ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1982

Seria: AUTOMATYKA z. 63 Nr kol. 735

Bogdan REMBOWSKI

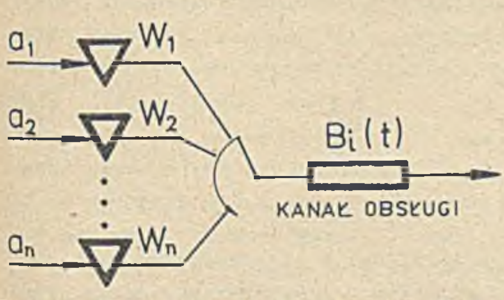
Instytut Elektroenergetyki i Automatyki
Politechnika Gdańska

OBSŁUGA PROCESÓW DYSKRETNYCH Z PRIORYTETAMI

Streszczenie. Dla systemu obsługi masowej o wielu strumieniach zgłoszeń i ograniczonej pojemności buforów sformułowano regułę określającą, w zależności od stanu systemu, które zgłoszenie trzeba obsłużyć i z którego strumienia, tak aby nie nastąpiło przepełnienie któregokolwiek z buforów. Zadanie rozwiązuje się metodą określenia priorytetów dla poszczególnych kanałów. Problem sprowadza się do rozwiązania zadania programowania liniowego.

1. SFORMUŁOWANIE PROBLEMU

Do systemu obsługi masowej napływa n niezdeterminowanych strumieni zgłoszeń z intensywnościami odpowiednio a_i zgłoszeń na jednostkę czasu. Jednocześnie może się odbywać obsługa tylko jednego zgłoszenia i nie dopuszcza się jej przerwania. Zgłoszenie zastające system pustym jest natychmiast obsługiwane, w przypadku zajętości kanału obsługi zgłoszenie ustawiane jest kolejce w buforze danego strumienia (patrz rys. 1). Bufory wejściowe dla każdego ze strumieni mają ograniczone pojemności W_i ($i = 1,2,\ldots,n$). Czas obsługi zgłoszenia strumienia i określony jest dystrybuantą $B_i(t)$.



Rys. 1. System obsługi masowej o n strumieniach zgłoszeń i jednym kanale obsługi

Należy określić porządek obsługi zgłoszeń, tzn. regułę określającą, które zgłoszenie należy wybrać do przetwarzania w kanale obsługi i z którego strumienia, tak aby nie nastąpiło przepełnienie któregokolwiek z buforów wejściowych.

W celu zapewnienia podstawowych warunków rozwiązywalności zadania należy wprowadzić następujące ograniczenia:

- pierwszy i drugi moment statystyczny czasu obsługi w dowolnym strumieniu są skończone, tzn.

$$\beta_{i1} = \int_0^\infty t dB_i(t) < \infty, \quad \beta_{i2} = \int_0^\infty t^2 dB_i(t) < \infty \qquad (i = 1,2,\dots,n)$$

- spełniony jest warunek ergodyczności, tzn. w systemie w stanie ustalonym nie tworzy się nieskończenie długa kolejka w którymkolwiek z buforów (system nie jest zablokowany zbyt dużą liczbą zgłoszeń) tzn.:

$$p = \sum_{i=1}^{n} a_i \cdot \beta_{i1} < 1.$$

Działaniom systemów obsługi masowej z ograniczoną pojemnością buforów i różnymi strumieniami wejściowymi poświęcony jest cykl prac w tym [1,2]. Rozpatrywane są algorytmy określenia prawdopodobieństwa stanu tych systemów przy określonej strategii nadawania priorytetów. Zadanie nadawania priorytetów w pracach tych nie było rozpatrywane. Zadanie nadawania priorytetów zależnych od stanu systemu (priorytety dynamiczne) najpierw było rozwiązane w pracach [3, 7]. W pracy [5] podano procedurę określania priorytetów dynamicznych dla systemów z ograniczoną kolejką, jednak wyłącznie dla specjalnej klasy systemów, w których czas obsługi jednego zgłoszenia jest zmienną losową z dystrybuantą wykładniczą. Problem nadawania priorytetów w systemie z nieograniczoną kolejką, strumieniami zgłoszeń typu Poissona i dowolnym skończonym czasem obsługi rozpatrywany był w pracy [4].

