ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1980
Seria: AUTOMATYKA z. 54 Nr kol. 650

Józef Goetz
Politechnika Wrocławska

MODELOWANIE NA GRAFIE DYSJUNKTYWNYM OGÓLNYCH ZAGADNIEŃ PRZYDZIAŁU ZASOBÓW I SZEREGOWANIA OPERACJI

Streszczenie. W pracy przedstawia się sformułowanie i model matematyczny ogólniejszych zagadnień przydziału zasobów i szeregowania operacji niż w literaturze. Sformułowanie to jest przedstawione dla systemu komputerowego, który jest ogólniejszy niż dyskretny system produkcyjny. Przedstawia się interpretacje, niezależnie od rodzaju systemu, na grafie dysjunktywnym klasy kryteriów opierających się na żądanych terminach wykonania zleceń.

1. Wprowadzenie

Do projektowania współczesnych systemów operacyjnych SO używa się metod przybliżonych opartych o reguły heurystyczne. Przy wyznaczaniu priorytetów nie uwzględnia się aktualnego i przyszłego środowiska przetwarzanych programów, wymagań tych programów, konfiguracji i zbioru zasobów systemu komputerowego w sposób globalny. Dalszy wzrost efektywności obsługi użytkowników wymaga opracowania ścisłych procedur optymalizacyjnych, uwzględniających możliwie globalnie parametry systemu komputerowego. Temu celowi m. in. jest podporządkowana ta praca.

Wyznaczanie priorytetów wiąże się z zagadnieniami kolejnościowymi, które są centralnymi zagadnieniami projektowania SO [1,2]. Można wyróżnić dwa współzależne zasadnicze zagadnienia:

a) przydziału zasobów, tj. wyboru dla każdej operacji podzbioru zasobów podstawowych /zasobu złożonego/ z rodziny zasobów dopuszczalnych,

b) uszeregowania operacji na każdym przydzielonym zasobie podstawowym, tj. wyboru kolejności wykonania operacji.

Wynikiem rozwiązania tych zagadnień jest taka kolejność wykonywania operacji, że zabezpieczona jest synchronizacja wykonywania operacji, brak jest blokad oraz jest takie wykorzystanie pamięci, procesorów, urządzeń we/wy itp., że spełnione jest określone kryterium jakości funkcjonowania systemu [2].

Dotychczasowe rozwiązania optymalne dotyczą co najwyżej wybranych, i to znacznie uproszczonych, modeli podsystemów systemu komputerowego, np. [9,10], gdzie wyróżnia się tylko procesory. Dla takich modeli znajduje się rozwiązanie, korzystając z klasycznych modeli zagadnienia: n detali uszeregować na m maszynach, gdzie nie bierze się pod uwagę powiązań między maszy-

nami bądź czasu pracy SO. Dalsze uproszczenia współczesnych modeli, takie jak brak wyróżnienia innych rodzajów zasobów systemu komputerowego oraz ograniczeń na poszczególne zasoby, spowodowały, że modele te w niewielkim stopniu przybliżają działanie współczesnych systemów komputerowych. Dlatego w pracy rozważania są oparte na oryginalnym modelu systemu komputerowego SK [5,6], w którym pominięte są wymienione uproszczenia.

Zagadnienia przydziału zasobów i szeregowania operacji formułowane na tym modelu wydają się być ogólniejsze od zagadnień kolejnościowych, formułowanych dla systemów produkcyjnych rozumianych jak w [7,8]. Wydaje się, że system komputerowy jest swego rodzaju też systemem produkcyjnym, który obrabia /produkuje wyniki/ na dostarczonych do niego danych. Wiele można znaleźć analogii między dyskretnym systemem produkcyjnym a systemem komputerowym /np. oba systemy są dyskretne/. W szczególności warto zwrócić uwagę, że system komputerowy może być jednym z elementów ogólnego systemu produkcyjnego. Dlatego zagadnienia poruszane w pracy nie mają ograniczenia tylko do systemów komputerowych.

Rozszerzając konwencję symbolicznego zapisu zagadnień kolejnościowych w [8,11] możemy przedstawić klasę sformułowanych zagadnień w pracy w postaci $n|m,G_m|J|m_h \subset m, D_h, r_j \geq 0, d_j \geq 0, RT|\Sigma f_j$, gdzie: n - zbiór zadań /zleceń/, m - zbiór maszyn, których połączenia tworzą graf G_m, J - uszeregowanie maszyn (dla każdego zlecenia może być dowolne) m_h - podzbiór maszyn przydzielany do wykonania operacji, D_h - zbiór danych systemowych potrzebny w m_h, symbol r_j orzeka o obecności terminu rozpoczęcia wykonywania poszczególnych zadań /operacji/, a d_j - o obecności żądanych terminów wykonania zadań, RT - częściowe uporządkowanie operacji wchodzących w skład każdego n, Σf_j - postać kryterium, gdzie $f_j(t)$ funkcja niemalejąca od t.

2. Model systemu

Ze względu na ograniczoną objętość pracy zostanie przedstawiony zarys modelu.

Przewodnią myślą konstrukcji modelu SK było takie spojrzenie na ten system, aby można go było przedstawić jako zbiór zasobów, które powinny być rozdzielone pomiędzy napływające zgłoszenia od użytkowników. System komputerowy /SK/ określamy trójką [5]: SK = (SR, SZ, SO), gdzie SR - system zasobów, SZ - system zleceń, SO - system operacyjny. W SK wyróżnia się trzy klasy zasobów podstawowych:

a) procesory /P/ - tj. szeroko rozumiane urządzenia fizyczne zdolne do przenoszenia i przetwarzania informacji,

b) pamięci - tj. urządzenia fizycznie zdolne do przechowywania informacji w czasie; pamięć nie istnieje w systemie samodzielnie, lecz związana jest z procesorem sterującym - zatem można mówić o procesorach bez i z pamięciami.

c/ dane systemowe /D/ - tj. zbiór wszystkich programów i danych do tych programów przechowywanych w pamięciach SK. Obszary nie zajęte przez dane systemowe są nazywane danymi wolnymi.

