeli probabilistycznych, przez zamianę losowych parametrów ich wartościami średnimi. Model wartości średnich do obliczania obciążenia systemu składa się z wielu jednakowych zadań, z charakterystykami równymi wartości średniej dla zadań modelowanego obciążenia. Zadania te przychodzą do systemu ze stałą intensywnością, równą wartości średniej intensywności z realnego obciążenia systemu. Inaczej to ujmując, można powiedzieć, że zadanie nie pojawia się i nie znika, a stała ich liczba krąży w systemie.

Bardziej dokładne rezultaty można otrzymać dla modeli, które są częściowo probabilistyczne a częściowo są modelami wartości średnich. Modele tego typu zaliczane są również do modeli deterministycznych.

Modele deterministyczne, wykorzystywane są często dla oceny przypadków granicznych (najbardziej niekorzystnych i najlepszych), w których losowe parametry zamieniane są ich wielkościami granicznymi. W tym przypadku wartości graniczne określane są jako wartość średnia plus, minus odchylenie standardowe. Ta klasa modeli nazywana jest modelami wartości granicznych.

Omawiając dwa proste modele deterministyczne ukażemy dalej przyjęte przybliżenia, sposoby układania równań i rezultaty, które mogą być otrzymane z tych modeli.

6.1 Model systemu jednoprogramowego

Jest model deterministyczny do oceny i porównywania różnych algorytmów pracy równoległej procesora centralnego i urządzeń wejścia - wyjścia. Jako że modelowany jest jednoprogramowy system, modelowana jest praca procesora centralnego i urządzeń we - wy tylko dla jednego i tego samego zadania. System taki składa się z procesora centralnego, bloku pamięci operacyjnej, jednego lub kilku urządzeń pamięci zewnętrznej i jednego lub kilku kanałów między tymi pamięciami i pamięcią główną. Zakłada się też, że na obciążenie systemu wpływa tylko to jedno zadanie, które nieskończenie długo opracowuje informacje z bazy danych, znajdujące się w pamięci zewnętrznej.

Pewne bloki informacji, podlegające opracowywaniu, ładowane są do pamięci operacyjnej, a po zakończeniu obsługi przesyłane są z powrotem

do pamięci zewnętrznych. Bloki te nie są rekordami, a składają się z rekordów o takiej samej długości, np. r słów, a rozmiar bloku, to b i jest to też wielokrotność r.

Tzw. współczynnik zblokowania, określany jest w następujący sposób:

$$g = \frac{b}{r} \tag{6.1}$$

Dostępna pamięć operacyjna M dzieli się na v obszarów, każdy z których zawiera informacje z jednego z bloków, czyli:

$$v = \frac{M}{b} \tag{6.2}$$

W czasie, gdy CPU przetwarza pewien blok informacji, ulokowany w jednym z tych obszarów, reszta obszarów pełni rolę buforów dla kanałów wejścia - wyjścia.

Pamięć zewnętrzna i kanały we - wy charakteryzują sie:
- stałym czasem dostępu t_a do każdego bloku,
- stałym czasem przesyłki słowa, równym t_w.

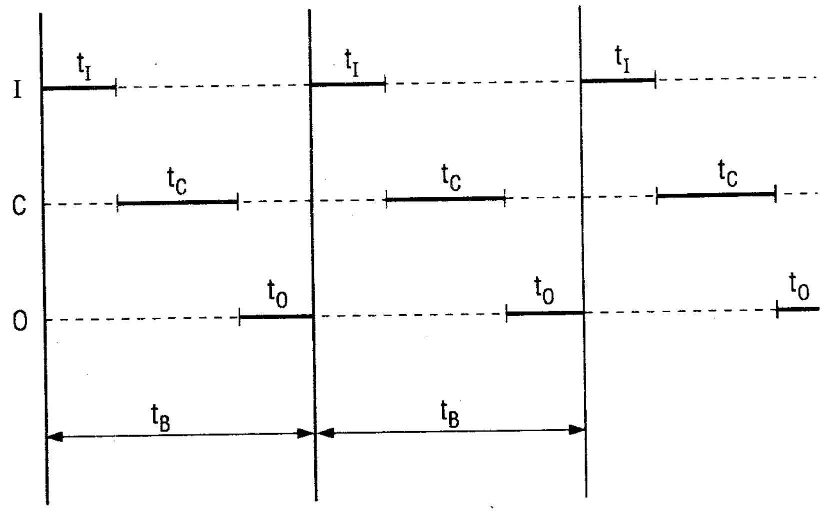
Oznacza to, że czas przesłania rekordu i bloku, wynosi odpowiednio: $r \cdot t_w$ i $b \cdot t_w$. Czas obsługi w CPU dowolnego rekordu, oznaczymy jako t_{CPU}.

