

Stanisław WOLEK

## METODA OPTYMALIZACJI FIZYCZNEJ ORGANIZACJI DANYCH W PAMIĘCI KASETOWEJ

Streszczenie. Rozpatrywany jest problem optymalnego doboru wartości parametrów fizycznej organizacji danych w pamięciach kasetowych, opisywanej przez strukturę fizyczną danych oraz metodę dostępu do nich. Jako kryterium oceny organizacji, ze względu na wartości jej parametrów, przyjęty został maksymalny czas dostępu do danych, tzn. czas dostępu do ostatniej, w sensie ich ułożenia, z danych.

Dla tak przedstawionego problemu zaproponowano dwuetapową metodę jego rozwiązania i wykazano jej istotną zaletę w stosunku do podejścia naturalnego, polegającego na bezpośredniej minimalizacji kryterium. Optymalizacja jest przeprowadzana dla sekwencyjnej oraz indeksowo-sekwencyjnej metody dostępu do danych.

Summary. The problem of optimal selection of parameter values for the physical organization of data in cassette memory, given by physical structure and access method, is considered in this paper. The maximum data access time (i.e. time of access to latest data in sense of their sequential recording) is considered as the performance criterion of the organization with regard to its parameter values.

For the described problem the two-stage method of its solution is proposed and its substantial advantage has been demonstrated in comparison with natural approach as the direct criterion minimization. The optimization is performed for two data access strategies: sequential and indexing-sequential.

Резюме. В работе решается проблему оптимального подбора значений параметров физической организации данных в кассетной памяти, описанной через структуру данных и стратегию доступа к ним. Рассмотрено движущую, секвенционную структуру данных, с разделением на плитки и блоки. В качестве критерия оценки организации, в отношении к значениям параметров, принято максимальное время доступа к данным.

Продложено двухэтапный метод решения проблемы оптимизации и указано его действительные достоинства по отношению к непосредственной минимизации критерия. Практическая оптимизация проведена для последовательной и индексно-последовательной стратегии доступа к данным.



## 1. WPROWADZENIE

W pracach [1,2] przedstawione są wyniki optymalizacji fizycznej organizacji danych w pamięci kasetowej dla sekwencyjnej struktury danych oraz sekwencyjnego i indeksowo-sekwencyjnego do nich dostępu. Dobór parametrów organizacji zrealizowany jest dla dwóch przypadków: zmiennej wielkości bloków (blok nagłówka pliku może mieć inną długość niż bloki operacyjne w pliku) oraz takiej samej wielkości dla wszystkich bloków w pliku.

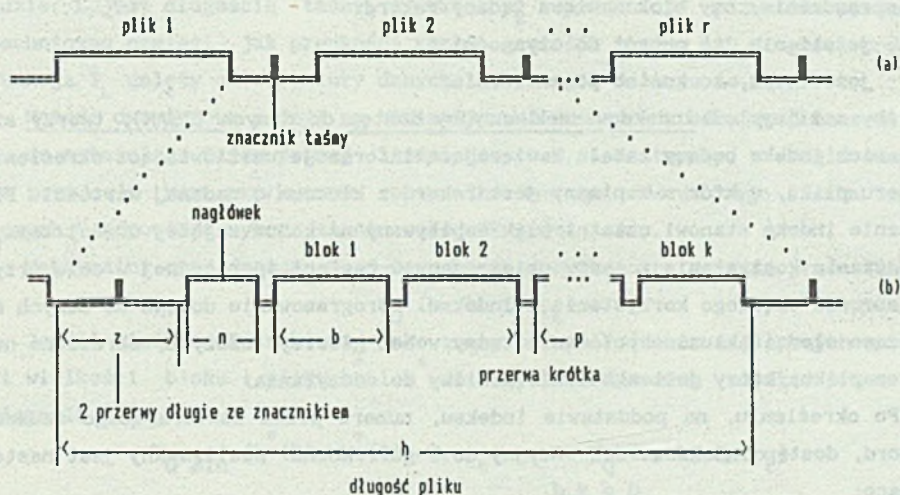
Istnienie zagadnienia optymalizacji wynika z faktu, że w pamięciach kasetowych stosuje się dwie prędkości ruchu taśmy. Odczyt, zapis albo pominięcie bloku lub znacznika wykonywane są z prędkością podstawową, natomiast z prędkością podwyższoną można wykonywać przewijanie taśmy o odcinki zawierające zadaną liczbę plików (do zadanego znacznika). Wszystkie operacje, poza zapisem, mogą być wykonywane w obu kierunkach ruchu taśmy. Dlatego przyjęta organizacja danych, wartości jej parametrów, jak i stosowana strategia dostępu do danych mają wpływ na szybkość pamięci.

Aby między operacjami było możliwe wstrzymanie ruchu taśmy z prędkością podstawową, bloki oddzielone są przerwami krótkimi. Wokół znacznika znajdują się natomiast przerwy długie, umożliwiające zatrzymanie taśmy po ruchu z prędkością podwyższoną. Stąd istnienie dwóch długości przerw wpływa na efektywne wykorzystanie pojemności pamięci. Optymalizacja organizacji danych przeprowadzana jest zaproponowaną w [1,2] metodą dwuetapową. W niniejszym artykule przeprowadza się optymalizację także metodą bezpośrednią, poprzez minimalizację funkcji określającej kryterium oceny modelu, tj. czasu dostępu do ostatniego bloku na taśmie. Porównanie wyników pozwala uwypuklić znaczenie metody dwuetapowej.