Proces przetwarzania danych dzieli się na pewne etapy, które dla swojego przebiegu wymagają niezmiennego zestawu zasobów podstawowych na pewien odcinek czasu. Zakres czynności o takiej właściwości określany jest mianem operacji, a zestaw zasobów podstawowych wymagany dla realizacji tej operacji zasobem złożonym.

Zasobem złożonym będziemy nazywać trójkę:

$$B = (P', D', D''), \qquad /1/$$

gdzie: $P' \subset P$, $D' \subset D$, $D'' \subset \bar{D}$ /zbiór danych wolnych/, jeżeli będą spełnione warunki:

i/ $P', D' \neq \emptyset$,

ii/ obcięcie relacji R_{pp} do zbioru $P' \times P'$ stanowi relację spójną,

iii/ D' przechowywany w pamięciach P',

iv/ objętość D'' nie przekracza wolnych obszarów pamięci procesorów.

Zbiór zasobów złożonych będziemy nazywać systemem zasobów SR, jeżeli dla każdych $p,q \in P$ takich, że $\langle p,q \rangle \in R_{pp}$ istnieje zasób złożony B należący do tego zbioru, który zawiera p oraz q /gdzie $R_{pp} \subset P \times P$ jest relacją połączeń/.

Wynikiem interpretacji zgłoszenia do SK jest zlecenie. Zleceniem będziemy nazywać trójkę: $J = (M, R_J, W)$, gdzie $R_J \subset M \times M$ jest relacją porządku technologicznego, tzn. $\langle i,j \rangle \in R_J$ iff wyniki operacji i są jednym z argumentów operacji j; $W \subset M \times SR \times \mathbb{R}^+$ - relacją wymagań technologicznych, tzn. $\langle j, B, c \rangle \in W$ iff $j \in M$ może być zrealizowane za pomocą $B \in SR$ w odcinku czasu $c \in \mathbb{R}^+$.

Systemem zleceń SZ będziemy nazywać klasą zleceń, które spełniają warunki realizowalności w danym SK. Zlecenie J jest realizowalne, jeżeli spełnione są warunki:

/i/ dla $\forall j \in M$ jest przydzielony jeden zasób złożony $B_j = (P_j, D'_j, D''_j) \in SR$ na odcinek czasu $[t^a_j, t^b_j]$,

/ii/ $\langle j, B_j, t^b_j - t^a_j \rangle \in W$, $j \in M$,

/iii/ jeżeli $\langle i,j \rangle \in R_J$, to $t^b_i \leq t^a_j$ $(i \neq j,\ i,j \in M)$,

/iv/ jeżeli $\langle i,j \rangle \in R_J$ $(i \neq j,\ i,j \in M)$, to pamięci procesorów $P_i \cap P_j$ zawierają tę część danych systemowych, które operacja i ma przekazać operacji j,

/v/ różne operacje realizowane w tym samym momencie czasu wymagają przydziału wzajemnie niezależnych zasobów złożonych.

W działaniu SO można wyróżnić dwie fazy [5]:

/i/ fazę akcji standardowej - rejestracja, identyfikacja przerwania oraz ewentualnie przydział zasobów złożonych dla następnej fazy,

/ii/ fazę akcji decyzyjnej - przydział zasobów i uszeregowanie operacji.

Faza /i/ jest oparta zasadniczo o realizację układową /układ przerwań/,

faza /ii/ jest w zasadniczej swej części algorytmem optymalizacyjnym i jest traktowana jako dana systemowa. Przedmiotem pracy jest przedstawienie sformułowania modelu matematycznego dla działania fazy /ii/. Zalety metody grafów dysjunktywnych [4, 7, 8] spowodowały, że została ona wykorzystana do tego celu.

3. Sformułowanie zagadnień optymalizacyjnych

Dla SK globalnym zleceniem jest układ operacji, w skład którego wchodzi zbiór operacji N wykonywanych i tych, które mają być wykonane. Zbiór N pochodzi od aktualnie wykonywanego zbioru zleceń K.

Układem operacji będziemy nazywać trójkę:

$$UO = \langle N, RT, W \rangle , \qquad /2/$$

gdzie $RT \subset N \times N$, a relacja $W \subset N \times SR \times \mathcal{R}^+$ jest relacją wymagań technologicznych związaną ze wszystkimi zleceniami zbioru K. Oznaczmy graf układu operacji przez $G_N = \langle N, RT \rangle$. Układem operacji bez przestrojeń będziemy nazywać taki układ operacji, dla którego są spełnione warunki realizowalności. Z każdym zleceniem /zgłoszeniem/ będzie zadane wymaganie realizacyjne, tj. żądany termin d_k wykonania zlecenia k, najwcześniejszy możliwy termin r_k zgłoszenia /rozpoczęcia/ zlecenia k oraz mogą być zadane najwcześniejsze terminy r_j rozpoczęcia wykonywania zadanych operacji zlecenia $k \in K$. Uwarunkowany czasowo układ operacji bez przestrojeń ma dołączone wymagania realizacyjne.

Niech $$B(j) = \{ B_h \in SR \mid \langle j, B_h, c_{jh} \rangle \in W,\ c_{jh} \in \mathcal{R}^+ \} , \quad j \in N \qquad /3/$$

będzie zbiorem dopuszczalnych zasobów złożonych dla $j \in N$, z których jeden zostanie przydzielony do wykonania tej operacji. Przydzielenie zasobu złożonego dla operacji $j \in N$ polega na wyborze ze zbioru zmiennych decyzyjnych $X_j = \{x_{j1}, x_{j2}, \ldots, x_{j|B(j)|}\}$, $j \in N$ jednej zmiennej decyzyjnej i nadaniu jej wartości $x_{jh}=1$, gdzie

$$x_{jh} = \begin{cases} 1, \text{ jeżeli dla wykonania operacji } j \text{ jest przydzielony } B_h\text{-ty zasób złożony oraz wynikiem są zbiory danych systemowych } D''_{jh} \text{ oraz wolnych } D'_{jh}, \\ 0, \text{ w przeciwnym przypadku} \end{cases} \quad B_h \in B(j),\ j \in N. \qquad /4/$$

Przydzielenie zasobu złożonego B_h ze zbioru $B(j)$ na odcinku czasu c_{jh} będziemy nazywać reprezentacją zbioru $B(j)$.