Miarą wydajności systemu, obliczoną przy zastosowaniu tego modelu, jest przepustowość T, wyrażona liczbą rekordów przetworzonych w jednostce czasu. Jeżeli wybrana jednostka czasu jest dużo większa niż t_{CPU}, to przepustowość uważana jest za stałą.

Podstawowe algorytmy pracy CPU i urządzeń zewnętrznych, odpowiadające liczbie obszarów w pamięci operacyjnej, to:

1. Praca szeregowa (jeden obszar w pamięci operacyjnej) - brak obsługi równoległej.

Wprowadzanie bloku to (oś I) faza obsługi zadania (oś C) i wyjście (oś O). Tak jak $v = 1$, to następny blok nie może być wprowadzony do czasu, gdy poprzedni blok nie zakończy pełnej obsługi:



Rys. 12 Diagram pracy szeregowej

Wartości czasu wejścia, przetwarzania i wyjścia równe są:

$$\begin{aligned} t_I &= t_a + b\ t_w \\ t_C &= g\ t_{CPU} \\ t_O &= t_a + b\ t_w = t_I \end{aligned} \tag{6.3}$$

Czas przetwarzania jednego rekordu $\frac{1}{T}$ można wyrazić jako prostą funkcję parametrów modelu. Określmy czas t_B jako iloraz stosunkowo długiego czasu pracy, do liczby przetworzonych bloków. Jest rzeczą oczywistą, że $\frac{1}{t_B}$ - to przepustowość systemu, wyrażona w liczbie bloków w jednostce czasu, czyli:

$$\frac{1}{T} = \frac{t_B}{g} \tag{6.4}$$

dlatego, że każdy blok zawiera g rekordów.

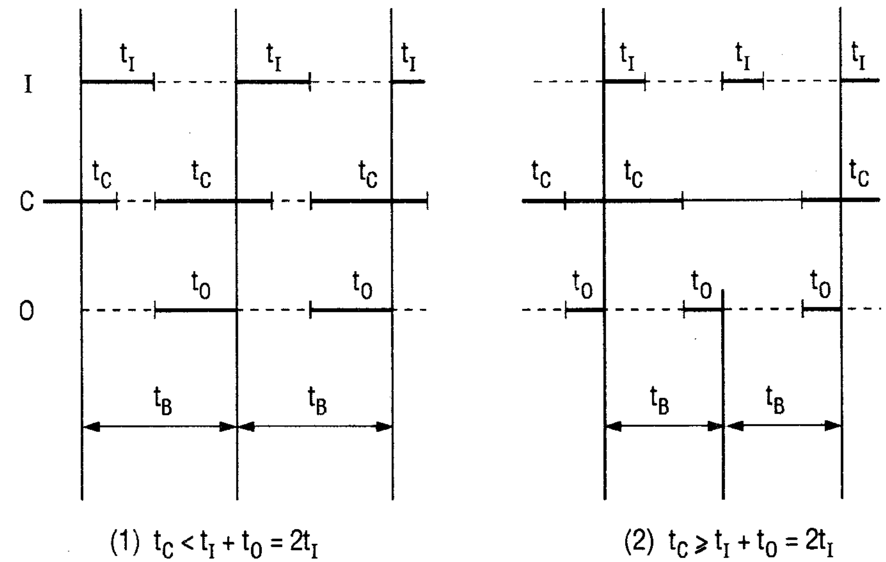
Z diagramu wynika też że:

$$t_B = t_I + t_C + t_O = g\ t_{CPU} + 2\ t_a + 2\ b\ t_w \tag{6.5}$$

i oznacza to, że:

$$\frac{1}{T} = t_{CPU} + \frac{2\ t_a}{g} + 2\ r\ t_w \tag{6.6}$$

2. Praca ograniczenie równoległa.



Rys. 13 Diagramy pracy ograniczenie równoległej

Algorytm ten pracuje jako: CPU - urządzenie wejścia lub jako urządzenie wyjścia - CPU. Dla $v = 2$ dopuszcza się tylko jeden kanał wejścia - wyjścia, który pracuje jako tylko wejście lub tylko jako wyjście.