## 2. SFORMUŁOWANIE PROBLEMU

Ze względu na charakter pamięci, najpowszechniejszą strukturą danych jest dwupoziomowa struktura sekwencyjna. Bloki są łączone w pliki, a te zaopatrywane w nagłówki zawierające informacje o pliku, np. jego nazwę, kolejny numer lub maksymalną wartość klucza dla rekordów zapisanych w pliku (rys.1).





Rys. 1. Fizyczna struktura danych na taśmie (a) i w pliku (b) pamięci kasetowej

Fig. 1. Physical structure of data for the tape (a) and the file (b) in cassette memory

Zagadnienie optymalizacji fizycznej organizacji danych w pamięci kasetowej polega na doborze wartości parametrów fizycznej struktury danych dla określonej strategii dostępu do danych. W naszym przypadku parametrami organizacji są: wielkość bloku wyrażona liczbą zapisanych w nim bajtów danych oraz wielkość pliku, wyrażona liczbą zawartych w nim bloków.

Rozpatrywane są dwie strategię dostępu do danych.

Strategia dostępu sekwencyjnego (SAM) do bloku, który zawiera rekord z kluczem o zadanej wartości, składa się z następujących akcji (przy założeniu, że początkowo taśma jest przewinięta do jej fizycznego początku):

- 1) odczyt nagłówka pliku z prędkością podstawową  $v_1$ ,
- 2) sprawdzenie, czy szukany rekord znajduje się w pliku:
  - jeżeli tak, to przejście do czynności 4,
  - jeżeli nie, to kontynuacja (czynność 3),
- 3) przewinięcie taśmy z prędkością podwyższoną  $v_2$  do początku następnego pliku i powrót do czynności 1,



4) odczyt bloku danych z prędkością podstawową  $v_1$ ,

5) sprawdzenie, czy blok zawiera żądany rekord:

- jeżeli nie, to powrót do czynności 4,
- jeżeli tak, to koniec poszukiwania.

Aby możliwy był indeksowo-sekwencyjny dostęp do danych (ISAM), tworzy się dla nich indeks będący tabelą zawierającą informacje umożliwiające określenie numeru pliku, w którym zapisany jest rekord z kluczem o zadanej wartości. Fizycznie indeks stanowi ostatni plik zapisywany na końcu zajętej części taśmy, a w czasie korzystania z taśmy umieszczany w pamięci operacyjnej w celu przyspieszenia częstego korzystania z indeksu. Oprogramowanie dotępu do danych na bieżąco śledzi aktualne położenie taśmy wobec głowicy odczytu, określone numerem pliku, który jest aktualnie możliwy do odczytania.

Po określeniu, na podstawie indeksu, numeru pliku zawierającego szukany rekord, dostęp indeksowo-sekwencyjny do tego rekordu realizowany jest następująco:

- 1) przewinięcie taśmy z prędkością podwyższoną  $v_2$  w odpowiednim kierunku, od aktualnego położenia, bezpośrednio do początku pliku zawierającego szukany rekord,
- 2) odczyt nagłówka pliku z prędkością podstawową  $v_1$  (ewentualne sprawdzenie poprawności wykonania przewinięcia),
- 3) odczyt bloku danych z prędkością podstawową  $v_1$ ,
- 4) sprawdzenie, czy blok zawiera żądany rekord:
  - jeżeli nie, to powrót do czynności 3,
  - jeżeli tak, to koniec poszukiwania.

Za podstawę oceny fizycznej organizacji danych przyjmujemy maksymalny czas dostępu do danych, tj. czas dostępu do ostatniego bloku zapisanego na taśmie przy jej całkowitym wykorzystaniu, rozumiejąc przez czas dostępu sumę czasu jego wyszukania oraz odczytania. Rozpatrywane przy tym będą dwie metody prowadzące do rozwiązania zagadnienia optymalizacji fizycznej organizacji danych:

Optymalizacja bezpośrednia, polegająca na minimalizacji funkcji określającej kryterium, będącej czasem dostępu do ostatniego bloku na taśmie. Jeżeli przez  $b$  i  $k$  oznaczymy odpowiednio wielkość bloku i pliku, to można uważać, że kryterium  $T$  jest funkcją



$$T_L(b, k, L, p_t) \quad (1)$$

gdzie:  $L$  jest długością taśmy, a wektor  $p_t$  reprezentuje pozostałe parametry techniczne pamięci, jak prędkości ruchu, gęstości zapisu lub długości przerw. Funkcja  $T_L$  zależy od struktury danych i strategii dostępu do nich.

Metoda dwuetapowa polega na minimalizacji funkcji określającej czas dostępu do ostatniego bajtu pewnej ilości danych  $Q$ , a następnie, w drugim etapie, na przyjęciu jako rozwiązanie optymalne takiego, dla którego liczba bajtów danych  $Q$  odpowiada wykorzystaniu całej długości taśmy.

Maksymalny czas dostępu do  $Q$ -tego bajtu danych można wyrazić przez funkcję

$$T_Q(b, k, Q, p_t) \quad (2)$$

Jeżeli wielkość  $Q$  traktować jako znany parametr, dobór najlepszych wartości wielkości bloku i pliku  $b^*$  i  $k^*$  zależnych od  $Q$ , można uzyskać przez rozwiązanie zadania

$$T_{Q, \min}(b^*(Q), k^*(Q), Q, p_t) = \min_{b, k \in D} T_Q(b, k, Q, p_t) \quad (3)$$

gdzie:  $D$  oznacza ograniczenia.