Niech $$B(i,j) = \{ \langle B_s, B_h \rangle \in B(i) \times B(j) \mid P_s \cap P_h \neq \emptyset \},\quad i,j \in N,\quad i \neq j \qquad /5/$$

będzie zbiorem możliwych par zależnych zasobów złożonych dla wykonania operacji i oraz j, gdzie: $B_s = (P_s, D'_s, D''_s)$, $B_h = (P_h, D'_h, D''_h)$.

W SK w każdej chwili po wykonaniu operacji $i \in \Gamma^{-1} j$ /zbiór poprzedników j/ w grafie G_N jest określony podzbiór danych systemowych, który ma zostać

przekazany operacji $j \in N$ do jej wykonania. Niech ten podzbiór danych jest przechowywany w pamięciach zasobu złożonego $B_s = (P_s, D'_s, D''_s) \in B(i)$ dostępnego przez zbiór procesorów $P_s^j \subset P_s$ oraz niech

$$B_{is}(j) = \{B_h \in B(j) \mid x_{is} = 1 \wedge P_s^j \subset P_h\}, \quad B_s \in B(i), \quad i \in \Gamma^{-1}j, \quad j \in N \qquad /6/$$

będzie zbiorem dopuszczalnych zasobów dla operacji $j \in N$, gdy jest przydzielony operacji $i \in \Gamma^{-1}j$ zasób złożony $B_s \in B(i)$.

Przydzielenie /wybór/ zasobów złożonych po jednym z każdego zbioru $B(j) \in SR$ dla wszystkich $j \in N$ na odpowiadające odcinki czasu będziemy nazywać <u>reprezentacją S^o zasobów złożonych</u> przydzielonych do wykonania układu operacji. Niech

$$\mathcal{R}_S^o = \{S_1^o, \ldots, S_r^o, \ldots, S_a^o\} \qquad /7/$$

będzie rodziną wszystkich takich reprezentacji oraz niech $S_r^o(j) \subset S_r^o$ jest reprezentacją zasobów złożonych dla wszystkich generacji $i \in \Gamma^{-1}j$ w grafie G_N. Niech

$$B_r(j) = \left\{B_h \in \bigcap_{i \in \Gamma^{-1}j} B_{is_i}(j) \,\middle|\, S_r^o(j) \in \mathcal{R}_S^o\right\}, \quad S_r^o \in \mathcal{R}_S^o, \quad j \in N \qquad /8/$$

będzie podzbiorem dopuszczalnych zasobów złożonych dla operacji j, gdy została wybrana reprezentacja $S_r^o(j)$ zasobów złożonych dla wszystkich $i \in \Gamma^{-1}j$ w grafie G_N.

Wprowadzimy jeszcze następujące pojęcia i oznaczenia:

t_j^a, t_j^b - odpowiednio, termin rozpoczęcia i zakończenia wykonywania operacji j;

c_{jh} - czas przydzielony dla wykonania operacji j na B_h-tym zasobie złożonym, $B_h \in B(j)$;

$t_o^a = t_o^b = t_o$ - termin, od którego liczy się rozpoczęcie wykonywania układu operacji;

$E = \bigcup_{k \in K}\{e_k\}$ - zbiór operacji fikcyjnych;

$C_k = t_{e_k}^a$ - termin rozpoczęcia wykonywania operacji fikcyjnej e_k, jest on równocześnie terminem zakończenia k-tego zlecenia, $k \in K$;

$L_k = t_{e_k}^b$ - algebraiczne spóźnienie wykonania k-tego zlecenia od żądanego terminu wykonania d_k;

cd_k - czas wykonywania operacji fikcyjnej e_k, $cd_k = -d_k \leq 0$, $k \in K$;

$\Theta^k = \max\{0, C_k - d_k\}$ - spóźnienie wykonania k-tego zlecenia od żądanego terminu wykonania d_k;

$\Theta = \sum_{k \in K} \Theta^k$ - spóźnienie wykonania uwarunkowanego czasowo układu operacji;

N^k - zbiór operacji k-tego zlecenia, $N^k \subset N$;

$N_o^k = \{j \in N^k \mid \forall i \in N^k \wedge \langle i, j\rangle \notin RT\}$ - zbiór operacji k-tego zlecenia nie posiadających poprzedników.

Niech $\Phi : \hat{N} \to \mathcal{R}^+ \times \mathcal{R}^+ \times [SR \cup \{0\}]$ jest funkcją taką, że dla $j \in N$ wartościami funkcji są: $\Phi(j) = \langle t_j^a, t_j^b, B_h\rangle \in \mathcal{R}^+ \times \mathcal{R}^+ \times B(j)$, zaś dla opera-

cji $e_k \in E$ wartościami są: $\Phi(e_k) = \langle C_k, \theta^k, B_k \rangle \in \mathcal{R}^+ \times \mathcal{R}^+ \times \{0\}$ /gdzie $\mathcal{R}^+$ - zbiór dodatnich liczb rzeczywistych z dołączonym zerem/.

Definicja 1. Funkcję Φ odwzorującą rozszerzony zbiór operacji $\hat{N} = N \cup E$ w produkt $\mathcal{R}^+ \times \mathcal{R}^+ \times [SR \cup \{0\}]$ będziemy nazywać funkcją rozdziału dla uwarunkowanego czasowo układu operacji bez przestrojeń, jeżeli spełnione są warunki:

$$t_j^b - t_j^a \geqslant \sum_{B_h \in B(j)} x_{jh} \cdot c_{jh}, \quad j \in N, \qquad /w.1/$$

$$\sum_{B_h \in B(j)} x_{jh} = 1, \quad j \in N, \qquad /w.2/$$

$$x_{jh} \in \{0,1\}, \quad B_h \in B(j), \quad j \in N, \qquad /w.3/$$

$$\bigcap_{i \in \Gamma^{-1} j} (x_{is_i} = 1) \Rightarrow \left[\sum_{B_h \in B_r(j)} x_{jh} = 1\right], \quad S_r^o \in \mathcal{R}_S^*, \quad j \in N, \qquad /w.4/$$

