FIFO-OPTIMAL PLACEMENT ON PAGES OF INDEPENDENTLY REFERENCED SECTORS

Arno ENDRES
Frankfurt, West Germany

Hans-Georg STORK
Institut für Informatik der Universität Stuttgart, West Germany

Received 3 December 1976

Optimal paginations, paging-algorithms, FIFO, locality principle

It has been widely recognized (and confirmed by experimental results, cf. [5]), that the placement of relocatable sectors of some program considerably affects its paging-behaviour. Therefore it is natural to ask for those placements of sectors on pages, given a sector-reference-string or a whole class of sector-reference-strings, which produce page-reference-strings (or classes thereof) which are in some sense optimal for a particular paging-algorithm. The model of reference-behaviour underlying the present note is known as the "independent reference model" (cf. [2]). It has been observed elsewhere [6] that under the assumption of independent, equally distributed sector-references the so called "frequency-order pagination" (i.e. place the most probable sectors on one page, the most probable of the remaining sectors on a second page and so on!) is optimal for an optimal demand-paging-algorithm – like A_0 of [1] – but does not need to be optimal for non-optimal algorithms, like for instance LRU [4]. Here, optimality means that the mean "page-fault-rate" is minimized. It is the purpose of this note to show that – under the said criterion of optimality – the "frequency-order pagination" is also optimal with respect to the "first-in-first-out"-(FIFO-) demand-paging-algorithm.

We shall briefly describe the formal setting necessary to prove our claim: let $X = \{x_1, ..., x_n\}$ be a set of pages and let $\not d \notin X$ be some extra symbol. In accordance with [1], the FIFO-algorithm will be described as a sequential state-machine with input-set X: Let $M < n$ be some natural number (the "page-capacity of main-memory"); the state-set Q is the set of all M-tuples

$$q = (y_1, ..., y_M)\,(y_i \in X \cup \{\not d\}\,, \quad 1 \leqslant i \leqslant M)\,,$$

whose first k (for any k, $0 \leqslant k \leqslant M$) entries all equal $\not d$ and whose last $M-k$ entries are pairwise distinct. Upon input of some symbol $x \in X$ state $q \in Q$ will change to

$$\delta(q, x) = \begin{cases} q, \text{ if } x = y_i & \text{for some } i,\ 1 \leqslant i \leqslant M\,, \\ (y_2, ..., y_M, x), & \text{otherwise.} \end{cases}$$

If the latter happens (i.e. "if x has to be added to the memory contents"), then costs $\kappa(q, x) = 1$ will arise; otherwise $\kappa(q, x) = 0$. After sequentially extending the mapping $\delta: Q \times X \to Q$ to the domain $Q \times X^*$ (where X^* denotes the set of all finite input-strings of elements of X), we can define the costs of processing symbol by symbol, some string $w = y_1 \cdots y_l \in X^*$ starting from state q, as the accumulated costs

$$K(q, w) := \kappa(q, y_1) + \kappa(\delta(q, y_1), y_2) + \cdots + \kappa(\delta(q, y_1 \cdots y_{l-1}), y_l)$$

(which is just the number of "page-faults"). Denote the machine defined above by FIFO(M).

Now let $S = \{s_1, ..., s_N\}$, where $N = f \cdot n$ for some natural number f, be a set of equal-sized sectors and let $\pi: S \to [0, 1]$ be a probability assigned to the elements of S in such a way that $\pi(s_1) \geqslant \pi(s_2) \geqslant \cdots \geqslant \pi(s_N)$.

A *pagination* of S is a total mapping $h: S \to X$, such that for all $x \in X$: $|h^{-1}(x)| = f$. h induces a probability $\pi_h: X \to [0, 1]$ as follows:

$$\pi_h(x) = \sum_{s \in h^{-1}(x)} \pi(s) .$$

Assuming sector-reference-strings to be generated by independent random variables (with values in S) distributed according to π and assuming S to be paginated by h, we obtain a probability $\Pi_h: X^l \to [0, 1]$ on the set of page-reference-strings of length l, generated by independent random variables (with values in X) distributed according to π_h.

Hence

$$\sum_{w \in X^l} \Pi_h(w)\, K(q, w)$$

(X^l denoting the set of input-strings of length l), is the expected number of "page-faults" while processing a sector-reference-string of length l, under the pagination h, and

$$F(\text{FIFO}(M), h, q) := \limsup_{l \to \infty} \frac{1}{l} \sum_{w \in X^l} \Pi_h(w) K(q,w)$$

is the asymptotic mean "page-fault-rate".

It has been shown in [7] (cf. also [2], p. 274) that $F(\text{FIFO}(M), h, q)$ is independent of q and that:

$$F(\text{FIFO}(M), h) = \frac{\sum \pi_h(y_1) \cdots \pi_h(y_M) \left(1 - \sum_{j=1}^{M} \pi_h(y_j)\right)}{\sum \pi_h(y_1) \cdots \pi_h(y_M)} ,$$

where the sums are over all $q = (y_1, ..., y_M) \in Q$ whose entries are different from $\not{q}$. Denote this subset by $\overline{Q}$.