W celu wyboru jednej optymalnej wartości parametrów  $b_{opt}$  i  $k_{opt}$  przyjmuje się w drugim etapie taką liczbę  $Q=Q_L(b, k, L, p_t)$  zapisanych bajtów danych, dla której przy podziale danych na  $k$  plików po  $b$  bloków zostaje wykorzystana cała długość  $L$  taśmy:

$$k_{opt} = k^*(Q_L(b^*, k^*, L, p_t)) \quad (4)$$

$$b_{opt} = b^*(Q_L(b^*, k^*, L, p_t))$$

### 3. ORGANIZACJA DANYCH Z DOSTĘPEM SEKWENCYJNYM

Wprowadźmy następujące oznaczenia parametrów technicznych pamięci oraz parametrów struktury danych (rys.1):

- $v_1, v_2$  - prędkość podstawowa i podwyższona ruchu taśmy,
- $g$  - gęstość zapisu,
- $p$  - długość przerwy krótkiej,
- $z$  - długość znacznika taśmy z przerwami długimi,
- $L$  - długość ścieżki zapisu,

- $n$  - długość nagłówka pliku,  
 $b$  - wielkość bloku operacyjnego, wyrażona liczbą zapisanych w nim bajtów informacji,  
 $k$  - liczba bloków operacyjnych w pliku,  
 $h$  - długość pliku,  
 $r$  - liczba plików na taśmie.

Operacja sekwencyjnego dostępu do  $k$ -tego bloku w  $r$ -tym pliku składa się z następujących czynności:

- $(r-1)$ -krotnego przeszukiwania plików początkowych, w którym nagłówek odczytywany jest z prędkością podstawową przez czas

$$\frac{n}{v_1}$$

a dalsza część pliku wraz ze znacznikiem pomijane są z prędkością podwyższoną przez czas

$$\frac{k(b/g + p) + z}{v_2}$$

- przeczytania z prędkością podstawową pełnego ostatniego pliku realizowanego w czasie

$$\frac{n + k(b/g + p)}{v_1}$$

Czas sekwencyjnego dostępu do  $k$ -tego bloku w  $r$ -tym pliku wyraża się więc następująco:

$$T(b, k, r, p_t) = (r-1) \left[ \frac{n}{v_1} + \frac{k(b + gp) + gz}{gv_2} \right] + \frac{gn + k(b + gp)}{gv_1} \quad (5)$$

### 3.1. Optymalizacja bezpośrednia

Jeżeli  $r$  plików po  $k$  bloków zajmuje całą taśmę o długości  $L$ , to zależność między długością taśmy a wielkościami bloków i plików ma postać

$$L = r(n + z + k(b/g + p)) \quad (6)$$



Po wyrugowaniu z wyrażeń (5) i (6) zmiennej  $r$  uzyskujemy wyrażenie określające czas  $T_L$  dostępu do ostatniego bloku zapisanego na taśmie, odpowiadające funkcji (1)

$$T_L(b, k, L, p_t) = \frac{aw_1}{k(b + gp) + c} + ak(b + gp) + \frac{L-z}{v_2} + \frac{n}{v_1} \quad (7)$$

gdzie 
$$a = \frac{1}{g} \left( \frac{1}{v_1} - \frac{1}{v_2} \right) \quad (8)$$

$$w_1 = g^2 nL \quad (9)$$

$$c = g(n + z) \quad (10)$$

Stosując podstawienie

$$x = k(b + gp) \quad (11)$$

sprowadzamy (7) do postaci

$$T_L(x, L, p_t) = \frac{aw_1}{x + c} + ax + \frac{L-z}{v_2} + \frac{n}{v_1}$$

Otrzymana funkcja posiada minimum, gdy

$$\frac{\delta T_L(x, L, p_t)}{\delta x} = - \frac{aw_1}{(x + c)^2} + a = 0$$

co prowadzi do rozwiązania

$$x = \sqrt{w_1} - c$$

które po uwzględnieniu (9), (10) i (11) przyjmuje postać

$$k(b + gp) = g(\sqrt{nL} - n - z) \quad (12)$$

Długość pliku wyraża się następująco

$$h = k(b/g + p) + n + z \quad (13)$$

co po uwzględnieniu (12) daje wartość optymalną długości pliku

$$h_{opt} = \sqrt{nL} \quad (14)$$

która zależy tylko od parametrów taśmy, a nie od przyjętej wielkości bloku i liczby bloków w pliku.

Uwzględniając równanie (12) we wzorze (6) uzyskujemy wzór na optymalną liczbę plików na taśmie

$$r_{opt} = \frac{\sqrt{gL}}{k(b + gp) + g(n + z)} = \sqrt{\frac{L}{n}} \quad (15)$$

zależną tylko od parametrów taśmy.

Minimalna wartość czasu dostępu do ostatniego bloku na taśmie, po uwzględnieniu równości (12) we wzorze (7), wynosi

$$T_{L,min} = \frac{L + n}{v_2} + 2\left(\frac{1}{v_1} - \frac{1}{v_2}\right)\sqrt{Ln} - \frac{n}{v_1} \quad (16)$$

Czas ten zależy także tylko od parametrów technicznych pamięci kasetowej.