$$t_j^a, t_j^b, C_k \geqslant 0, \quad j \in N, \quad k \in K, \qquad /w.5/$$

$$t_j^a - t_i^b \geqslant 0, \quad \langle i,j \rangle \in RT, \qquad /w.6/$$

$$t_j^a - t_o \geqslant r_j, \quad j \in N, \qquad /w.7/$$

$$t_j^a - t_o \geqslant r_k, \quad j \in N_o^k, \quad k \in K, \qquad /w.7a/$$

$$(x_{is} = 1) \wedge (x_{jh} = 1) \Rightarrow (t_j^a - t_i^b \geqslant 0) \vee (t_i^a - t_j^b \geqslant 0), \quad \langle B_s, B_h \rangle \in B(i,j), \quad i,j \in N, \ i \neq j, \qquad /w.8/$$

$$t_{e_k}^b = C_k + cd_k, \quad k \in K, \qquad /w.9/$$

$$\theta^k - t_{e_k}^b \geqslant 0, \quad k \in K, \qquad /w.10/$$

$$\theta^k \geqslant 0, \quad k \in K. \qquad /w.11/$$

Warunki /w.1/ - /w.11/ posiadają odpowiednią interpretację fizyczną. Warunek /w.2/ jest dysjunkcją wieloczłonową i wyraża, że operacji $j \in N$ może być przydzielony tylko jeden zasób złożony ze zbioru dopuszczalnych dla tej operacji. Ograniczenie /w.1/ przy uwzględnieniu warunków /w.2/ i /w.3/ wyraża warunek, że różnica między terminem zakończenia i rozpoczęcia operacji $j \in N$ nie może być mniejsza niż czas, na jaki został przydzielony zasób złożony do wykonywania tej operacji. Warunek /w.4/ oznacza, że jeżeli dla wszystkich operacji $i \in \Gamma^{-1} j$ zostały przydzielone zasoby złożone $B_{s_i} \in B(i)$, to dla wykonania operacji j musi pozostać dopuszczalny co najmniej jeden zasób złożony dla jej wykonania. Warunek /w.4/ jest związany z ograniczeniem zasobów dopuszczalnych dla operacji j do podzbioru $B_r(j) \subset B(j)$, - wynika to z warunku /iv/ p.2 realizowalności zleceń. Warunek /w.6/zapewnia zachowanie kolejności wykonania układu operacji bez przestrojeń. Warunek /w.7/ i /w.7a/ zapewnia wykonanie odpowiednich operacji nie wcześniej

niż dane terminy r_j i r_k. Warunek /w.8/ narzuca szeregową kolejność wykonania operacji i oraz j w przypadku gdy dla wykonania tych operacji przydzielone są zależne zasoby złożone. Z równości /w.9/ wynika, że $t^b_{e_k} \in \mathcal{R}$, $k \in K$. Ograniczenia /w.9/ i /w.10/ podają warunki, jakie musi spełniać wielkość spóźnienia dla każdego zlecenia i układu operacji. Warto zauważyć, że ograniczenia /w.1/ - /w.3/, /w.4/, /w.6/ i /w.8/ są równoważne odpowiednio warunkom /i/ - /ii/, /iv/, /iii/ i /v/ p.2 realizowalności zleceń.

Funkcja rozdziału Φ z ograniczeniami /w.1/ - /w.11/ jest równoważna określeniu funkcji SO działania systemu operacyjnego dla uwarunkowanego czasowo układu operacji bez przestrojeń dla fazy akcji decyzyjnej [4,5].

Oznaczmy przez $(\bar{t},\bar{x})$ rozwiązanie dopuszczalne ogólnego zagadnienia P przydziału zasobów i szeregowania operacji, w którym $\bar{t}$ jest wektorem wartości czasów t^a_j, t^b_j, $j \in N$, Θ^k, $k \in K$ oraz $\bar{x}$ jest wektorem wartości zmiennych x_{jh}, $B_h \in B(j)$, $j \in N$. Celem ogólnego zagadnienia P jest znalezienie $(\bar{t},\bar{x})_o$ minimalizującego $\sum_{k \in K} f_k(\Theta^k)$ przy ograniczeniach /w.1/ - /w.11/, zatem opracowanie algorytmu SO_o dla fazy akcji decyzyjnej systemu operacyjnego. Funkcja f_k może być interpretowana jako funkcja reprezentująca koszty związane ze spóźnieniem wykonywania zleceń.

Postępując tak samo jak w pracy [4] możemy zagadnieniu P przyporządkować ogólny graf dysjunktywny

$$D^o = \langle \bar{N}, U, V^o, V \rangle , \qquad /9/$$

gdzie $G = \langle \bar{N}, U \rangle$ jest grafem bezkonturowym, w którym zbiorem węzłów jest

$$\bar{N} = A \cup B \cup \bar{E} \cup \{0\}, \qquad /10/$$

$$A = \bigcup_{j \in N} \{j^a\}, \quad B = \bigcup_{j \in N} \{j^b\}, \quad \bar{E} = \bigcup_{k \in K} \{e^a_k, e^b_k\},$$

gdzie węzły j^a, j^b odpowiadają zmiennym t^a_j, t^b_j. Węzeł 0 odpowiada zmiennej t_o, a węzły e^a_k, e^b_k odpowiednio zmiennym C_k oraz Θ^k. Zbiorem łuków ustalonych w grafie D^o jest

$$U = U_1 \cup U_2 \cup U_3 \cup U_4, \qquad /11/$$

przy czym $U_1 = \bigcup_{\langle i,j \rangle \in RT} \{\langle i^b, j^a \rangle\}$, $U_2 = \bigcup_{k \in K} \{\langle e^a_k, e^b_k \rangle\}$,

$U_3 = \bigcup_{k \in K} \{\langle 0, e^b_k \rangle\}$, $U_4 = \bigcup_{j \in N} \{\langle 0, j^a \rangle\}$.