W algorytmie tym dopuszczalne są tylko dwa ograniczenia na pracę systemu (rys. 13):

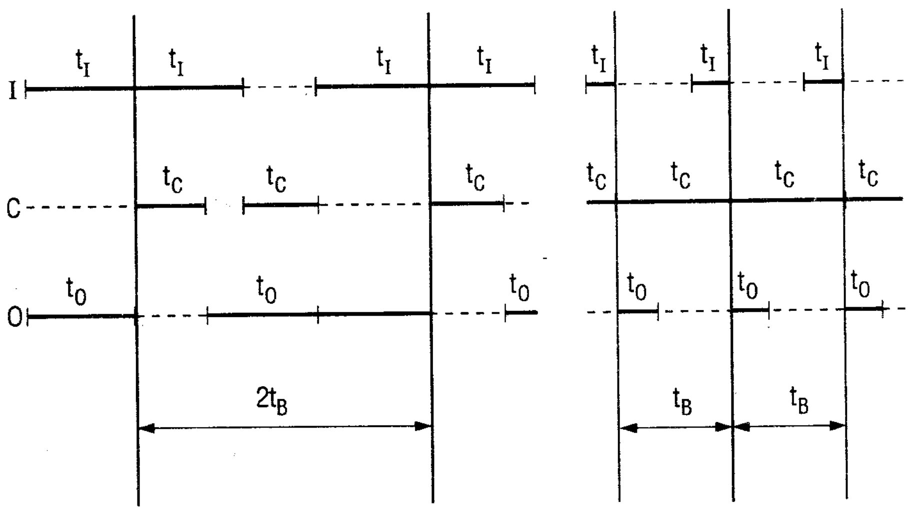
- ograniczenia dla operacji wejścia - wyjścia (1) i
- ograniczenie dla obliczeń (2).

Tutaj też, jak i w poprzednim algorytmie, łatwo można otrzymać wyrażenia dla t_B, dlatego że wyrażenia (6.3) i (6.4) odnoszą się i do tego przypadku; wzory więc dla wartości $\frac{1}{T}$ przjmują następującą postać:

$$\text{dla (1)} \qquad \frac{1}{T} = \frac{2\,t_a}{g} + 2\,r\,t_w \tag{6.7}$$

$$\text{dla (2)} \qquad \frac{1}{T} = t_{CPU} \tag{6.8}$$

3. Algorytm dwukanałowy równoległy, $v = 2$.



(1) $t_C < t_I + t_O = 2t_I$ (2) $t_C \geqslant t_I + t_O = 2t_I$

Rys. 14 Diagram pracy dwukanałowej równoległej

W tym konkretnym przypadku mamy dwa kanały wejścia - wyjścia i minimum dwa urządzenia pamięci zewnętrznej, tak że wprowadzanie i wyprowadzanie informacji może odbywać się równolegle, a w tej sytuacji CPU nie jest obciążone przetwarzaniem informacji. Tak jak i dla algorytmu 2 wystąpią tu dwa ograniczenia (rys. 14).

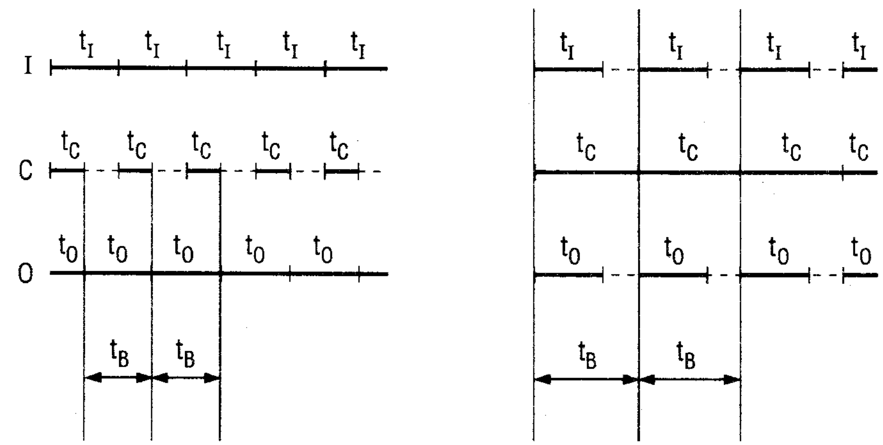
Z diagramów tych wynika, że dodatkowe możliwości tego algorytmu, w porównaniu z algorytmem 2, wpływają na wydajność systemu tylko w przypadku ograniczeń na wejście wyjście, tak jak ten algorytm zwiększa tylko wydajność urządzeń we - wy, a nie szybkość obliczeń CPU.