Na podstawie uzyskanych wyników można sformułować następujące

#### Twierdzenie 1

Metodą bezpośredniej optymalizacji uzyskuje się nieskończenie wiele rozwiązań zagadnienia doboru wielkości bloku i pliku. Każda para tych wielkości, spełniająca równość (12), jest rozwiązaniem zadania, dając jednocześnie podział taśmy na stałą liczbą pików (15) o długości (14) i stały czas dostępu do ostatniego bloku na taśmie (16).

Wśród poprawnych optymalnych rozwiązań znajduje się też trywialne, w którym długość bloku jest równa zero ( $b=0$ ), liczba takich pustych bloków (składających się tylko z przerw) w pliku wynosi wg (12)  $k=(\sqrt{nL} - n - z)/p$  i uzyskuje się taki sam czas (16) dostępu do ostatniego bloku na taśmie, mimo że taśma nie zawiera żadnych danych operacyjnych, a jedynie organizacyjne.

Drugi skrajny przypadek optymalnego rozwiązania odpowiada założeniu, że plik składa się tylko z jednego bloku operacyjnego ( $k=1$ ). Wielkość tego bloku wynosi wg (12)

$$b_{sup} = g(\sqrt{nL} - n - z - p) \quad (17)$$

Aby wybrać jedno z optymalnych rozwiązań, należy zastosować dodatkowe kryterium, np. maksymalizację uzyskanej pojemności użytecznej taśmy, którą można wyrazić jako

$$Q_L = rkb \quad (18)$$

Ze względu na możliwość istnienia przekłamań przy wykonywaniu operacji odczytu bloku i wynikającej stąd konieczności powtarzania tych operacji, jak i ze względu na buforowanie treści bloku w pamięci operacyjnej przy sprowadzaniu



danych z pamięci zewnętrznej, należy ograniczyć wielkość bloku do pewnej wartości, oznaczonej  $b_{\max}$ . Z charakteru wielkości  $k$  oznaczającej liczbę bloków w pliku, wynika też ograniczenie  $k \geq 1$ .

Maksymalizację użytecznej pojemności taśmy można więc, po uwzględnieniu (18), (15) i (12), przedstawić następująco

$$\text{przy } Q_{L, \max} = \max_{k, b} Q_L = \max_{k, b} kb\sqrt{L/n}$$

$$k(b + gp) = g(\sqrt{Ln} - n - z)$$

$$k \geq 1$$

$$b \leq b_{\max}$$

Po zastosowaniu dla optymalizacji metody Kuhna-Tuckera [3] uzyskuje się następujące rozwiązanie

$$\text{Jeżeli } b_{\max} \geq b_{\text{sup}}, \text{ to } k_{\text{opt}} = 1, \quad b_{\text{opt}} = g(\sqrt{Ln} - n - z - p), \quad (18)$$

$$Q_L = g(L - (n + z + p)\sqrt{Ln}),$$

$$\text{Jeżeli } b_{\max} < b_{\text{sup}}, \text{ to } k_{\text{opt}} = \frac{g(\sqrt{Ln} - n - z)}{b_{\max} + gp}, \quad b_{\text{opt}} = b_{\max}, \quad (19)$$

$$Q_L = \frac{gb_{\max}}{b_{\max} + gp} (L - \sqrt{Ln} - z\sqrt{L/n})$$

### Przykład 1

Rozpatrzmy pamięć kasetową PK-1 produkcji "Meramat" w Warszawie, która ma następujące parametry techniczne:

- prędkości ruchu taśmy  $v_1 = 127$  [mm/s],  $v_2 = 1500$  [mm/s],
- gęstość zapisu  $g = 4$  [B/mm],
- długość taśmy  $L = 85$  [m]

i przyjmijmy następujące wartości stałych parametrów fizycznego modelu danych:

- długość przerwy krótkiej  $p = 20$  [mm],
- długość nagłówka pliku  $n = 24$  [mm],



- długość znacznika taśmy z przerwami długimi  $z=181$  [mm].

Uzyskujemy następujące optymalne wartości parametrów fizycznej struktury danych:

- długość pliku  $h_{opt} = 1428$  [mm],
- liczba plików na taśmie  $r_{opt} = 60$ ,
- czas dostępu do ostatniego bloku  $T_{L,min} = 75,9$  [s],
- skrajnie dopuszczalna wielkość bloku  $b_{sup} = 4813$  [B],

natomiast dla różnych ograniczeń na wielkość bloku, mniejszych od wartości skrajnej, uzyskuje się następujące optymalne liczby bloków w pliku i użyteczne pojemności taśmy (tab.1).