2. SPOSÓB ROZWIĄZANIA ZADANIA

Porządek obsługi zgłoszeń zostanie określony w zależności od stanu systemu. W każdej chwili czasu kolejka w systemie scharakteryzowana jest wektorem:

$$l = \left\{ l_i,\ i \in \Omega \right\}, \qquad \Omega = \left\{ 1,2,\dots,n \right\},$$

gdzie l_i jest długością kolejki w buforze strumienia i. Po zakończeniu obsługi zgłoszenia z pewnego strumienia, wybór kolejnego strumienia obsługi dokonuje się w zależności od liczb l_i, pozostałych zgłoszeń w systemie według odwzorowania $l \mapsto u(l)$. Funkcję $u = u(l)$ określającą porządek obsługi zgłoszeń w systemie, nazywa się funkcją przełączeń. Do określenia funkcji przełączeń posłużono się metodą priorytetów zdefiniowanych następująco. Mówimy, że strumień α ma wyższy priorytet niż strumień γ, jeśli w dowolnej chwili rozpoczęcia obsługi zgłoszenia strumienia γ liczba zgłoszeń w kolejce strumienia α jest równa zeru.

Na wstępie rozwiązany zostanie problem uproszczony, w którym strumienie zgłoszeń są strumieniami Poissona z intensywnościami a_i $(i = 1,2,\ldots,n)$ i dopuszcza się nieograniczoną kolejkę w każdym z buforów. Po rozwiązaniu zadania uproszczonego zostaną nałożone ograniczenia na pojemności buforów i określona zostanie wrażliwość funkcji przełączeń na zmianę tej pojemności.

Rozwiązanie zadania przebiega w sposób następujący. Definiuje się zmienne x_{ij}, które pozwalają obliczyć priorytety. Określają one priorytety w sposób następujący: strumień i ma wyższy priorytet od strumienia j, gdy $x_{ij} \neq 0$.

Korzystając z metody funkcji tworzącej buduje się podstawową zależność wiążącą parametry obsługiwanego procesu: intensywność napływu zgłoszeń a_i z pierwszym i drugim momentem statystycznym czasu obsługi. Równanie zbudowane metodą funkcji tworzącej i definicja zmiennych x_{ij} pozwalają na wzajemne ich powiązanie z parametrami obsługiwanego procesu w układ równań liniowych.

3. ROZWIAZANIE ZADANIA

Rozpatrzmy na wstępie następujący wielowymiarowy proces stochastyczny:

$$\eta(t) = \left\{n(t),\ i(t)\right\}, \qquad t \geqslant 0,$$

gdzie $n(t) = \left\{n_i(t),\ i \in \Omega\right\}$; $n_i(t^*)$ jest zmienną losową określającą liczbę zgłoszeń w chwili t^* w buforze strumienia i; $i(t^*)$ jest zmienną losową określającą numer strumienia, na którym dokonuje się obsługi; gdy nie ma zgłoszeń w systemie $i(t^*)$ przyjmuje wartość zero.

Niech n_{iN} będzie zmienną losową określającą liczbę zgłoszeń pozostających w buforze strumienia i po zakończeniu N operacji obsługi; i_N określa strumień, na którym dokonuje się obsługi.

Proces $\eta_N = \left\{n_N,\ i_N\right\}$ jest włożony, wg terminologii Kendalla, w proces $\eta(t)$ i tworzy łańcuch Markowa.

Do rozwiązania zadania określenia priorytetów posłużono się metodą funkcji tworzącej. Interpretacja funkcji tworzącej metodą zdarzenia dodatkowego ułatwia opis fizyczny równań i ułatwia ich dowody.

Zakładamy, że każde zgłoszenie przybyłe do bufora i-tego strumienia uznajemy za czerwone z prawdopodobieństwem Z_i i niebieskie z prawdopodobieństwem $(1-Z_i)$. Przyjmujemy następujące oznaczenia i ich interpretacje:

$P_{iN}(Z) = E\, Z^{n_N}\delta_{i,iN}$ - prawdopodobieństwo tego, że N-ta obsługa dotyczy zgłoszenia strumienia i-tego i po zakończeniu obsługi w systemie nie ma niebieskich zgłoszeń.

Przyjmuje się, że δ_{ij} jest deltą Kronekera: $\delta_{ij} = 1$, gdy i=j; $\delta_{ij} = 0$, gdy $i \neq j$.