Łukom $\langle i^b, j^a \rangle \in U_1$ odpowiada ograniczenie /w.6/ i ich obciążenie wynosi zero. Łuki $\langle e^a_k, e^b_k \rangle \in U_2$ przedstawiają relację między rozpoczęciem i zakończeniem operacji fikcyjnej e_k wyrażoną ograniczeniem /w.9/, stąd każdemu z łuków U_2 przyporządkujemy liczbę $cd_k = -d_k \leq 0$, $k \in K$. Łuki $\langle 0, e^b_k \rangle \in U_3$ odpowiadają ograniczeniu /w.10/, stąd ich obciążenie wynosi zero. Łuki ze zbioru U_3 będziemy nazywać łukami fikcyjnymi. Pozwalają one uzyskać drogę krytyczną od 0 do e^b_k równą zero w przypadku gdy $C_k - d_k < 0$, $k \in K$. Zatem wprowadzenie łuków $\langle e^a_k, e^b_k \rangle$ i $\langle 0, e^b_k \rangle$ pozwala wyznaczyć bezpośrednio na grafie wartość spóźnienia $\Theta^k = \max\{0, C_k - d_k\} = \max\{0, L_k\}$, $k \in K$. Łuki $\langle 0, j^a \rangle \in U_4$ odpowiadają ograniczeniu /w.7/ i ich obciążalność wynosi r_j.

Ograniczenia /w.2/ i /w.3/ możemy zastąpić równoważnym ograniczeniem

w postaci dysjunkcji rozłącznej

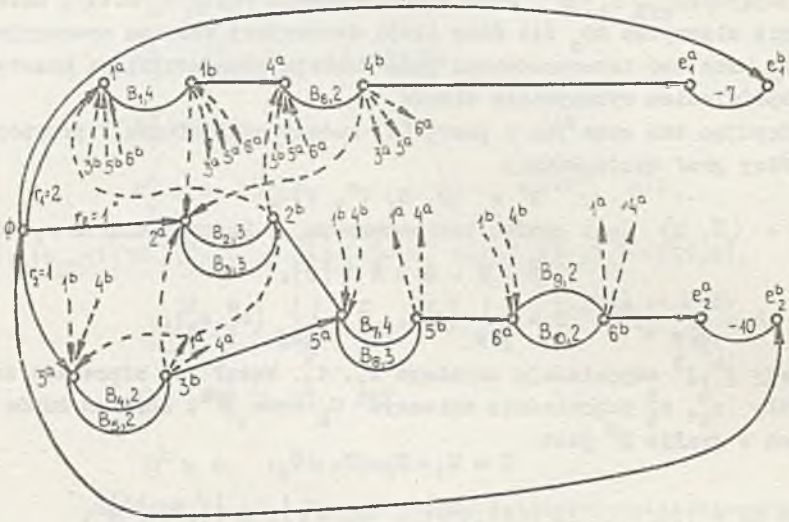
$$\left(x_{j1} = 1\right) \vee \left(x_{j2} = 1\right) \vee \ldots \vee \left(x_{jv_j} = 1\right), \quad j \in N \qquad /12/$$

gdzie $|B(j)| = v_j$. Zatem jeżeli zmienna $x_{jh_j} = 1$, to pozostałe zmienne będą równe 0 i ograniczenie /w.1/ przyjmie postać:

$$t_j^b - t_j^a \geqslant c_{jh_j}, \quad j \in N. \qquad /13/$$

Każdej zmiennej x_{jh}, która przyjmuje wartość równą jeden, możemy przyporządkować dokładnie jedno ograniczenie w postaci /13/. Ograniczeniu /13/ możemy przyporządkować łuk z węzłem początkowym j^a i końcowym j^b oraz obciążeniem c_{jh}. Zatem ograniczeniu /12/ możemy przyporządkować dysjunkcję v_j elementową łuków o różnych obciążeniach /rys. 1/.

$$\left[\left\langle j^a, j^b\right\rangle_1 \quad \left\langle j^a, j^b\right\rangle_2 \quad \ldots \quad \left\langle j^a, j^b\right\rangle_{v_j}\right], \quad j \in N.$$



Rys. 1. Przykładowy graf dysjunktywny D^o

Zbiorem <u>dysjunktywnych łuków operacyjnych V^o</u> /zbiorem wszystkich realizacji operacji/ będziemy nazywać

$$V^o = \bigcup_{j \in N} V_j^o, \quad \text{gdzie } V_j^o = \bigcup_{B_h \in B(j)} \left\{\left\langle j^a, j^b\right\rangle_h\right\}, \qquad /14/$$

przy czym długość każdego łuku $\left\langle j^a, j^b\right\rangle_h$ wynosi $c_{jh} > 0$ i $|V_j^o| = v(j)$, $j \in N$. Zbiór łuków V^o odpowiada ograniczeniom /w.1/ - /w.3/.

Uwzględniając implikacyjny warunek przydziału /w.4/ w ograniczeniu /w.2/, otrzymujemy dysjunkcję zawężoną do $v_r(j) \leq v(j)$ łuków, gdzie $v_r(j) = |B_r(j)|$ /8/. Zatem ograniczenie /w.4/ /po przenumerowaniu łuków po prawej stronie warunku implikacyjnego/ możemy przedstawić w równoważnej postaci:

$$S_r^o(j) \Longleftrightarrow \left[\langle j^a,j^b\rangle_1 \quad \langle j^a,j^b\rangle_2 \quad \dots \quad \langle j^a,j^b\rangle_{v_r(j)}\right], \quad S_r^o \in \mathcal{R}_S^o, \; j \in N-\Gamma O \qquad /15/$$

Ograniczenie to będziemy nazywać <u>implikacyjnym warunkiem przydziału łuków operacyjnych</u>.

Dla warunku /w.8/ dysjunkcji rozłącznej ograniczeń $\left(t_j^a - t_i^b \geqslant 0\right) \vee$ $\vee \left(t_i^a - t_j^b \geqslant 0\right)$ możemy przyporządkować w grafie G <u>dysjunktywną parę łuków</u> $\left[\langle i^b,j^a\rangle, \langle j^b,i^a\rangle\right]$ o obciążeniach równych zeru. Łuk $\langle i^b,j^a\rangle$ oznacza wykonanie operacji i przed j, a $\langle j^b,i^a\rangle$ oznacza wykonanie j przed i. Zatem każdemu ograniczeniu /w.8/ możemy przyporządkować <u>implikacyjny warunek dysjunktywnych łuków kolejnościowych</u>.

$$\left[\langle j^a,j^b\rangle_s \wedge \langle i^a,i^b\rangle_h\right] \Rightarrow \left[\langle i^b,j^a\rangle \vee \langle j^b,i^a\rangle\right],$$

$$\langle B_s,B_h\rangle \in B(i,j), \quad i,j \in N, \quad i \neq j \qquad /16/$$