Wyrażenia dla $\frac{1}{T}$:

$$\text{dla (1)} \qquad \frac{1}{T} = \frac{t_{CPU}}{2} + \frac{t_a}{g} + r\,t_w \tag{6.9}$$

$$\text{dla (2)} \qquad \frac{1}{T} = t_{CPU} \tag{6.10}$$

4. Praca równoległa, $v = 3$.



(1) $t_C < t_I = t_O$ (2) $t_C \geqslant t_I = t_O$

Rys. 15 Diagram pracy równoległej

Algorytm ten zakłada, że CPU przetwarza jeden blok. W tym czasie inny blok jest wprowadzany, zaś trzeci jest wyprowadzany. Wymaga to zwiększonej pamięci operacyjnej systemu.

Jeżeli CPU jest w pełni wykorzystany (ograniczenia dla obliczeń), to wydajność systemu nie różni się zasadniczo od algorytmów z punktów 3 i 2, ale w przypadku ograniczeń dla wejścia - wyjścia (diagram (1)), czas przetwarzania rekordu jest mniejszy o $\frac{t_{CPU}}{2}$ od czasu z algorytmu 3 (rys. 15).

Wyrażenia dla $\frac{1}{T}$:

$$\text{dla (1)} \qquad \frac{1}{T} = \frac{t_a}{g} + r\, t_w \tag{6.11}$$

$$\text{dla (2)} \qquad \frac{1}{T} = t_{CPU} \tag{6.12}$$

Gdy każdy z algorytmów ma rozwiązanie w formie analitycznej, to obliczenie miar wydajności systemu, dla dowolnej, zadanej, kombinacji parametrów modelu, jest zadaniem oczywistym.

Również nie powinno być trudności w obliczaniu miar wydajności, jako funkcji od parametrów, np. jeżeli chcemy ocenić, który z algorytmów jest najlepszy, gdy wielkość pamięci operacyjnej i długość rekordu jest zadana i stała, a wymiary bloków zmieniają się wraz z parametrem v (patrz (6.2)).

Aby odpowiedzieć na to pytanie, znormalizujmy całkowity czas przetwarzania rekordu, dzieląc go przez czas t_{CPU}, tj. czas przetwarzania jednego rekordu w procesorze centralnym:

$$z = \frac{1}{T\, t_{CPU}} \tag{6.13}$$

Wartość ta równa jest 1, gdy współczynnik obciążenia CPU = 1 i gdy jest pełna równoległość w pracy CPU i urządzeń we - wy. Tymczasem najczęściej $z > 1$ (np. przy ograniczeniach dla we - wy) i właściwie nie trudno jest pokazać, że z jest wielkością odwrotną do ρ (współczynnika obciążenia - wykorzystania procesora).

Utwórzmy teraz znormalizowane równania dla każdego z algorytmów, jako funkcje od następujących parametrów:

$$x = \frac{r\, t_w}{t_{CPU}} \qquad \text{i} \qquad y = \frac{r\, t_a}{M\, t_{CPU}} \tag{6.14}$$