Now let $h_0: S \to X$ be the "frequency-order pagination", i.e.

$$h_0(s_i) = x_{\lceil \frac{i}{f} \rceil}$$

(where $\lceil \frac{i}{f} \rceil$ is the smallest integer larger than $\frac{i}{f}$), for $i = 1, ..., N$. Our claim is:

$$F(\text{FIFO}(M), h_0) \leqslant F(\text{FIFO}(M), h) ,$$

for any other pagination h.

To show this, let h be some pagination such that $\pi_h \neq \pi_{h_0}$ (if no such pagination exists then our claim holds trivially with equality for all paginations!). It is then clear that there exist sectors $s_k, s_l \in S$ such that $h(s_k) \neq h(s_l)$, $\pi(s_k) < \pi(s_l)$ and $\pi_h(h(s_k)) \geqslant \pi_h(h(s_l))$. Let h' be the pagination obtained from h by "interchanging s_k and s_l", i.e. $h'(s_k) = h(s_l)$, $h'(s_l) = h(s_k)$ and $h'(s_i) = h(s_i)$ for $i \neq k, l$.

Moreover, it is easy to show [3] that by repeating this operation one eventually gets a pagination $\bar{h}$ that is equivalent to h_0 in the sense that $\pi_{\bar{h}} = \pi_{h_0}$. Hence we only have to show:

$$F(\text{FIFO}(M), h') \leqslant F(\text{FIFO}(M), h) . \tag{1}$$

To this end, let $h(s_k) = x_r$, $h(s_l) = x_s$ and $\Delta := \pi(s_l) - \pi(s_k)$. Then:

$h'(s_k) = x_s$, $h'(s_l) = x_r$ and

$$\pi_{h'}(x_r) = \pi_h(x_r) + \Delta, \pi_{h'}(x_s) = \pi_h(x_s) - \Delta, \pi_{h'}(x_i) = \pi_h(x_i) \qquad \text{for } i \neq r, s .$$

For $q = (y_1, ..., y_M) \in \overline{Q}$ we put:

$$P(h, q) = \prod_{i=1}^{M} \pi_h(y_i), S(h, q) = \sum_{i=1}^{M} \pi_h(y_i);$$

$$C(h) := \sum_{\overline{Q}} P(h, q), B(h) := \sum_{\overline{Q}} P(h, q) \cdot S(h, q),$$

and $\text{set}\,(q) := \{y_1, ..., y_M\}$.

With these abbreviations, one readily verifies that inequality (1) is equivalent to:

$$C(h) \cdot B(h') - C(h') \cdot B(h) \geqslant 0 .$$

We now partition the state-subset $\overline{Q}$ according to whether set(q) contains exactly one of x_r, x_s, both, or none of them:

$\overline{Q} = Q_1 \cup Q_2 \cup Q_3 \cup Q_4$, where

$$Q_1 = \{q | q \in \overline{Q}; x_r \in \text{set}(q), x_s \notin \text{set}(q)\} ,$$

$$Q_2 = \{q | q \in \overline{Q}; x_s \in \text{set}(q), x_r \notin \text{set}(q)\} ,$$

$$Q_3 = \{q | q \in \overline{Q}; x_r, x_s \in \text{set}(q)\} ,$$

$$Q_4 = \{q | q \in \overline{Q}; x_r, x_s \notin \text{set}(q)\} .$$

Furthermore, let Q_5 (resp. Q_6) be the set of $(M-1)$-tuples (resp. $(M-2)$-tuples) of pairwise distinct elements of $X - \{x_r, x_s\}$.

$C_i(h)$ and $B_i(h)$ $(i = 1, ..., 6)$ will denote the sums above with summation ranging over Q_i only.

Setting $p_1 := \pi_h(x_r)$, $p_2 := \pi_h(x_s)$, $p'_1 := \pi_{h'}(x_r)$ and $p'_2 := \pi_{h'}(x_s)$, the following relations hold for the quantities $C_i(h)$, $C_i(h')$, $B_i(h)$ and $B_i(h')$:

(I) $C_i(h) = C_i(h') =: C_i$ and

$$B_i(h) = B_i(h') =: B_i \qquad \text{for } i = 4, 5, 6\ ,$$

(II) $C_i(h) = p_i C_5$, $C_i(h') = p'_i C_5$ for $i = 1, 2$,

$$C_3(h) = p_1 p_2 C_6,\ C_3(h') = p'_1 p'_2 C_6\ ,$$

(III) $B_i(h) = p_i^2 C_5 + p_i B_5$,

$$B_i(h') = p_i'^2 C_5 + p'_i B_5 \qquad \text{for } i = 1, 2\ ,$$

$$B_3(h) = (p_1 + p_2) p_1 p_2 C_6 + p_1 p_2 B_6\ ,$$

$$B_3(h') = (p'_1 + p'_2) p'_1 p'_2 C_6 + p'_1 p'_2 B_6\ .$$

Since

$$C(h) = \sum_{i=1}^{4} C_i(h)\ ,\ B(h) = \sum_{i=1}^{4} B_i(h)$$

and similarly for h', a straightforward calculation yields:

$$C(h)B(h') - C(h')B(h) = Ud(2C_5 - aC_6 - B_6) + WdC_6\ ,$$

where we have employed the following abbreviations:

$$a := p_1 + p_2 = p'_1 + p'_2\ ; \qquad b := p_1 p_2\ ;$$

$$c := p_1^2 + p_2^2\ ; \qquad d := (p_1 - p_2 + \Delta)\Delta\ ;$$

$$U := C_4 + aC_5 + bC_6\ ;$$

$$W := B_4 + aB_5 + bB_6 + cC_5 + abC_6\ .$$

Since $d > 0$, it suffices to show:

$$R := 2UC_5 - aUC_6 - UB_6 + WC_6 \geqslant 0\ .$$

But:

$$R = 2C_4C_5 + B_4C_6 - C_4B_6 + a(2C_5^2 - C_4C_6 + B_5C_6 - C_5B_6)\ .$$

We are done if we can prove:

(i) $B_4C_6 \geqslant C_4B_6$;

(ii) $C_5^2 \geqslant C_4C_6$;

(iii) $B_5C_6 \geqslant C_5B_6$.

As an example, we will prove (ii), the proofs of the remaining two inequalities being similar. We have

$$C_5^2 = \sum_{(q_i, q_j) \in Q_5 \times Q_5} P(h, q_i) P(h, q_j)\ ;$$

$$C_4C_6 = \sum_{(q_i, q_j) \in Q_4 \times Q_6} P(h, q_i) P(h, q_j)\ .$$

Obviously, by definition of the Q_i, there is an injection from $Q_4 \times Q_6$ into $Q_5 \times Q_5$; hence any summand of C_4C_6 is also a summand of C_5^2, which proves (ii).

Thus $R \geqslant 0$ and inequality (1) holds. Taking into account the remarks preceding (1) we have thus established a formal proof of the claim stated informally at the conclusion of our introductory remarks.

Final remarks

1. The result proved above has a straightforward extension to the case of non-equal-sized sectors if one replaces sector-reference probabilities by "weighted" probabilities (i.e. probabilities divided by sector size, cf. [6]).

2. It has been pointed out in [8] that "given a sector-reference-pattern and changing the pagination" is formally equivalent to: "given a pagination and changing the sector-reference-pattern". Hence, our result may be considered as an exact formulation of a "locality principle": "Given a pagination, try to make references to each sector on the first page as likely as possible, then on the second page, and so on, in order to achieve optimal FIFO-performance!" Since the "frequency-order pagination" does not need to be optimal for LRU (cf. [4]) that very principle does not have to apply in this case. In this connection it is interesting to note that a similar situation exists for a strictly, deterministic reference-pattern, termed "simple loop": here, the references are over and over again to a sequence of addresses (as is the case with matrix-operations for instance); in this case, the "locality principle" that is valid for FIFO and for any realization of Belady's optimal algirithm as well [8], can be put as follows: "Given a pagination, choose a

sequencing of addresses that generates as many consecutive references to a single page as possible!" The same principle does not necessarily hold for LRU, as has also been shown in [8].

References

[1] A.V. Aho, P.J. Denning and J.D. Ullman, Principles of optimal page replacement, Assoc. Comput. Mach. 18 (1) (1971) 80–93.

[2] E.G. Coffman, P.J. Denning, Operating Systems Theory, Series in automatic computation (Prentice-Hall, Englewood Cliffs, NJ, 1973).

[3] A. Endres, Pagination für einfache Seitenwechselalgorithmen bei unabhängigen und identisch verteilten Adressreferenzen, Diploma Thesis, Darmstadt 1976, unpublished.

[4] R. Fagir, A counter-intuitive example of computer paging, IBM Research Report RC 5021, Yorktown Heights.

[5] D.J. Hatfield, J. Gerald, Program restructuring for virtual memory, IBM Systems J., 10 (3) (1971) 168–192.

[6] D.R. Innes, On the relationship of sector references, sector placement and paging performance, Proceedings GI-Annual Meeting 1976, Informatik Fachberichte 5 (Springer Verlag, Berlin, 1976).

[7] W.F. King (III), Analysis of paging-algorithms, IBM Research Report RC 3288, 1971.

[8] H.G. Stork, On the paging-complexity of periodic arrangements, to appear.