Polskie Towarzystwo Badań Operacyjnych i Systemowych

Instytut Badań Systemowych Polskiej Akademii Nauk

Wojskowa Akademia Techniczna

Redaktorzy:

Zbigniew Nahorski

Marian Chudy

Andrzej Straszak

OPTYMALIZACJA

Warszawa 1991

POLSKIE TOWARZYSTWO
BADAŃ OPERACYJNYCH I SYSTEMOWYCH

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

WOJSKOWA AKADEMIA TECHNICZNA

OPTYMALIZACJA

ZADANIA, METODY, ALGORYTMY

Redaktorzy

Zbigniew Nahorski, Marian Chudy, Andrzej Straszak

WARSZAWA 1991

OPTYMALIZACJA:
Zadania, metody, algorytmy
PTBOiS, Warszawa 1991.

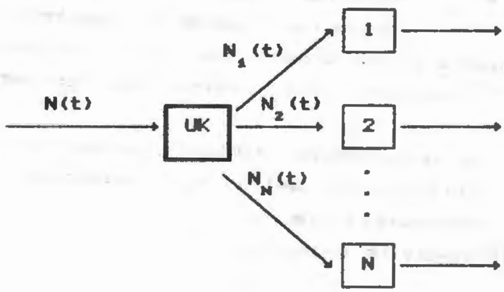
WYZNACZANIE ILOŚCI KOMPUTERÓW DO REALIZACJI ZADANEGO CIĄGU ZADAŃ OBLICZENIOWYCH

Marian Chudy
Wojskowa Akademia Techniczna. Warszawa

Streszczenie: Wyznaczono oszacowanie granicznych charakterystyk kosztowych systemu składającego się z N komputerów, realizującego zadania obliczeniowe o losowej długości obliczeń i napływające w losowych odstępach czasowych. Sformułowano zadanie wyznaczenia ilości komputerów minimalizującej oczekiwane koszty jednostkowe.

1. Wstęp.

Dany jest system (rys.1) składający się z N komputerów przeznaczonych do realizacji ciągu zadań obliczeniowych.



rys.1.

Napływające w losowych odstępach czasowych zadania obliczeniowe są przydzielane przez urządzenie kierujące (UK) komputerom według zasady, że kolejne zadanie jest przydzielone komputerowi o numerze o jeden większym od numeru komputera, do którego przydzielono zadanie poprzednie - gdy ten numer jest mniejszy od N; w przeciwnym przypadku przydzielone jest komputerowi o numerze 1.

Jeśli komputer, do którego przydzielono zadanie jest zajęty realizacją poprzednich zadań, to aktualnie przydzielone zadanie wchodzi do kolejki typu FIFO przy tym komputerze. Wprowadzimy następujące oznaczenia:

t_i - zmienna losowa oznaczająca chwilę napływu do UK i-tego zadania obliczeniowego (zamówienia),

$t_{i+1} - t_i = x_i$ - odległość czasowa między (i+1) a i-tym zamówieniem.

Założymy, że

1° (x_i) - ciąg niezależnych zmiennych losowych o jednakowym rozkładzie i skończoną wartością oczekiwaną równą a.

Dalsze oznaczenia

N(t) - ilość zamówień, które napłyneły do UK do chwili t,

t_k^n - chwila pojawienia się w n-tym komputerze k-tego zadania obliczeniowego,

gdzie

$$t_k^n = t_i \ , \ i = (k-1)N+n \ ,$$

$t_{k+1}^n - t_k^n = x_k^n$ - odległośc czasowa między (k+1)-szym a k-tym zamówieniem do komputera o numerze n,

η_k^n - zmienna losowa oznaczająca czas wykonania przez n-ty komputer k-tego zadania obliczeniowego.

Założymy, że

2° (η_k^n) - ciąg niezależnych zmiennych losowych o jednakowym rozkładzie dla $n=\overline{1,N}$ oraz skończonej wartości oczekiwanej równej b.

Z zasady działania UK wynika,że

$$x_k^n = \sum_{i=(k-1)N+n+1}^{k\cdot N+n} x_i \tag{1}$$

Zatem (x_k^n) - ciąg niezależnych zmiennych losowych o jednakowym rozkładzie, dla $n=\overline{1,N}$.