$S_N(Z) = E\, Z^{n_N} = \sum_{i=1}^{n} P_{iN}(Z)$ - prawdopodobieństwo tego, że po zakończeniu obsługi N-tego zgłoszenia w systemie nie ma niebieskich zgłoszeń. Zgodnie z powyższą definicją $S_N(0)$ oznacza, że w systemie nie ma zgłoszeń.

$R_{iN}(Z) = E\, Z^{n_N}\cdot\delta_{i,u(n_N)}$ - prawdopodobieństwo tego, że po zakończeniu N operacji obsługi następna obsługa będzie dotyczyła zgłoszenia strumienia i-tego i że w systemie pozostałe zgłoszenia są koloru czerwonego.

$b_i\left(\sum_{i=1}^{n} a_i(1-Z_i)\right) \overset{df}{=} \int_0^{\infty} e^{-\sum_{i=1}^{n} a_i(1-Z_i)t}\, dB_i(t)$ - prawdopodobieństwo tego, że podczas obsługi zgłoszenia strumienia i-tego do buforów nie napłyną zgłoszenia koloru niebieskiego.

Po wprowadzeniu powyższych oznaczeń można zapisać podstawową zależność:

$$Z_i P_{iN+1}(Z) = \left[R_{iN}(Z) + S_N(0)\frac{a_i}{\sum_{i=1}^{n} a_i} Z_i\right] b_i\left(\sum_{i=1}^{n} a_i(1-Z_i)\right) \tag{1}$$

Dowód słuszności równania (1) można przeprowadzić następująco.

Na to, aby N+1 operacja obsługi przebiegała dla strumienia i, a po jej zakończeniu w systemie nie zostały niebieskie zgłoszenia i zgłoszenie obsługiwane było czerwone (czego prawdopodobieństwo jest równe $Z_i P_{i,N+1}(Z)$) potrzeba i wystarcza by:

- po zakończeniu N operacji obsługi w systemie nie pozostały niebieskie zgłoszenia, a w następnej operacji było w obsłudze zgłoszenie czerwone ze strumienia i:

$$\left(R_{iN}(Z) + S_N(O)\frac{a_i}{\sum_{i=1}^{n} a_i} Z_i\right)$$

- podczas N+1 operacji obsługi w strumieniu i do systemu nie przybyły nowe niebieskie zgłoszenia:

$$\left(b_i \sum_{i=1}^{n} a_i(1-Z_i)\right).$$

Rozpatrywany jest wyłącznie stan stacjonarny. W równaniu (1) zapisujemy wobec tego $N \to \infty$ (dowód istnienia stanu stacjonarnego znaleźć można w pracy [6]).

Korzystamy z następujących oznaczeń:

$$x_{ij} = \lim_{N \to \infty} E\, n_{jN} \cdot \delta_{i,u(n_N)}$$

$x_j = \lim\limits_{N \to \infty} E\, n_{jN}$ - średnia liczba zgłoszeń w buforze strumienia j po zakończeniu N operacji obsługi, gdy $N \to \infty$.

Łatwo zauważyć, że strumień i ma wyższy priorytet od strumienia j, gdy $x_{ij} \neq 0$. Może nastąpić przełączenie kanału obsługi z obsługi strumienia j na strumień i, chociaż liczba zgłoszeń w buforze strumienia j jest różna od zera.

Słuszna jest też zależność:

$$x_j = \sum_{i=1}^{n} x_{ij}.$$

Dwukrotne różniczkowanie lewej i prawej strony równania (1) pozwala określić zależność (układ równań) wiążącą parametry obsługiwanego procesu: intensywność napływu zgłoszeń a_i, pierwszy i drugi moment statystyczny czasu obsługi, z wielkościami x_{ij}.