Tabela 1

Optymalne liczby bloków w pliku  
i użyteczne pojemności taśmy

$b_{mx}$ [B]	32	64	128	256	512
$k_{opt}$	44	34	24	15	8
$Q_L$ [kB]	83,2	129,4	179,2	221,9	251,9

### 3.2. Metoda dwuetapowa

Jeżeli w równości (5) wyrugujemy zmienną  $r$  wyrażając ją przez wielkość  $Q$  zapisanych bajtów danych, spełniających zależność

$$Q = rkb \quad (20)$$

to uzyskamy następującą zależność czasu dostępu do  $Q$ -tego bajtu danych  $T_Q$  od zmiennych  $b$  i  $k$ , odpowiadającą funkcji (2).

$$T_Q(b, k, Q, p_t) = awQ \frac{1}{bk} + \frac{pQ}{v_2} \frac{1}{b} + ak(b + gp) + \frac{Q - gz}{gv_2} \quad (21)$$

gdzie  $a$  jest dane przez (8) oraz

$$w = \frac{q (nv_2 + zv_1)}{v_2 - v_1} \quad (22)$$



W pierwszym etapie zaproponowanej metody dobiera się optymalne wartości wielkości  $b$  i  $k$ , w funkcji parametru  $Q$ , czyli rozwiązuje zadanie (3). Minimalizację funkcji (21) dwóch zmiennych  $b$  i  $k$ , przy ograniczeniach

$$\begin{aligned} b &\leq b_{\max} \\ k &\geq 1 \end{aligned} \quad (23)$$

przeprowadza się metodą Kuhna-Tuckera, w sposób przedstawiony w [1,2], uzyskując wynik:

Tabela 2

Wynik minimalizacji funkcji dwóch zmiennych  $b$  i  $k$

Przypadek	$Q \leq Q_1$	$Q_1 \leq Q \leq Q_2$	$Q \geq Q_2$
Wielkość boku	$b^* = \sqrt{uQ}$	$b^* = b_{\max}$	$b^* = b_{\max}$
Liczba bloków w pliku	$k^* = 1$	$k^* = 1$	$k^* = \sqrt{wQ/b_{\max}/(b_{\max} + gp)}$
Liczba plików	$r^* = \sqrt{Q/u}$	$r^* = Q/b_{\max}$	$r^* = \sqrt{Q(b_{\max} + gp)/w/b_{\max}}$

gdzie

$$u = w + \frac{p}{av_2}, \quad Q_1 = \frac{b_{\max}^2}{u}, \quad Q_2 = \frac{b_{\max} (b_{\max} + gp)}{w}$$

W drugim etapie omawianej metody określa się taką wartość  $Q_L$  zapisanych bajtów danych, która odpowiada wykorzystaniu całej długości  $L$  taśmy przy podziale danych na bloki i pliki wg wzorów z tabeli 2. Rugując z wyrażeń (6) oraz (20) zmienną  $r$  uzyskujemy wzór na ilość informacji zapisanej na całej taśmie

$$Q_L = k b \frac{gL}{kb + g(kp + n + z)} \quad (24)$$

Zastąpienie parametru  $Q$  przez  $Q_L$  w rozwiązaniu przedstawionym w tabeli 2 ograniczymy do ostatniego przypadku rozwiązania



$$Q \approx \frac{b_{mx} (b_{mx} + gp)}{w} \quad (25)$$

ponieważ ten odpowiada praktycznym wartościom parametrów pamięci kasetowej i ograniczeniom na wielkość bloku.

Podstawiając w wyrażeniu (24)  $b=b_{mx}$  oraz  $k$  wg tabeli 2 uzyskujemy równanie kwadratowe

$$(b_{mx} + gp) k^2 + g(n + z) k - \frac{gwL}{(b_{mx} + gp)} = 0$$

mające rozwiązanie

$$k_{opt} = \frac{d - c}{2(b_{mx} + gp)} \quad (26)$$

gdzie

$$c = g(n + z), \quad d = \sqrt{4gwL + c^2} \quad (27)$$

Warunek (25), po uwzględnieniu  $b=b_{mx}$  oraz (24) i (26), można przedstawić w postaci

$$b_{mx} \leq b_{sup} = \frac{d - c}{2} - gp \quad (28)$$

Wielkości dalszych parametrów struktury danych wyrażają się następująco:

$$\text{długość pliku} \quad h_{opt} = \frac{d + c}{2g} \quad (29)$$

$$\text{liczba plików} \quad r_{opt} = \frac{2gL}{d + c} \quad (30)$$

$$\text{pojemność taśmy} \quad Q_L = \frac{d - c}{d + c} \frac{b_{mx} gL}{b_{mx} + gp} \quad (31)$$

$$\text{czas dostępu} \quad T_{L, min} = \left( 2agw + \frac{d - c}{v_2} \right) \frac{L}{d + c} + a \frac{d - c}{2} - \frac{z}{v_2} \quad (32)$$

W przeciwieństwie do wyniku z metody bezpośredniej uzyskane zostało jednoznaczne rozwiązanie, które można sformułować następująco:

## Twierdzenie 2

Jeżeli parametry pamięci kasetowej, taśmy magnetycznej oraz przyjęte ograniczenie wielkości bloku  $b_{mx}$  spełniają warunek (28), to minimalny czas dostępu do ostatniego zapisanego na taśmie bajtu danych nie zależy od wielkości bloku i liczby bloków w pliku i uzyskuje się go przy podziale taśmy na pliki o stałej długości (29), w liczbie (30). Wielkość bloku powinna przyjmować maksymalną dopuszczalną wartość  $b_{mx}$ , od której zależy liczba bloków w pliku (26) oraz uzyskiwana efektywna pojemność taśmy (31).