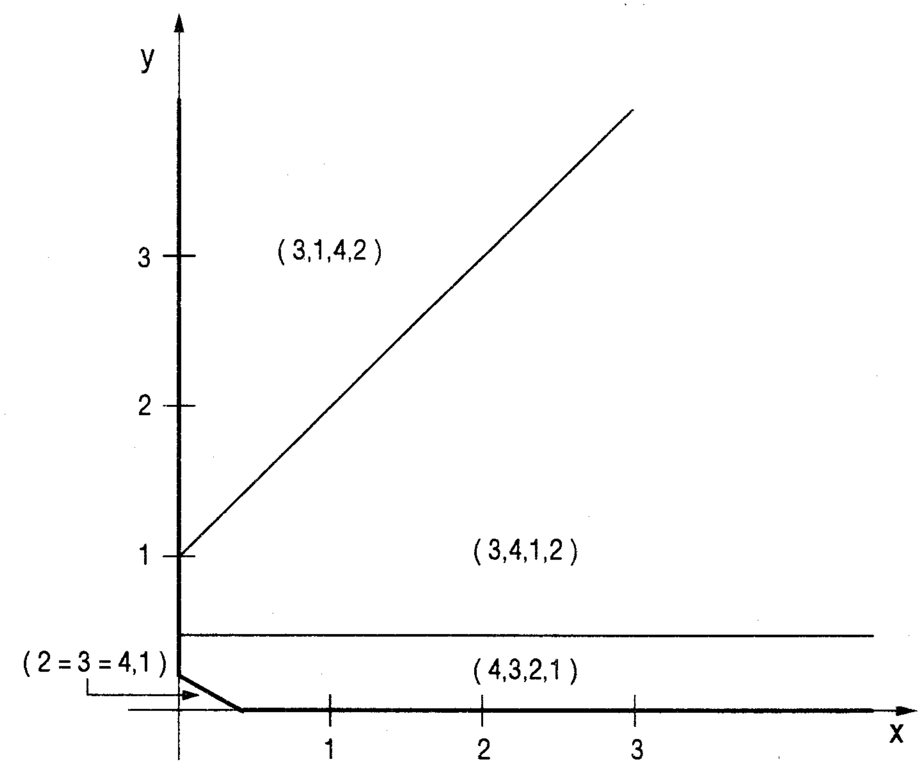
Parametry te nie zależą od wielkości b lub v, a ze wzorów (6.2) i (6.14) mamy:

$$\frac{t_a}{g\, t_{CPU}} = v\, y \tag{6.15}$$

Więc nowe równania dla algorytmów pracy CPU i urządzeń we - wy przyjmują następującą postać:

$$\left.\begin{array}{lrcl}
\text{dla 1} & z & = & 1 + 2\,y + 2\,x \\
\text{dla 2} & z & = & 4\,y + 2\,x \\
 & z & = & 1 \\
\text{dla 3} & z & = & \frac{1}{2} + 2\,y + x \\
 & z & = & 1 \\
\text{dla 4} & z & = & 3\,y + x \\
 & z & = & 1
\end{array}\right\} \tag{6.16}$$

Rysunek 16 zawiera w sobie odpowiedź, na to pytanie.



Rys. 16 Klasyfikacja algorytmów pracy CPU i urządzeń we - wy

Wielkości x i y, charakteryzujące parametry zadania, są to punkty w pierwszym kwadracie płaszczyzny (x, y). Na rysunku tym przedstawiona jest klasyfikacja algorytmów wg wydajności, w zależności od wartości wielkości x i y. Na przykład dla $x = 0,4$ i $y = 0.3$ odpowiada obszar przedstawiony jako (4, 3, 2, 1), a tutaj algorytmy ustawione są w kolejności zmniejszania się wydajności, czyli algorytm z numerem 4 jest najlepszy dla zadanych parametrów.

I jeszcze jedna ważna konkluzja wynika z tego rysunku, a mianowicie taka, że komplikując algorytm nie zawsze osiągamy wzrost wydajności, np. gdy $y > x + 1$ najlepszy okazuje się algorytm 3, a praca równoległa (algorytm 4) jest lepsza niż praca ograniczenie równoległa (algorytm 2), itp.

