Journal of Engineering Mathematics, Vol. 10, No. 3, July 1976
Noordhoff International Publishing–Leyden
Printed in The Netherlands

Analysis of storage hierarchy

J. W. COHEN

Department of Mathematics, University of Utrecht, Utrecht, The Netherlands

E. W. B. VAN MARION

*Philips Electrologica, Apeldoorn, The Netherlands**

(Received October 10, 1975)

SUMMARY

The performance of storage hierarchies is an important problem in modern computer systems. This problem is studied here and an analytical method is developed to evaluate the memory sizes in "page on demand" systems for various page replacement procedures.

The first part of the study is devoted to a fundamental description of the statistical characteristics of program traces. The second part describes stochastic models, which provide the possibility to evaluate the operational behaviour of the storage system due to the statistical properties of the trace.

The method is applied to a real situation of virtual storage in an operating system. The theoretical results agree excellent with results obtained by measurements.

1. Introduction

The evaluation of operational characteristics of storage hierarchies is of great importance in computer design. In the present study a number of basic points concerning storage hierarchies will be analysed.

The simplest form of storage hierarchy, i.e. a two level system, is considered, but the methods to be developed are also applicable to more complicated hierarchies. A CPU (central processing unit) accesses a local store (LS) with a capacity of b pages during the execution of programs, which need altogether $v > b$ separate pages. The last ones are stored in main store (MS) and transported to LS on demand.

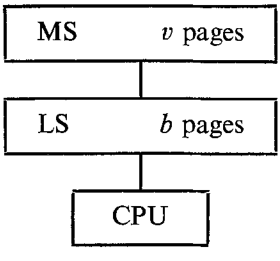

Figure 1.

* Present address: Unilever, Rotterdam, The Netherlands.

The behaviour of this system under various replacement procedures is investigated below. For a comprehensive description of these procedures see [1], [2], and [3].

The performance of the storage hierarchy depends on

i. the replacement procedures;
ii. the statistical properties of the program traces;
iii. the capacities and access times of the memories.

The first part of this study discusses those statistical properties of the trace, which are relevant for the determination of the page fault probability. The concepts of concentration and traffic are introduced.

The second part is devoted to the study of the LRU (least recently used), FIFO (first in first out) and RO (random out) page replacement procedures. For these disciplines analytical models are developed, which may be considered to be sufficiently close to the actual situation; exact and approximative formulas for the page fault probability are derived for these models. The statistics of the trace filtered by the LS and offered to the MS is investigated.

In Section (3.9) an example from practice is discussed, in particular, the page fault probability and the number of page swaps are determined. The results obtained so far show that the present approach is very promising.

The method developed starts from a number of assumptions. In the conclusion some questions and problems concerning the validity of these assumptions and other implications are indicated and may be subjects for further study.

2. On the statistics of the trace

2.1. *Characterization of the trace*

Traces are ordered sequences of pages, every page being element of a set of v separate pages.

If m is the number of occurrences of a certain page in a trace containing M pages, then the *frequency* p of this page for the given trace is defined as:

$$p = \frac{m}{M}. \tag{2.1}$$

The *concentration* c of this page for the given trace is defined by:

$$c = vp. \tag{2.2}$$

By $\xi(c)$ will be denoted the quotient of the number of separate pages with concentration not less than c in the given trace and the number v. Clearly $\xi(c)$ is a nonincreasing function of c,

$$\xi(0) = 1,\ \xi(\infty) = 0, \tag{2.3}$$

and $\xi(c) - \xi(c + \Delta c)$ is the fraction of separate pages in the trace with concentration not less than c and less than $c + \Delta c$.

We introduce the concept of traffic:

$$Q(c) \stackrel{\text{def}}{=} -\int_c^\infty \gamma\, d\xi(\gamma) \tag{2.4}$$

is for the given trace the traffic delivered by the pages with concentration not less than c. From (2.2), (2.4) and

$$dQ(c) = -c\,d\xi(c), \tag{2.5}$$

it follows immediately that $Q(0) = 1$.

Let $\boldsymbol{c}$ be a non-negative stochastic variable with distribution function

$$F(c) \stackrel{\text{def}}{=} \Pr\{\boldsymbol{c} < c\}, \tag{2.6}$$

and suppose that

$$F(0+) = 0 \text{ and } E(\boldsymbol{c}) = 1; \tag{2.7}$$

$F(.)$ is assumed to be continuous from the left.

If we take

$$\xi(c) = 1 - F(c), \tag{2.8}$$

then

$$Q(c) = -\int_c^\infty \gamma d(1 - F(\gamma)),\ Q(0) = E(\boldsymbol{c}) = 1. \tag{2.9}$$

Consequently, as far as the concentrations of the various types of pages in a given trace are concerned, it suffices to specify the distribution $F(c)$ with mean equal to one.

We discuss a few examples.