## Przykład 2

Dla pamięci kasetowej z przykładu 1 uzyskuje się następujące wartości parametrów fizycznej struktury danych, otrzymane metodą dwuetapową:

- długość pliku  $h_{opt} = 2016$  [mm],
- liczba plików na taśmie  $r_{opt} = 42$ ,
- czas dostępu do ostatniego bloku  $T_{L,min} = 76,9$  [s],
- dopuszczalna wielkość bloku  $b_{sup} = 7169$  [B],

natomiast dla różnych ograniczeń na wielkość bloku uzyskuje się następujące optymalne liczby bloków w pliku i użyteczne pojemności taśmy:

Tabela 3

Optymalne liczby bloków w pliku i użyteczne pojemności taśmy

$b_{mx}$ [B]	32	64	128	256	512
$k_{opt}$	65	50	35	22	12
$Q_L$ [kB]	87,3	135,8	188,0	232,7	264,2

## 3.3. Porównanie metod i wyników

Dla obydwu metod uzyskano rozwiązania o tym samym jakościowo charakterze. W obu przypadkach długość pliku, liczba plików na taśmie i czas dostępu do ostatniego bloku danych nie zależą od doboranych parametrów  $b$  i  $k$ . Istnieje jednak różnica ilościowa między uzyskanymi wynikami, ponieważ wymienione



wielkości przyjmują dla każdej metody inne wartości, jak i inne są też wzory na optymalną liczbę blików w pliku i uzyskiwaną pojemność użyteczną.

Dla przykładowej pamięci kasetowej (z porównania wyników przedstawionych w tabeli 1 i 3) uzyskano metodą bezpośrednią mniejszą wartość czasu dostępu do ostatniej danej zapisanej na taśmie, ale jednocześnie też mniejszą pojemność użyteczną taśmy, dlatego bezpośrednie porównanie czasów dostępu nie jest miarodajne.

Można wykazać, że metodą bezpośrednią uzyskuje się stale mniejszą pojemność użyteczną taśmy. Wystarczy w tym celu porównać wyrażenia na  $Q_L$  ze wzoru (19) i (31).

$$\frac{d-c}{d+c} \frac{b_{mx}gL}{b_{mx}+gp} > \frac{gb_{mx}}{b_{mx}+gp} (L - \sqrt{nL} - z\sqrt{L/n})$$

Podana nierówność jest stale spełniona, gdyż po przekształceniach uzyskuje postać

$$4Lv_1 + (n+z)(v_2 - v_1) > 0$$

Aby pomóc porównać czasy dostępu, należy je wyznaczyć dla tych samych liczb zapisywanych na taśmie bajtów danych, np. dla mniejszej z dwóch pojemności maksymalnych, czyli pojemności (19), uzyskanej pierwszą metodą. Czas dostępu dla pierwszej metody podany jest przez (16). Czas dostępu dla drugiej metody uzyskuje się wg (21), podstawiając tam  $b-b_{mx}$ ,  $k$  wg (26) oraz  $Q=Q_L$  wg (19)

$$T_2 = 2a\sqrt{wg(L - \sqrt{nL} - z\sqrt{L/n})} + \frac{1}{v_2} (L - \sqrt{nL} - z\sqrt{L/n}) - \frac{z}{v_2}$$

Po przekształceniach można wykazać, że różnica czasu (16) i  $T_2$  jest z pewnym zapasem dodatnia, gdy

$$(\sqrt{L/n} + 1) v_1 / v_2 > 1 \quad (33)$$

co jest praktycznie stale spełnione.

## Wniosek

Metoda dwuetapowa daje korzystniejsze wyniki, gdyż dla wyznaczonych według niej parametrów organizacji danych uzyskuje się mniejsze czasy dostępu do tego samego bajtu danych.

## Przykład 3

Dla pamięci kasetowej z przykładów 1 i 2 lewa strona nierówności (33) przyjmuje wartość 5.1. W tabeli 4 przedstawione są wartości parametrów organizacji danych, obliczone dwiema metodami i wyznaczone dla tych samych wartości  $Q_1$  bajtów danych (kolumna 1 oznacza długość części taśmy potrzebnej do zapisania  $Q_1$  bajtów danych wg wyznaczonych parametrów ich organizacji).

Wartości parametrów organizacji danych

Tabela 4

					$b_{mx}$ [B]	32	64	128	256	512
Metoda	$h_{opt}$ [mm]	$r_{opt}$	$T_{L,min}$ [S]	$l$ [m]	$Q_1$ [kB]	83.2	129.4	179.2	221.9	251.9
Bezpośrednia	1428	60	75.9	85	$k_{opt}$	44	34	24	15	8
Dwuetapowa	1974	41	73.9	81	$k_{opt}$	63	49	34	21	12

Wiersz dla metody bezpośredniej odpowiada wynikom z przykładu 1. Wiersz dla metody dwuetapowej wyznaczono następująco:

- dla takich samych  $Q_1$ , jak w metodzie bezpośredniej, wyznacza się  $k_{opt}$  jako  $k$  z drugiego wiersza tabeli 2,
- $r_{opt}$  wyznacza się z (20), dla znanych  $b=b_{mx}$ ,  $Q=Q_1$  i  $k=k_{opt}$ ,
- $h_{opt}$  wyznacza się z (13), dla znanych  $k$  i  $b$ ,
- $l$  wyznacza się z (6), dla znanych  $r$ ,  $k$  i  $b$ ,
- $T_{L,min}$  wyznacza się z (21), dla znanych  $Q$ ,  $b$  i  $k$ .