Zbiór wszystkich <u>łuków dysjunktywnych w D</u>° będzie

$$V = \left\{\langle i^b,j^a\rangle \in B \times A \;\middle|\; i \; Rd \; j\right\}, \qquad /17/$$

gdzie: $i \; Rd \; j \Leftrightarrow \sim \exists \left[d(i,j) \wedge d(j,i) \text{ w } G\right] \wedge \underset{B_s \in B(i)}{\exists} \; \underset{B_h \in B(j)}{\exists} \left(P_s \cap P_h \neq \emptyset\right)$,

$i,j \in N$, $i \neq j$. Zbiór V jest zbiorem łuków łączących operacje i oraz j, $i \neq j$, dla których nie istnieje droga w G_1 i do wykonania tych operacji mogą być przydzielone zasoby zależne. W skład zbioru V wchodzą wszystkie dysjunktywne pary łuków, gdyż jeżeli $\langle i^b,j^a\rangle \in V$, to także jego <u>partner</u> $\langle j^b,i^a\rangle \in V$, co wynika z /17/. Z ograniczenia /w.8/ i /16/ wynika, że każdemu rozwiązaniu dopuszczalnemu zagadnienia P odpowiada podzbiór łuków dysjunktywnych $S \subset V$, który zawiera co najwyżej jeden łuk z każdej pary dysjunktywnej $\left[\langle i^b,j^a\rangle, \langle j^b,i^a\rangle\right]$. Taki podzbiór będziemy nazywać <u>reprezentacja zbioru V</u>. Niech $\mathcal{R}_S = \{S_1,S_2,\dots,S_a\}$ będzie rodziną wszystkich reprezentacji. Każda reprezentacja $S_i \in \mathcal{R}_S$ przedstawia pewną kolejność wykonywania operacji.

Z ograniczenia /w.2/ wynika, że również każdemu rozwiązaniu dopuszczalnemu odpowiada podzbiór łuków $S^o \subset V^o$, który zawiera dokładnie po jednym łuku z każdego zbioru dysjunktywnego V_j^o. Taki podzbiór będziemy nazywać <u>reprezentacja zbioru V^o</u> i jest on równoważny z wcześniej wprowadzoną reprezentacją zasobów złożonych, zatem $\mathcal{R}_S^o$ /7/ będzie rodziną wszystkich takich reprezentacji. Każda para reprezentacji $\langle S_r^o,S_r\rangle \in \mathcal{R}_S^o \times \mathcal{R}_S$ generuje graf

$$G_r^o = \langle \bar{N}, U \cup S_r^o \cup S_r\rangle. \qquad /18/$$

Niech $\mathcal{R}^o$ będzie rodziną grafów postaci /18/, które są bezkonturowe oraz niech $\mathcal{R}_1^o$ będzie podzbiorem grafów $G_r^o \in \mathcal{R}^o$ takich, że G_r^o spełnia implikacyjny warunek kolejności i przydziału.

<u>Definicja 2</u>. Sumę wartości funkcji dróg krytycznych f_o od wierzchołka 0 do wszystkich wierzchołków końcowych $e_k^b \in \bar{E}$ grafu $G_o^o \in \mathcal{R}_1^o$ będziemy nazywać <u>minimaksymalną z implikacyjnymi warunkami przydziału</u> /15/

i kolejności /16/ w grafie dysjunktywnym $D^o = \langle \bar{N}, U, V^o, V \rangle$, a związane z nią reprezentacje S_o i S_o^o nazywamy optymalnymi, jeżeli

$$f_o = \min_{G_r^o \in \mathcal{R}_1^o} \sum_{k \in K} f_k\left(L_r^k\right) = \min_{G_r^o \in \mathcal{R}_1^o} f_k^r , \qquad /19/$$

przy czym L_r^k jest długością drogi krytycznej od 0 do e_k^b, $k \in K$ w $G_r^o \in \mathcal{R}_1^o$. Postępując podobnie jak w [4,7] możemy udowodnić, że zagadnienie P jest równoważne zagadnieniu znalezienia grafu G_o^o z reprezentacjami optymalnymi S_o i S_o^o. Optymalne wartości t_j^a, t_j^b, $j \in N$ mogą być otrzymane przez stosowanie metody dróg krytycznych do grafu G_o^o.

W [4] uzupełniono i udowodniono nowe własności i twierdzenia dla ogólnego grafu dysjunktywnego D^o. Pozwoliło to zbudować najbardziej ogólne modele optymalizacyjne zagadnień przydziału i szeregowania oraz uzyskać nowe algorytmy optymalizacyjne.

Minimaksymalną sumę f_o w grafie G_o^o, w którym są spełnione implikacyjne warunki przydziału i kolejności otrzymamy drogą sterowanego przeglądania grafów $G_r^o \in \mathcal{R}^o$ metodą podziału i ograniczeń z mieszaną strategią przeglądania drzewa rozwiązań dopuszczalnych [4].

Istotnym problemem, niezależnym od wyboru strategii poszukiwań, jest sposób określenia dolnych ograniczeń dla bezpośredniego i dolnego następnika w drzewie rozwiązań. Przedstawienie analitycznej postaci tych ograniczeń wymagałoby wprowadzenia dalszych pojęć i oznaczeń, dlatego czytelnika odsyłamy do [4]. Podane tam ograniczenia dla kryterium sumy spóźnień są najsilniejsze, a sprowadzone do kryterium czasu zakończenia należą do najmocniejszych, znanych w literaturze. Warto zaznaczyć, że te ograniczenia łatwo jest przedstawić w postaci przydatnej dla ogólniejszego kryterium

$$\sum_{k \in K} f_k\left(\theta^k\right) \longrightarrow \min . \qquad /20/$$

3. Kryteria optymalizacji

W pracy sformułowano najbardziej ogólne kryterium opierające się na żądanych terminach wykonania zleceń. Podamy, jakie należy wykonać czynności na ogólnym grafie dysjunktywnym, aby tak przekształcony graf służył do obliczenia szczegółowych funkcji celu.