Poniżej określono przekształcenia niezbędne do budowy układu równań. Są one następujące:

a) $$\frac{\partial^2}{\partial z_j \partial z_i} S_N(z) = \frac{\partial^2}{\partial z_j \partial z_i} \prod_{\alpha \in \Omega} z_\alpha^{n_{\alpha N}} = n_{iN} \cdot z_i^{n_{iN}-1} \cdot n_{jN} \cdot z_j^{n_{jN}-1} \cdot$$

$$\cdot \prod_{\substack{\alpha \in \Omega \\ \alpha \neq i \\ \alpha \neq j}} z_\alpha^{n_{\alpha N}} (1-\delta_{ij}) + n_{iN}(n_{jN}-1) z_j^{n_{jN}-2} \cdot \prod_{\substack{\alpha \in \Omega \\ \alpha \neq i \\ \alpha \neq j}} z_\alpha^{n_{\alpha N}} \cdot \delta_{ij} \Big|_{z=1} =$$

$$= n_{iN} \cdot n_{jN} \cdot (1-\delta_{ij}) + n_{iN} \cdot (n_{jN}-1) \cdot \delta_{ij},$$

gdzie: $\delta_{ij} = 1$, gdy $i = j$; $\delta_{ij} = 0$, gdy $i \neq j$.

b) $$\frac{\partial}{\partial z_j} R_{\alpha N}(z) \Big|_{z=1} = \frac{\partial}{\partial z_j} \left[z^{n_N} \delta_{\alpha}, u(n_N) \right] = n_{jN} \cdot z_j^{n_{jN}-1} \prod_{\substack{\alpha \in \Omega \\ \alpha \neq j}} z_\alpha^{n_{\alpha N}} \delta_{\alpha}, u(n_N) \Big|_{z=1} =$$

$$= n_{jN} \delta_{\alpha}, u(n_N) = X_{\alpha j N}$$

c) $$\frac{\partial}{\partial z_j} \frac{R_{\alpha N}(z)}{z_\alpha} \Big|_{z=1} = \frac{X_{\alpha j N}}{z_\alpha} - \frac{R_{\alpha N}(z)}{z_\alpha^2} \delta_{\alpha j} \Big|_{z=1} = X_{\alpha j N} - R_{\alpha N}(1) \delta_{ij}.$$

d) $$\frac{\partial^2}{\partial z_i \partial z_j} b_\alpha(z) \Big|_{z=1} = a_i \cdot a_j \cdot \beta_{\alpha 2}.$$

e) $$\frac{\partial^2}{\partial z_i \partial z_j} \frac{R_{\alpha N}(z)}{z_\alpha} \Big|_{z=1} = \frac{\partial}{\partial z_i} \left[\frac{z_\alpha \frac{\partial}{\partial z_j} R_{\alpha N}(z) - R_{\alpha N}(z) \frac{\partial}{\partial z_j} z_\alpha}{z_\alpha^2} \right]_{z=1} =$$

$$= \frac{\partial^2}{\partial z_i \partial z_j} R_{\alpha N}(z) \Big|_{z=1} - X_{\alpha i N} \delta_{\alpha j} - X_{\alpha j N} \delta_{\alpha i} + 2R_{\alpha N}(1) \delta_{\alpha i} \delta_{\alpha j}.$$

f) $$\frac{\partial^2}{\partial z_i \partial z_j} R_{\alpha N}(z) \Big|_{z=1} = \frac{\partial}{\partial z_i} \left[\frac{\partial}{\partial z_j} R_{\alpha N}(z) \right]_{z=1} = n_{iN}(1-\delta_{ij}) X_{\alpha j N} +$$

$$+ (n_{iN}-1) \delta_{ij} X_{\alpha j N}.$$

Wykorzystanie wyżej wyprowadzonych zależności przy $N \to \infty$ prowadzi do układu równań liniowych.

Poniżej podano jego szczególny przypadek, gdy czas obsługi zgłoszenia jest zdeterminowany, tzn. $\beta_{12} = 0$. Układ równań jest następujący:

$$x_{ij} + x_{ji} = \sum_{\alpha \in \Omega} (x_{\alpha i} d_{\alpha j} + x_{\alpha j} d_{\alpha i} + C_{ij}, \qquad (2)$$

gdzie:

$$d_{ij} = a_j \cdot \beta_{il}$$

$$c_{ij} = \frac{2a_i \delta_{ij} - a_i d_{ij} - a_j d_{ji}}{\sum_{i=1}^{n} a_i}.$$

Do układu równań (2) można dołączyć wskaźnik jakości porządku obsługi zgłoszeń w postaci sumarycznego kosztu gromadzenia zgłoszeń w buforach. Po wprowadzeniu kosztu przechowywania jednego zgłoszenia w buforze strumienia i jako c_i uzyskamy:

$$I = \sum_{i \in \Omega} c_i \cdot \sum_{\alpha \in \Omega} x_{\alpha i} \qquad (3)$$

Układ równań liniowych (2) wraz ze wskaźnikiem jakości (3), który należy minimalizować tworzy zadanie programowania liniowego.