Dla parametrów organizacji danych wyznaczonych metodą dwuetapową otrzymano krótszy czas dostępu do tych samych  $Q_1$  bajtów danych zapisanych na taśmie uzyskując także mniejszą długość  $l$  wykorzystywanej na to części taśmy.



## 4. ORGANIZACJA DANYCH Z DOSTĘPEM INDEKSOWO-SEKWENCYJNYM

Przy indeksowo-sekwencyjnej strategii dostępu do danych, czas dostępu do bloku zawierającego szukany rekord danych zależy jest od położenia taśmy wobec głowicy odczytu przed rozpoczęciem operacji poszukiwania. Poszukiwanie to nie jest bowiem robione sekwencyjnie od początku taśmy, lecz (wykorzystując informacje z indeksu) przechodzi się bezpośrednio od aktualnego położenia do początku szukanego pliku. Maksymalny czas dostępu osiąga się wtedy, gdy położenie głowicy w stosunku do położenia szukanego bloku jest skrajnie odległe, np. głowica na początku taśmy, a szukany blok jako ostatni w pliku zapisanym na końcu taśmy.

Operacja indeksowo-sekwencyjnego przejścia od początku taśmy do k-tego bloku w r-tym pliku składa się z:

- przewinięcia  $(r-1)$  plików początkowych z prędkością podwyższoną  $v_2$  przez czas

$$(r-1) \frac{n + k(b/g + p) + z}{v_2}$$

- przeczytania z prędkością podstawową  $v_1$  pełnego ostatniego pliku realizowanego w czasie

$$\frac{n + k(b/g + p)}{v_1}$$

Maksymalny czas indeksowo-sekwencyjnego dostępu do danych wyraża się więc (przy pominięciu czasu konsultowania indeksu) następująco:

$$T(b, k, r, p_t) = (r-1) \frac{k(b + gp) + g(n + z)}{gv_2} + \frac{gn + k(b + gp)}{gv_1} \quad (34)$$

## 4.1. Optymalizacja bezpośrednia

Z równania (34) można wyrugować zmienną  $r$  korzystając z zależności (6) wiążącej tę zmienną z wielkością bloku i pliku oraz długością taśmy. Uzyskuje się następujące wyrażenie na czas dostępu do ostatniego bloku zapisanego na taśmie

$$T_L(b, k, L, p_t) = ak(b + gp) + \frac{L}{v_2} + agn - \frac{z}{v_2} \quad (35)$$

Jeżeli założy się, że  $k$  i  $b$  mogą mieć wartości nieujemne, to otrzymana funkcja posiada minimum ze względu na te wielkości, gdy  $k(b + gp) = 0$ . Odpowiada to rozwiązaniu:  $k = 0$  i  $b$  dowolne, co oznacza, że pliki nie zawierają bloków operacyjnych, których wielkość może być wobec tego dowolna. Minimalny czas dostępu do końca taśmy uzyskuje się więc dla struktury trywialnej, w której pliki zawierają tylko nagłówki i żadnych danych operacyjnych.

Jeżeli przyjmie się ograniczenie  $k \geq 1$  i  $b \geq 0$ , to minimum funkcji (35) osiąga się dla  $k = 1$  i  $b = 0$ , co oznacza, że w każdym pliku jest tylko jeden pusty blok operacyjny, a więc plik składa się tylko z nagłówka i przerwy, co jest także rozwiązaniem trywialnym.

#### Wniosek

Metodą bezpośredniej optymalizacji uzyskuje się minimum czasu dostępu do ostatniej danej wtedy, gdy taśma nie zawiera żadnych danych operacyjnych, a jedynie organizacyjne, co jest niemożliwym do przyjęcia rozwiązaniem trywialnym.

#### 4.2. Metoda dwuetapowa

Jeżeli w równości (34) wyrugujemy zmienną  $r$  korzystając z zależności (20), to po przekształceniach uzyskujemy

$$T_Q(b, k, Q, p_t) = a w_2 Q \frac{1}{bk} + \frac{pQ}{v_2} \frac{1}{b} + ak(b + gp) + \frac{Q - gz}{g v_2} + a g n \quad (35)$$

gdzie

$$w_2 = \frac{(n + z)v_1}{v_2 - v_1} \quad (36)$$

Otrzymane wyrażenie (35) ma tę samą postać, co (21) dla modelu sekwencyjnego, co prowadzi do podobnego rozwiązania.

#### Twierdzenie 3

Jeżeli parametry pamięci kasetowej, taśmy magnetycznej oraz przyjęte ograniczenie wielkości bloku  $b_{\max}$  spełniają warunek

$$b_{\max} \leq b_{\text{sup}} = \frac{e - c}{2} - gp \quad (37)$$

gdzie

$$c = g(n + z), \quad e = \sqrt{4g w_2 L + c^2} \quad (38)$$



to minimalny czas dostępu do ostatniego zapisanego na taśmie bajtu danych nie zależy od wielkości bloku i liczby bloków w pliku i uzyskuje się go przy podziale taśmy na pliki o stałej długości

$$h_{opt} = \frac{e + c}{2g} \quad (39)$$

w liczbie

$$r_{opt} = \frac{2gL}{e + c} \quad (40)$$

Wielkość bloku powinna przyjmować maksymalną dopuszczalną wartość  $b_{max}$ , od której zależy liczba bloków w pliku

$$k_{opt} = \frac{e - c}{2(b_{max} + gp)} \quad (41)$$

oraz uzyskiwana efektywna pojemność taśmy

$$Q_L = \frac{e - c}{e + c} \frac{b_{max} gL}{b_{max} + gp} \quad (42)$$

#### Przykład 4

Dla przykładowej pamięci kasetowej uzyskuje się:

- długość pliku  $h_{opt} = 1376$  [mm],
- liczbę plików na taśmie  $r_{opt} = 62$ ,
- czas dostępu do ostatniego bloku  $T_{L,min} = 65,0$  [s],
- dopuszczalną wielkość bloku  $b_{sup} = 4605$  [B]

natomiast różnym ograniczeniom na wielkość bloku odpowiadają następujące optymalne liczby bloków w pliku i użyteczne pojemności taśmy:

Tabela 5

Optymalne liczby bloków w pliku i użyteczne pojemności taśmy

$b_{max}$ [B]	32	64	128	256	512
$k_{opt}$	42	33	23	14	8
$Q_L$ [kB]	82,7	128,6	178,0	220,5	250,3



## 5. PODSUMOWANIE

Z porównania dwóch metod optymalizacji organizacji danych w pamięci kasetowej, tj. bezpośredniej minimalizacji kryterium oceny modelu oraz zaproponowanej metody dwuetapowej wynika, że optymalizacja bezpośrednia daje rozwiązania trywialne, praktycznie nieprzydatne, natomiast do zadowalających wyników prowadzi metoda dwuetapowa. Uzyskuje się nią takie same jakościowo rozwiązanie zarówno dla sekwencyjnej, jak i indeksowo-sekwencyjnej strategii dostępu do danych, przedstawione w twierdzeniach 2 i 3.

Wyników praktycznych dla przykładowej pamięci kasetowej, uzyskanych dla obydwóch strategii i przedstawionych w tabelach 3 i 5, nie można porównać bezpośrednio, gdyż uzyskuje się dla nich różne pojemności użyteczne taśmy.

Określone tam czasy dostępu dotyczą przeszukiwania różnych ilości informacji, ponieważ oznaczają one czasy dostępu do ostatniego bloku danych zapisanego na taśmie. W tabeli 6 przedstawione są wyniki porównywalne, wyznaczone dla tych samych wartości  $Q_1$  zapisanych bajtów danych, odpowiadających mniejszym pojemnościom użytecznym, uzyskiwanym dla dostępu indeksowo-sekwencyjnego.

Wiersz dla dostępu indeksowo-sekwencyjnego odpowiada wynikom z przykładu. 4. Wiersz dla dostępu sekwencyjnego został wyznaczony w sposób identyczny z przykładem 3.

Tabela 6

Wyniki porównania metod

					$b_{mx}$ [B]	32	64	128	256	512
Dostęp	$h_{opt}$ [mm]	$r_{opt}$	$T_{L,min}$ [s]	$l$ [m]	$Q_1$ [kB]	82.7	128.6	178.0	220.5	250.3
Indeks.-sekw.	1376	62	65.0	85.0	$K_{opt}$	42	33	23	14	8
Sekwencyjny	1968	41	73.5	80.7	$K_{opt}$	63	49	34	21	12



Dla dostępu indeksowo-sekwencyjnego uzyskuje się krótszy czas dostępu do ostatniego bloku danych, co jest zgodne z intuicyjnym oczekiwaniem. Potrzebne jest jednak wtedy większe zużycie długości taśmy dla zapisu tej samej ilości danych.

#### LITERATURA

1. S.Wołek: Fizyczny model danych w pamięci kasetowej dla struktury i dostępu sekwencyjnych. "Podstawy Sterowania" 1988 r. tom 18, z. 1-2.
2. S.Wołek: Optymalizacja fizycznego modelu danych w pamięci kasetowej systemów komputerowych. Zeszyty Naukowe Pol.Śl., seria Informatyka, Gliwice 1989, z. 13.
3. W.Findeisen, J.Szymanowski, A.Wierzbicki: Teoria i metody obliczeniowe optymalizacji. PWN, Warszawa 1977.

Recenzent: doc.dr hab.inż. Adam Mrózek

Wpłynęło do Redakcji: 20.11.1990 r.

#### METHODE OF THE OPTIMIZATION OF THE PHYSICAL DATA ORGANIZATION IN CASSETTE MEMORY

##### Abstract

The paper deals with the problem of optimal selection of parameter values for the physical organization of data in cassette memory. The optimization problem stems from the fact that a cassette memory uses two different tape speeds with different tape-stopping times for each speed, which makes that a given data organization may influence the practical speed of the tape and the effectiveness of the tape use.

A two-level sequential physical data structure and two data access strategies: sequential and indexing-sequential, are considered. Size for data block

One file are selected in the case of a sequential access assuming that the gathered data are of a homogeneous character, i.e. they can be collected into blocks and/or files in any way.

The maximum data access time (i.e. time of access to latest data in sense of their sequential recording) is considered as the performance criterion of the organization with regard of its parameter values.

For the described problem the two-stage method of its solution is proposed and its substantial advantage has been demonstrated in comparison with natural approach as the direct criterion minimization.

It has been demonstrated that a maximal possible block size and a constant physical size of file, the latter depending only on technical parameters of the memory device, should be used. As for the block size it influences only the effective capacity of memory and not data access time.