Poprzednio pokazaliśmy jak bezpośrednio na grafie zamodelować spóźnienie $\theta^k = \max\{0, C_k - d_k\}$ dla każdego zlecenia $k \in K$ /rys. 1/. Obliczenie drogi krytycznej od 0 do e_k^b było równoważne z obliczeniem θ^k. Następnie obliczone wartości θ^k, $k \in K$ metodami dróg krytycznych były odwzorowywane na wartości $\sum_{k \in K} f_k(\theta^k)$. Jeżeli $f_k(\theta^k) = \omega_k \theta^k$ to kryterium /20/ przybierze postać $\sum_{k \in K} \omega_k \theta^k$. Wartości ω_k mogą oznaczać koszty jednostkowego spóźnienia /lub koszty jednostkowego czasu wykonania/ bądź wartości

pilności dla zlecenia $k \in K$. Sformułowane kryterium wymusza, aby spóźnienie było mniejsze dla większych ω_k.

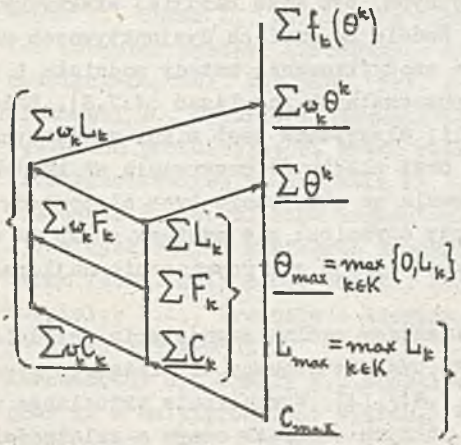
Jeżeli w ogólnym grafie dysjunktywnym odrzucimy łuki $\langle 0, e_k^b \rangle \in U_3$, to otrzymamy kryterium $\sum_{k \in K} f_k(L_k)$ lub $\sum_{k \in K} \omega_k L_k$ /, gdzie L_k algebraiczne spóźnienie/, bądź jeżeli $d_k = 0$, $k \in K$, to otrzymamy $\sum_{k \in K} f_k(C_k)$ lub $\sum_{k \in K} \omega_k C_k$. W szczególności dla $\omega_k = 1$, $k \in K$ odpowiednie funkcje kryterialne mają postać $\sum \theta^k$, $\sum L_k$ i $\sum C_k$. Dla stałej liczby zleceń powyższe funkcje są równoważne odpowiednio kryterium średniego /ważonego/: spóźnienia, algebraicznego spóźnienia i czasu wykonania [4].

W przypadku, gdy dane są terminy przybywania zleceń do systemu, tj. $r_k = 0$ oraz $d_k = 0$, $k \in K$, to jak łatwo zauważyć, kryterium sumy czasów przebywania zleceń w systemie można zapisać w postaci [4,11]:

$$\sum_{k \in K} F_k = \sum_{k \in K} (C_k - r_k) = \sum_{k \in K} \theta^k - \sum_{k \in K} r_k . \qquad /21/$$

Zatem kryterium /21/ jest równoważne kryterium $\sum_{k \in K} \theta^k$ dla $d_k = 0$, $k \in K$. Podobnie kryteria $\sum_{k \in K} f_k(F_k)$, $\sum_{k \in K} \omega_k F_k$ są odpowiednio równoważne kryteriom $\sum_{k \in K} f_k(\theta^k)$, $\sum_{k \in K} \omega_k \theta^k$ dla $d_k = 0$, $k \in K$. Dla stałej liczby zleceń kryterium /21/ jest równoważne kryterium średniego czasu przebywania zleceń w systemie.

Innym kryterium, istotnym w praktyce, jest kryterium maksymalnego spóźnienia $\theta_{max} = \max_{k \in K} \theta^k$. Zamodelowanie tej funkcji na ogólnym grafie dysjunktywnym polega na zwarciu wszystkich węzłów $e_k^b \in \bar{E}$ w jeden węzeł e^b. W przypadku, gdy odrzucimy łuki $\langle 0, e_k^b \rangle \in U_3$, to otrzymamy kryterium $L_{max} = \max_{k \in K} L_k$. Dla $d_k = 0$, $k \in K$ otrzymamy kryterium $C_{max} = \max_{k \in K} C_k$. Warto zwrócić uwagę, że zastosowanie kryterium θ_{max} dla jednakowych wartości $d_k = d_1$, $k \in K$ daje w wyniku szeregowanie operacji takie, że rozkład odchyleń od żądanego terminu realizacji wszystkich zleceń jest możliwie równomierny.



Rys. 2. Hierarchia ogólności jednowymiarowych funkcji kryterialnych

Przedstawione kryteria mieszczą się w dwóch klasach funkcji kryterialnych:

$$\sum_{k \in K} f_k(\theta^k) , \qquad /22/$$

$$\max_{k \in K} f_k(\theta^k) . \qquad /23/$$

Na rys. 2 jest przedstawiony w hierarchii ogólności zbiór szczególnych przypadków funkcji celu mieszczący się w powyższych klasach funkcji kryterialnych. Klamerkami oznaczono równoważność odpowiednich funkcji kryterialnych wcześniej uzasadnionych.

W pracy [8] wykazano, że kryteria podkreślone na rys. 2 reprezentują /są równoważne/ jednowymiarowe kryteria opierające się na czasie zakończenia, żądanym terminie wykonania zleceń i wykorzystania zasobów.

4. Uwagi końcowe

Zagadnienia ze szczegółowymi funkcjami celu przedstawionymi powyżej nie muszą spełniać wszystkich ograniczeń /w.1/ - /w.11/. Dla zagadnień sformułowanych ze szczegółowymi funkcjami celu, modele i algorytmy w zasadzie pozostają niezmienione, jedynie oszacowania ograniczeń dolnych zostają w ten sposób zmienione, że dla klasy funkcji kryterialnej /23/ zamiast $\sum_{k\in K}$ należy wpisać symbol $\max_{k\in K}$ [4]. Również łatwo jest wprowadzić obliczenie ograniczeń dolnych przedstawionych w [4], ze względu na funkcję celu $f_k(\theta^k)$.