Należy podkreślić, że przedstawione zadanie programowania liniowego słuszne jest dla chwili czasowej zakończenia dowolnej obsługi. Słuszne jest jednak twierdzenie, że średnia liczba zgłoszeń w buforze strumienia i (i = 1,2,...,n) w chwili zakończenia obsługi zgłoszenia jest na pewno większa niż średnia liczba zgłoszeń w buforze tegoż strumienia, liczona za okres działania systemu. Skoro należy określić porządek obsługi zgłoszeń, tak aby nie nastąpiło przepełnienie któregokolwiek z buforów wejściowych, należy wziąć pod uwagę najbardziej niekorzystny przypadek.

Rozwiążmy zadanie ograniczonej pojemności buforów. Do układu równań (2) dołączamy dodatkowy warunek, który zapewnia, że średnia liczba zgłoszeń w buforze jest mniejsza od zadanej liczby określanej przez pojemność bufora W_i (i = 1,2,...,n). W takiej sytuacji zakładamy, że prawdopodobieństwo odrzucenia zgłoszenia z powodu zapełnienia bufora jest pomijalne. Dołączamy układ nierówności postaci:

$$\sum_{j=1}^{n} x_{ij} \leq W_i \qquad (4)$$

Obliczony porządek priorytetów w przypadku buforów o nieograniczonej pojemności zostanie teraz przekształcony w wyniku wprowadzenia warunków (4). Brak rozwiązania układu (2) i (4) świadczy o nieistnieniu takiego porządku obsługi zgłoszeń, który zapewniałby nieprzepełnienie się buforów.

4. PRZYKŁAD OBLICZENIOWY

Do systemu obsługi napływają cztery strumienie zgłoszeń typu Poissona z intensywnościami odpowiednio $a_1 = 2$, $a_2 = 4$, $a_3 = 1$ i $a_4 = 2{,}5$. Zgłoszenia przechowywane są w odpowiednich buforach o pojemnościach odpowiednio W_i ($i = 1,2,3,4$). Czasy potrzebne na obsługę zgłoszenia są zdeterminowane i wynoszą odpowiednio $\beta_{11} = 0{,}2$, $\beta_{21} = 0{,}05$, $\beta_{31} = 0{,}1$ i $\beta_{41} = 0{,}1$.

Rozwiązanie zadania podzielono na dwa etapy. W pierwszej kolejności nie wprowadzono do zadania ograniczeń pojemności buforów.

Przykładowe rozwiązanie zadania programowania liniowego dla $c_i = 1$ ($i = 1,2,3,4$) doprowadziło do następujących rezultatów:

$x_{11} = 0{,}21$	$x_{23} = 5{,}0 \cdot 10^{-7}$
$x_{21} = 2{,}4 \cdot 10^{-5}$	$x_{33} = 0{,}1$
$x_{31} = 8{,}4 \cdot 10^{-7}$	$x_{43} = 8{,}9 \cdot 10^{-7}$
$x_{41} = 1{,}5 \cdot 10^{-6}$	$x_{24} = 3{,}4 \cdot 10^{-8}$
$x_{22} = 0{,}42$	$x_{44} = 0{,}26.$

Pozwala to ustalić następującą sekwencję priorytetów (wzrastają od lewej do prawej strony):

$$1 \text{ ----- } 2 \text{ ----- } 3 \text{ ----- } 4.$$

Średnie liczby detali w każdym z buforów są następujące:

$x_1 = 0{,}21$	$x_3 = 0{,}1$
$x_2 = 0{,}42$	$x_4 = 0{,}26$

W następnej kolejności wprowadzono ograniczenia pojemności buforów. W tym celu układy (2) i (4) przekształcono w następujący sposób:

$$x_{ij} + x_{ji} = \sum_{\alpha=1}^{4} (x_{\alpha i} d_{\alpha j} + x_{\alpha j} d_{\alpha i}) + c_{ij} \qquad i,j = 1,2,3,4$$

$$\sum_{j=1}^{4} x_{ij} \leq W_i - \beta_i. \tag{5}$$

Zmieniająca się wartość współczynników $\varnothing_i$ powoduje zmianę pojemności buforów uwzględnianych w obliczeniach. Dla $\varnothing_i = 0$ mamy do czynienia z przypadkiem pojemności W_i. Zwiększająca się wartość $\varnothing$ powoduje zmniejszanie pojemności bufora.