6.2 Model wartości średnich

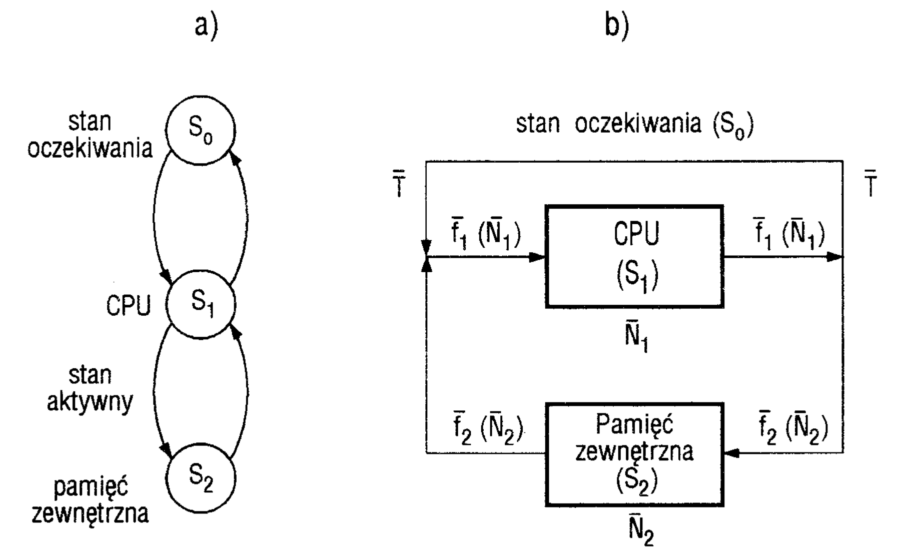
Jest to model systemu w którym zakłada się, że proces obliczeniowy i proces wejścia - wyjścia nie kolidują ze sobą i jest to model będący szczególnym przypadkiem modelu probabilistycznego.

Modelowany system jest uproszczoną wersją systemu z jednym urządzeniem pamięci zewnętrznej, ze stałym współczynnikiem wieloprogramowości N. W dowolnym momencie zadanie może mieć jeden z dwu statusów: aktywne lub pasywne. Zadanie jest w stanie pasywnym (stan S_0), jeżeli nie znajduje się ono w pamięci operacyjnej; zadanie ma status aktywnego (stan S_1), gdy oczekuje przetwarzania lub jest przetwarzane w CPU lub gdy jest w stanie S_2 - oczekując przetwarzania w pamięci zewnętrznej lub będąc w niej przetwarzane (rys. 17).

Zakłada się również, że gdy jakiekolwiek zadanie przechodzi do stanu pasywnego, inne zadanie od razu jest aktywizowane; oznacza to, że stopień wieloprogramowości jest stały. Zakłada się również, że wszystkie zadania są tego samego typu i charakteryzują się jednakowym czasem przetwarzania w CPU równym t_{CPU} oraz stałą liczbą odwołań do pamięci zewnętrznej, równą n_z.

Zadania te generowane są w regularnych odstępach czasu pracy procesora, tzn. zawsze, gdy zadanie jest przetwarzane w CPU, w odstępie czasu $\frac{t_{CPU}}{n_z+1}$, zaczynając od zakończenia ostatniej operacji z pamięcią zewnętrzną lub od początku przetwarzania zadania.

Każde odwołanie się do pamięci zewnętrznej powoduje przesyłanie bloku informacji o stałej wielkości, jednakowej dla wszystkich bloków, a pod koniec $(n_z + 1)$ - ego odstępu czasu pracy procesora zadanie przechodzi do stanu pasywnego.



Rys. 17 Model wartości średnich wieloprogramowego systemu: a) diagram dla pojedynczego zadania; b) model systemu

Oznaczmy przez $\overline{N_1}$ - średnią liczbę zadań w stanie S_1 i przez $\overline{N_2}$ - średnią liczbę zadań w stanie S_2 i wtedy:

$$\overline{N_1} + \overline{N_2} = N \tag{6.17}$$

W stanie równowagi średnia intensywnosć strumienia wchodzącego do CPU i pamięci zewnętrznej równa jest średniej intensywności strumienia wychodzącego. Intensywności te zależne są od obciążeń N_1 i N_2, więc oznaczmy je przez $\overline{f_1}(N_1)$ i $\overline{f_2}(N_2)$.

Równość średnich intensywności strumienia wchodzącego i wychodzącego odnosi się również do stanu pasywnego, dlatego że liczba N jest stała. Intensywności dla stanu pasywnego równe są wydajności $\overline{T}$, tj. średniej przepustowości systemu.