W pracy [4] przedstawiono szereg modeli oraz algorytmów o wzrastających stopniach ogólności z kryterium minimalizacji $\sum_{k\in K}\theta^k$ oraz przedstawiono możliwości zastosowania tych algorytmów do zagadnień optymalizacyjnych z różnymi kryteriami. Sformułowano m.in. ogólne zagadnienie, w którym uwzględnia się operacje przestrojenia dokonywane przez SO, tzn. takie operacje, które przygotowują środowisko SK do wykonania kolejnej operacji. Są to czynności związane z alokacją danych, a także związane z obliczeniem samego algorytmu szeregowania. Postawione w pracy zagadnienia są ogólniejszej postaci niż dotychczas rozwiązane w literaturze, np. [7,8] ogólne zagadnienia kolejnościowe postaci $n|m|J,r_j \geq 0, RT|L_{max}$.

Modele na grafach dysjunktywnych wydają się być obecnie najbardziej efektywnym i wygodnym narzędziem do rozwiązywania ogólnych zagadnień szeregowania. Modele te dostarczają wiele informacji i mogą być punktem wyjścia do opracowania modeli permutacyjnych, być może bardziej efektywnych w sensie czasochłonności obliczeń. Modele na grafach dysjunktywnych umożliwiają w łatwy sposób zastosowanie zmodyfikowanej metody podziału i ograniczeń do sterowanego przeglądu dopuszczalnych rozwiązań [4,7,8]. Zaletami opracowywanych, na bazie tych modeli, algorytmów jest m.in. automatyczna generacja dopuszczalnych rozwiązań oraz możliwość przerwania na każdym etapie obliczeń. Ostatnia cecha pozwala na stosowanie tych algorytmów przy sterowaniu w czasie rzeczywistym, gdy ogranicza się odcinek czasu na wypracowanie decyzji. W tym przypadku otrzymuje się rozwiązanie najlepsze jakie uzyskano do tej pory.

Algorytm optymalizacyjny rozwiązujący ogólne zagadnienia przydziału zasobów i szeregowania operacji może być zastosowany do bieżącego sterowania w SK /w systemie produkcyjnym [7,8]/ [4]. Wyznaczanie aktualnego wektora sterowania może odbywać się w różnych chwilach czasu w zależności od potrzeb, znajomości parametrów zleceń, specyfiki SK, bądź zmian pilności poszczególnych zleceń. Proces obliczania algorytmu może zachodzić na bie-

żąco w SK, np. na wydzielonym minikomputerze.

Algorytm przydziału zasobów i szeregowania operacji - zasadnicza część SO - do których potrzebna jest odpowiednia znajomość charakterystyk zleceń, mogą znaleźć zastosowanie np. w projektowaniu SO dla komputerów pracujących metodą wsadową /w trybie wielodostępu/ oraz w sieciach komputerowych z wyspecjalizowanymi podsystemami komputerowymi. Systemy operacyjne mogą być ukierunkowane na zastosowanie do sterowania procesami technologicznymi, obiektami przemysłowymi, organizacjami, np. do harmonogramowania prac dla poszczególnych komputerów ośrodka obliczeniowego, dalej w cyklicznych obliczeniach w bankowości, gospodarce materiałowej itp., tj. wszędzie tam, gdzie istnieje znajomość wymagań technologicznych na poszczególne operacje.

LITERATURA

[1] Brinch Hansen P., Operating System Principles, Englewod/Cliffs, New Jersey, Prentice-Hall 1973 /w j. polskim, WNT, 1979/.

[2] Coffman E.G., A Survey of Mathematical Results in Flow-Time for Computer Systems, in: Lect. Notes in Computer Science ed. by E. Goos, Springer Verlag 1973, pp. 25-46.

[3] Coffman E.G., Computer and Job-Shop Scheduling Theory, New York, Wiley, 1976.

[4] Goetz J., Zagadnienia przydziału zasobów i szeregowania operacji w systemie cyfrowym /praca doktorska/, Inst. Cybern. Techn. Politechniki Wrocławskiej, Komunikat nr 774, Wrocław 1978.

[5] Goetz J., Huzar Z., Nowe podejście do opisu modelu systemu cyfrowego, Postępy Cybernetyki, No. 2, 1980.

[6] Goetz J., Huzar Z., Systematic Formulation of Management Problems for Computer System, Proceedings of the 4-th Symposium on System-Modeling-Control, Zakopane, oct. 8-12, 1979, part II, 42-47.

[7] Grabowski J., Algorytmy optymalizacji i sterowania w dyskretnych systemach produkcyjnych, Prace nauk. Inst. Cybern. Techn. Politechniki Wrocławskiej, Monografie nr 7, Wrocław 1977.

[8] Grabowski J., Uogólnione zagadnienia optymalizacji kolejności operacji w dyskretnych systemach produkcyjnych, Prace naukowe Inst. Cybern. Techn. Politechn. Wrocł., Monografie nr 9, Wrocław 1979.

[9] Pospielow D.A., Vvedenie v teoriju vycislitelnych sistem, Sovetskoje Radio, Moskva 1972.

[10] Ramamoorthy C., Chandy K., Gonzales M., Optimal Scheduling Strategies in a Multiprocessor System, IEEE Trans. Comput., vol. C-21, No. 2, 1972, pp. 137-146.

[11] Rinnoy Kan A.H.G., Machine Scheduling Problems: Classification, Complexity and Computations, Leinden 1976.

МОДЕЛИРОВАНИЕ БОЛЕЕ ШИРОКИХ ЗАДАЧ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ И ОЧЕРЕДНОСТИ ОПЕРАЦИЙ НА ДИЗЪЮНКТИВНОЙ ГРАФЕ

Р е з ю м е

В работе представлена формулировка и математическая модель более широких чем в литературе задач распределения ресурсов и очередности операций. Для вычислительной системы представлена более широкая дискретная система, чем система производства. На дизъюнктивном графе представлена интерпретация /независимая от рода системы/ критериев опирающихся на требуемых сроках выполнения задач.

MODELLING OF THE GENERAL PROBLEMS OF RESOURCES ALLOCATION AND OPERATIONS SCHEDULING THE DISJUNCTIVE GRAPH

S u m m a r y

The formulation and the mathematical model of resources allocation problems and operations scheduling problems, more general than in the literature, are presented in the paper. The problems concern the computer system, which is more general than the discrete production system described in [7,8] . The goal functions based on the due-date completion of jobs are pointed out. They are interpreted on the disjunctive graphs independently from a kind of a system.