Niech ponadto

c_o - opłata za realizację zadania obliczeniowego przez

Uwzględniając (1) oraz założenie 1°otrzymamy z podstawowego twierdzenia teorii odnowy

$$\lim_{t\to\infty} \frac{N_n(t)}{t} \cdot \frac{1}{1} = \frac{1}{N \cdot a} \tag{6}$$

gdyż

$$Ex_k^n = N \cdot a$$

Uwzględniając założenie (2), z silnego prawa wielkich liczb [1] mamy

$$\lim_{t\to\infty} \frac{c_o}{N_n(t)} \sum_{k=1}^{N_n(t)} \eta_k^n = c_o \cdot b. \tag{7}$$

Z (6),(7) oraz twierdzenia 1 [3] dotyczącego granicy z p-tem 1 iloczynu procesów otrzymamy

$$\lim_{t\to\infty} \frac{K_o(t,N)}{t} \cdot \frac{1}{1} = \sum_{n=1}^{N} \frac{c_o \cdot b}{N \cdot a} = \frac{c_o \cdot b}{a} \tag{8}$$

Jak widać (8) nie zależy od N. Nie jest to wynik zaskakujący gdyż z przyjętej reguły działania UK oraz kolejki typu FIFO wynika, że niezależnie na ilu komputerach obsłużymy wpływający strumień zadań obliczeniowych opłata na jednostkę czasu za obliczenia będzie w granicy taka sama. Znacznie trudniejsze jest wyznaczenie granicy z p-twem 1 wyrażenia (5).

Ostatecznie, tej granicy nie wyznaczymy a tylko jej oszacowanie.

Rozpatrzymy w tym celu wyrażenie określające W_{k+1}^n dla pewnego $n \in \{1,2...,N\}$

$$w_{k+1}^n = \max\left\{o, w_k^n + \eta_k^n - x_k^n\right\} = \max\left\{o, w_k^n + \xi_k^n\right\} \tag{9}$$

gdzie

$$\xi_k^n = \eta_k^n - x_k^n$$

komputer w jednostce czasu (zakładamy, że komputery są jednorodne),

c_v - koszt wypłacony zamówieniu (dokładniej właścicielowi zadania obliczeniowego) za oczekiwanie przez jednostkę czasu na realizację obliczenia.

Przy pomocy wprowadzonych wielkości będziemy mogli określić interesujące nas charakterystyki kosztowe systemu.

2. Charakterystyki kosztowe systemu.

Całkowita opłata za wykonanie zadań obliczeniowych, które napłynęły do systemu do chwili t będzie wynosiła.

$$K_o(t,N) = \sum_{n=1}^{N} \sum_{k=1}^{N_n(t)} c_o \eta_k^n \tag{2}$$

gdzie

$N_n(t)$ - ilość zadań obliczeniowych, które napłynęły do n-tego komputera do chwili t.

Całkowity koszt oczekiwania zadań, które napłynęły do systemu (UK) do chwili t będzie wynosił

$$K_v(t,N) = \sum_{n=1}^{N} \sum_{k=1}^{N_n(t)} c_v w_k^n \tag{3}$$

gdzie

w_k^n - czas oczekiwania k-tego zadania obliczeniowego na rozpoczęcie obliczeń w n-tym komputerze.

Interesować nas będą granice lub ich oszacowanie następujących wielkości przy $t \to \infty$

$$\frac{K_o(t,N)}{t} = \frac{1}{t} \sum_{n=1}^{N} \sum_{k=1}^{N_n(t)} c_o \eta_k^u \tag{4}$$

oraz

$$\frac{K_w(t,N)}{t} = \frac{1}{t} \sum_{n=1}^{N} \sum_{k=1}^{N_n(t)} c_v w_k^u \tag{5}$$

Przyjmując $w_1^n = 0$ można wykazać, że

$$w_k^n = \max\left\{0, \xi_k^n, \xi_k^n+\xi_{k-1}^n, \ldots, \xi_k^n+ \ldots +\xi_2^n\right\} \tag{10}$$

Ciąg (w_k^n) zdefiniowany wyrażeniem (10) nie jest stacjonarny w wąskim sensie, co uniemożliwia zastosowanie twierdzenia ergodycznego.
Z twierdzenia Kołmogorowa [1] wynika natomiast, że (ξ_k^n) można dopełnić do postaci

$$\left\{\xi_k^n\right\}, \quad -\infty < k < \infty \quad , \quad n = \overline{1,N} \tag{11}$$

Rozpatrzmy ciąg zmiennych losowych postaci

$$\bar{w}_k^n = \sup\left\{0, \xi_k^n, \xi_k^n+\xi_{k-1}^n, \xi_k^n+\xi_{k-1}^n+\xi_{k-2}^n, \ldots\right\} \quad -\infty < k < \infty \tag{12}$$