Każde z zadań przychodzi do stanu S_1, $(n_z + 1)$ razy. Pierwsze n_z przejść, wychodzac ze stanu S_1 trafia do stanu S_2, ostatnie zaś przejście przesyła zadanie do stanu S_0, czyli:

$$\overline{f_2}(\overline{N_2}) = \frac{n_z}{n_z + 1}\,\overline{f_1}(\overline{N_1}) \tag{6.18}$$

oraz

$$\overline{T} = \overline{f_1}(\overline{N_1}) - \overline{f_2}(\overline{N_2}) = \frac{\overline{f_1}(\overline{N_1})}{n_z + 1} \tag{6.19}$$

Jeżeli wartość $\overline{f_1}(\overline{N_1})$ jest znana, to ze wzoru (6.19) można obliczyć średnią przepustowość. Mamy tutaj:

$$\overline{f_1}(\overline{N_1}) = \begin{cases} 0 & \text{dla} \quad \overline{N_1} = 0 \\ \frac{n_z+1}{t_{CPU}} & \text{dla} \quad \overline{N_1} \geq 1 \\ \overline{N_1}\,\frac{n_z+1}{t_{CPU}} & \text{dla} \quad 0 < \overline{N_1} < 1 \end{cases} \tag{6.20}$$

Dwa pierwsze przypadki są trywialne. Gdy $\overline{N_1} = 0$, wszystkie aktywne zadania oczekują przetwarzania w pamięci zewnętrznej, która potrzebuje nieskończonej wielkości czasu dla przetworzenia zadań, czyli obciążenie pamięci zewnętrznej osiąga wartość graniczną i przepustowość systemu równa jest 0.

Dla $\overline{N_1} \geq 1$ procesor ciągle jest zajęty i system ma maksymalną przepustowość, równą:

$$\overline{T}_{\max} = \frac{1}{t_{CPU}} \tag{6.21}$$

Najbardziej interesujący jest trzeci przypadek, gdy obciążenie procesora zawarte jest w przedziale od 0 do 1. Do rozwiązania tego zadania potrzebna jest wielkość $\overline{f_2}(\overline{N_2})$.

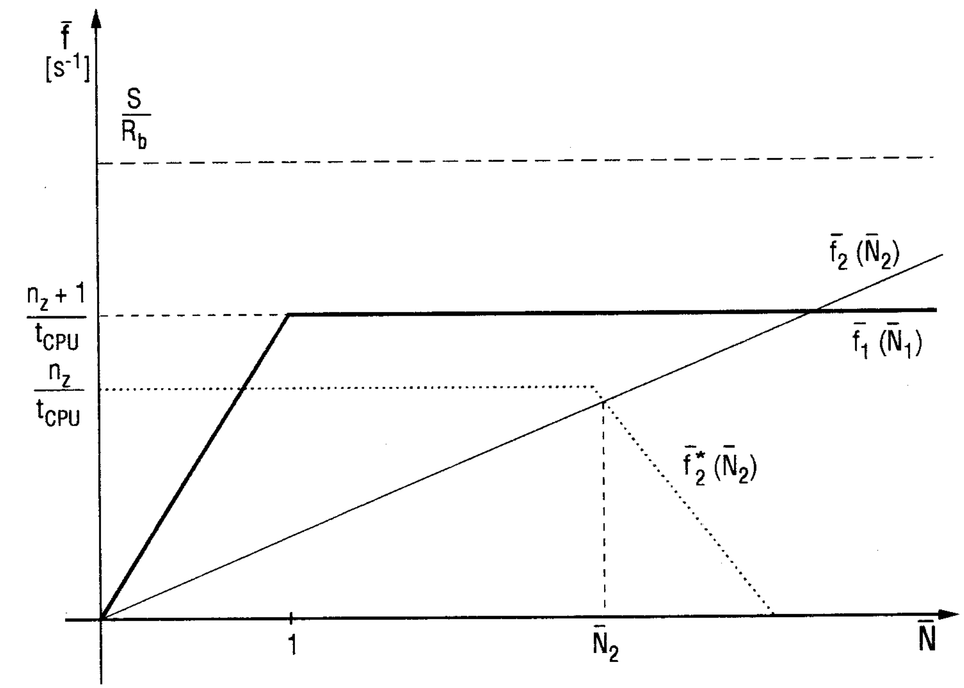
Wielkość tę można otrzymać z pomiarów lub przez modelowanie, albo z modelu analitycznego dla pamięci zewnętrznej.