Zmniejszanie pojemności bufora 4 powoduje nierozwiązywalność zadania przy średniej liczbie zgłoszeń w buforze równej 0,26. Jest to oczywiste z tego względu, że strumień 4 posiadał najwyższy priorytet.

Zmniejszanie pojemności bufora 2 do wartości 0,42 powoduje zmianę priorytetów wg kolejności:

$$1 \text{ ----- } 3 \text{ ----- } 4 \text{ ----- } 2,$$

przy następujących wartościach zmiennych x_{ij}:

$$x_{11} = 0{,}21 \qquad x_{23} = 5{,}0 \cdot 10^{-7}$$

$$x_{21} = 2{,}4 \cdot 10^{-6} \qquad x_{33} = 0{,}1$$

$$x_{31} = 8{,}4 \cdot 10^{-7} \qquad x_{43} = 8{,}9 \cdot 10^{-7}$$

$$x_{21} = 0{,}42 \qquad x_{24} = 3{,}4 \cdot 10^{-8}$$

$$x_{41} = 1{,}5 \cdot 10^{-6} \qquad x_{44} = 0{,}26$$

Trzeba zauważyć, że do rozwiązania ograniczonej pojemności buforów posłużono się wyłącznie wartościami średnimi liczby zgłoszeń w buforach. Maksymalna liczba zgłoszeń w buforze jest większa. Wysoka wartość średniej liczby zgłoszeń w buforze świadczy o jego dużym obciążeniu.

Uzyskane wyniki z rozwiązania zadania (5) należy skorygować metodami symulacyjnymi.

Przeprowadzone przez autora badania symulacyjne potwierdziły możliwość takiego rozwiązania problemu. Szczupłe ramy niniejszego opracowania nie pozwalają na dokładniejsze przedstawienie tego problemu.

LITERATURA

[1] Башарин Г.П.: Об обслуживании двух потоков с относительным приоритетом на полнодоступной системе с ограниченным числом мест для ожидания. Изв. АН СССР, Техническая Кибернетика, 1967, № 2.

[2] Башарин Г.П., Лысенкова В.Т.: Об обслуживании нескольких неоднородных потоков полнодоступным пучком с ограниченной очередью. В сб. Системы распределения информации. "Наука", 1972.

[3] Зеклеров Е.Б.: Об оптимальных абсолютных динамических приоритетах в системах массового обслуживания. Изв. АН СССР, Техническая Кибернетика, 1967, № 1.

[4] Klimow G.P.: Procesy obsługi masowej, WNT, 1979.

[5] Мова В.В., Пономаренко Н.А.: Об оптимальном назначении приоритетов зависящих от состояния обслуживающей системы с ограниченным числом мест для ожидания. Изв. АН СССР, Техническая Кибернетика, 1974, № 5.

[6] Panayiotopoulos J.C.: Solving queuing systems with increasing priority numbers. J. Oper. Res. Soc. (GB) vol. 31, no 7, pp 637-46.

[7] Рыков В.В., Лемберг Э.Е.: Об оптимальных динамических приоритетах в одноканальных системах массового обслуживания. Изв. АН СССР, Техническая кибернетика, 1967, № 1.

Recenzent: prof. dr hab. inż. Antoni NIEDERLINSKI

Wpłynęło do Redakcji 15.05.82 r.

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ПРИОРИТЕТАМИ

Р е з ю м е

Для системы массового обслуживания с несколькими потоками требований и с ограниченной очередью нужно определить правило обслуживания, зависящее от состояния системы, так чтобы не произошло переполнение буферов. Эту задачу автор решает методом определения приоритетов для потоков. Показано, что проблему можно свести к решению задачи линейного программирования.

THE SERVING OF QUEUING SYSTEMS WITH PRIORITIES

S u m m a r y

This paper considers the n queues system with single server and finite waiting rooms. For the above system the best rule of serving, depending of the state of the system, should be found. The problem is solved by the assignement of increasing priority numbers. It is shown that if interarrival distribution is exponential the task can be reduced to the linear programming problem.