Dla każdego zdarzenia elementarnego zachodzi

$$w_k^n \le \bar{w}_k^n \quad , \ k \ge 1 \ , \ n = \overline{1,N} \tag{13}$$

Jest to ważne dla dalszych rozważań spostrzeżenie.
Jak pokazano w [2], ciąg $(\bar{w}_k^n)$ jest stacjonarny w wąskim sensie, a ponadto jeśli $E\xi_k^n < 0$, to $\bar{w}_k^n$ są skończone z p-twem 1. Można więc zastosować dla ciągu $(\bar{w}^n)$, $k \ge 1$ twierdzenie ergodyczne [4] w rezultacie czego otrzymamy

$$\lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \bar{w}_k^n \stackrel{1}{=} w(N), \ Ew(N) = E\bar{w}_1^n \tag{14}$$

Z podstawowego twierdzenia odnowy oraz (13) i (14) otrzymujemy

$$\lim_{t\to\infty} \frac{K_v(t,N)}{t} = \lim_{t\to\infty} \sum_{n=1}^{N} \frac{N_n(t)}{t} \cdot \frac{c_v}{N_n(t)} \sum_{k=1}^{N_n(t)} w_k^n \le \tag{15}$$

$$\le \lim_{t\to\infty} \sum_{n=1}^{N} \frac{N_n(t)}{t} \cdot \frac{c_v}{N_n(t)} \sum_{k=1}^{N_n(t)} \bar{w}_k^n \stackrel{1}{=} \frac{c_v \cdot w(N)}{a}$$

stąd

$$E\left[\lim_{t\to\infty} \frac{K_v(t,N)}{t}\right] \le \frac{c_v Ew(N)}{a} = \frac{c_v E\bar{w}^n_1}{a} \tag{16}$$

Jednocześnie z [2] wiemy, że przy $k\to\infty$ rozkład zmiennej losowej w^n_k dąży do rozkładu zmiennej losowej $\bar{w}^n_1$. Zatem aby obliczyć $E\bar{w}^n_1$ należy obliczyć wartość oczekiwaną zmiennej losowej o rozkładzie granicznym

$$P\{ w^n(N) < y \} = \lim P\{ w^n_k < y \} \tag{17}$$

czyli

$$Ew^n(N) = E\bar{w}^n_1 = Ew(N)$$

Wartość $EW^n(N)$ jest funkcją zmiennej N (ilości komputerów w systemie) i można ją obliczyć wykorzystując znane modele SMO z oczekiwaniem dla różnych rozkładów zmiennych losowych x_i oraz η^n_k.

3. Sformułowanie zadania wyznaczania ilości komputerów minimalizującej koszty jednostkowe systemu.

Oznaczamy przez:

C – koszt eksploatacji komputera w jednostce czasu.

Oszacowanie wielkości zysku na jednostkę czasu wynosi

$$\frac{C_o \cdot b}{a} - \left[\frac{C_w E\bar{w}^n_1}{a} + N \cdot C\right] \tag{18}$$

Przy przyjętych założeniach i regułach realizacji zamówień zadanie wyzanczenia optymalnej ilości komputerów do realizacji zadań obliczeniowych można sformułować następująco.

$$\max\left[\frac{C_o \cdot b}{a} - \frac{C_o \cdot Ew(N)}{a} - N \cdot C\right] \tag{19}$$

przy ogr.

(20) $b - N \cdot a < 0$

(21) $N \geq 0$, N – całkowita

gdzie ograniczenie (20) wynika z warunku $E\,\zeta_k^n < 0$.
Powyższe zadanie jest zadaniem programowania dyskretnego, które przyjmie postać zależną od postaci wyrażenia EW(N). Należy zauważyć, że uzyskane rozwiązanie tego zadania będzie oszacowaniem optymalnej ilości komputerów w sytemie.

Literatura.

[1] Боровков А.А.: Курс теории вероятностей. Москва 1972.
[2] Боровков А.А.: Вероятностные процессы в терии массового обслуюування. Москва 1972
[3] Chudy M.: Zależności między charakterystykami procesów uzupełniania zasobów. Biul.WAT Nr.1/91
[4] Коваленко И.Н., Кузнецов Н.Н., Шуренков В.М.: Случайные процессы. Куев 1983..

ISBN 83-900412-1-9