Załóżmy, że pamięcią zewnętrzną jest pamięć bębnowa. Wtedy maksymalna intensywność strumienia wyjściowego, dla podzielonego na sektory bębna, charakteryzującego się czasem obsługi R_b i liczbą sektorów S, równa jest $\frac{S}{R_b}$ bloków w jednostce czasu. Odpowiada to przesyłce jednego bloku.

Jeżeli założymy, że odwołanie się do bębna składa się z jednego bloku, to:

$$\overline{f_2}_{\max} = \overline{f_2}(\infty) = \frac{S}{R_b} \tag{6.22}$$

Postać funkcji $\overline{f_2}(\overline{N_2})$ wygląda zwykle jak na rys. 18.



Rys. 18 Wykresy funkcji f

Najprostszym rozwiązanie dla tego modelu jest rozwiązanie graficzne, sprowadzające się do narysowania w płaszczyźnie $(\overline{N_2}, \overline{f_2})$ następującej funkcji:

$$\overline{f_z^*} = \begin{cases} 0 & \text{dla} \quad \overline{N_2} = N \\ \frac{n_z}{t_{CPU}} & \text{dla} \quad \overline{N_2} \leq N - 1 \\ \frac{n_z}{t_{CPU}}\,(N - \overline{N_2}) & \text{dla} \quad N - 1 < \overline{N_2} < N \end{cases} \tag{6.23}$$

Funkcja $\overline{f_2^*}$ powstaje z podstawienia (6.17) do (6.20) i (6.20) do (6.18). Punkt przecięcia krzywej $\overline{f_2^*}$ z krzywą $\overline{f_2}$ daje wartość $\overline{N_2}$, co jest rozwiązaniem modelu. Znając $\overline{N_2}$ można obliczyć $\overline{N_1}$, a z wyrażenia (6.19) wartość $\overline{T}$.

Ponadto analiza rysunku pozwala wysnuć bezpośrednie wnioski o rezultatach ewentualnych zmian modelowanych parametrów, i tak np. przy zwiększaniu N rośnie wartość $\overline{N_1}$ do 1 oraz wartość $\overline{T}$ do $\overline{T}_{\max}$. Rezultat taki można otrzymać dla bardziej wydajnego bębna z mniejszym R_b, lub przy większej gęstości zapisu (wzrost S) prowadzącym do szybkiego wzrostu wartości krzywej $\overline{f_2}(\overline{N_2})$.

Zwiększanie zaś $\overline{T}$ może być osiągnięte przez zmniejszanie n_z. Zwiększenie t_{CPU} prowadzić będzie do wzrostu $\overline{N_1}$ i przybliżania $\overline{T}$ do $\overline{T}_{\max}$, przy jednoczesnym zmniejszaniu wartości $\overline{T}_{\max}$.

Rozdział 7

Literatura

1. Gross D., Harris C. M. *Fundamentals of queueing theory*, John Wiley & Sons, Inc., 1974.

2. Kleinrock L. *Queueing Systems*, Volume I: *Theory*, John Wiley, 1975.

3. Kleinrock L. *Queueing Systems,* Volume II: *Computer Applications,* John Wiley, 1976.

4. Jaiswal N. *Priority Queues,* Academic Press, 1968.

5. Ferrari D. *Computer Systems Performance Evaluation,* Prentice - Hall, 1978.

6. Klimow G. P. *Procesy obsługi masowej,* WNT, 1979.

7. Węgrzyn S. *Podstawy informatyki,* PWN, 1982.

8. Rajski J., Tyszner J. *Modelowanie i symulacja cyfrowa,* Wyd. Politechniki Poznańskiej, 1986.

9. Zeigler B. P. *Teoria modelowania i symulacji,* PWN, 1984.

10. Czachórski T. *Numeryczne i przybliżone metody obliczeniowe w statycznych modelach systemów komputerowych,* Podstawy sterowania, tom 13, z. 4, 1983, pp. 293 - 314.

11. Oniszczuk W. *Analiza działania złożonych systemów i sieci komputerowych jako sieci kolejkowych z priorytetowym szeregowaniem zadań,* Archiwum Automatyki i Telemechaniki, z. 1, 1988, pp. 165 - 181.

12. Gelenbe E., Mitrani I. *Analysis and Synthesis of Computer Systems,* Academic Press, 1980.

13. Gelembe E., Pujolle G., Nelson J. C. C. *Introduction to Queueing Networks,* John Wiley & Sons, 1987.