Example 2.1. There are 120 separate pages, so $v = 120$, of which, for a given trace, 20 ones have each a frequency $p = 0.04$ and the remaining 100 pages each have frequency 0.002. Consequently, for this trace the concentration $\boldsymbol{c}$ has only two values, viz. 4.8 and 0.24, and

$$\begin{aligned} 1 - F(c) = \xi(c) &= 1 \text{ for } && c < 0.24, \\ &= \tfrac{1}{6} \text{ ,, } && 0.24 \leqq c < 4.8, \\ &= 0 \text{ ,, } && 4.8 \leqq c, \end{aligned}$$

whereas

$$\begin{aligned} Q(c) &= 1 \quad \text{for} && c < 0.24, \\ &= 0.8 \text{ ,, } && 0.24 \leqq c < 4.8, \\ &= 0 \text{ ,, } && 4.8 \leqq c. \end{aligned}$$

The graphs for $\xi(c)$ and $Q(c)$ are given in Figure 2.

In this example 80% of the traffic is originated by 20% of the pages (those with $c = 8$), the remaining 20% of the traffic items forms 80% of the pages.

Example 2.2. Here it is assumed that $\boldsymbol{c}$ has a negative exponential distribution, so that $1 - F(c) = e^{-c}$, and hence

$$Q(c) = \int_c^\infty e^{-\gamma}\gamma d\gamma = (c + 1)e^{-c}.$$

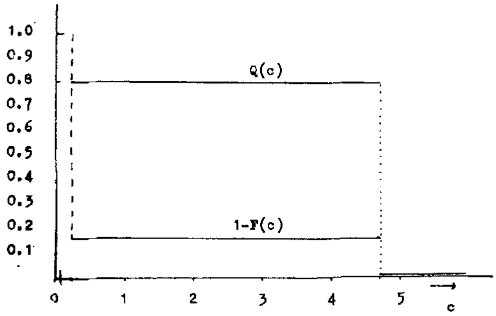

Figure 2.

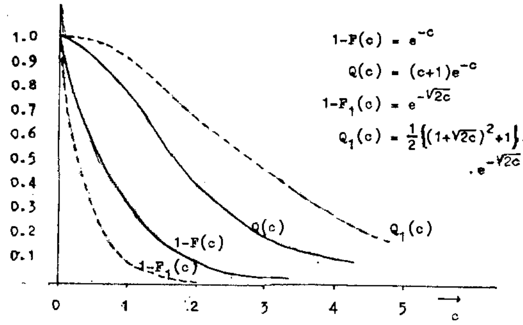

Figure 3.

The graphs for these functions are shown in Figure 3 (full lines). From the curves it is seen that 10% of the pages has a concentration not les than 2.5 and that about 33% of the total traffic is delivered by these pages.

Example 2.3. Here we take

$$1 - F_1(c) = e^{-\sqrt{2c}},$$

and

$$Q_1(c) = \tfrac{1}{2}\{(1 + \sqrt{2c})^2 + 1\}\, e^{-\sqrt{2c}}.$$

The graphs are also shown in Figure 3. It is seen that 10% of the pages have concentrations larger than 1.2; they deliver about 89% of the traffic.

2.2. *Mixing of traces*

Suppose two traces type 1 and type 2 with distributions $F_1(c)$ and $F_2(c)$ for the concentration are given, these traces are composed to a single trace as follows.

In some order the CPU chooses the successive pages to be processed from the traces of type 1 and type 2 until all pages of both traces have been processed. Since $dF_1(c)$ ($dF_2(c)$) is the fraction of pages with concentration $\boldsymbol{c}$ such that $c \leqq \boldsymbol{c} \leqq c + dc$ for trace of type 1 (type 2), it is seen that for the resulting trace

$$\tfrac{1}{2}dF_1(c) + \tfrac{1}{2}dF_2(c)$$

is the fraction of pages with concentration $\boldsymbol{c}$ $(c, c + dc)$.

Hence for the resulting trace the distribution $F_3(c)$ is given by

$$F_3(c) = \tfrac{1}{2}F_1(c) + \tfrac{1}{2}F_2(c),$$

so that

$$Q_3(c) = \tfrac{1}{2}Q_1(c) + \tfrac{1}{2}Q_2(c);$$

more generally, if the resulting trace is such that trace of type 1 is processed n_1 times and that of type 2 n_2 times then for the resulting trace

$$F_3(c) = \frac{n_1}{n_1 + n_2} F_1(c) + \frac{n_2}{n_1 + n_2} F_2(c),$$
$$Q_3(c) = \frac{n_1}{n_1 + n_2} Q_1(c) + \frac{n_2}{n_1 + n_2} Q_2(c). \tag{2.10}$$

Similarly, the resulting distribution of the concentration c can be found if a trace is composed of a number of traces.

Note. If traces with the same distribution of concentration are mixed the resulting trace has also the same distribution for its concentration.

2.3. *Truncating of traces*

Whenever from a given trace certain pages are deleted, so that also the set v is diminished, a new trace results, a so called "truncated trace". It is of some interest to know the relation between the distributions of the concentration in the original and the truncated trace. We discuss here an example showing how such a relation can be obtained.

From the original trace with distribution $F(c)$ of the concentration all pages with concentration less than or equal to c_0 are deleted. Denote by $F_t(c)$ the distribution for the truncated trace, be further

$$Q_0 = Q(c_0) = \int_{c_0}^{\infty} c\, dF(c),$$
$$\xi_0 = \xi(c_0) = 1 - F(c_0). \tag{2.11}$$

The total number v_t of separate pages, which may occur in the truncated trace, is given by

$$v_t = v\xi_0, \tag{2.12}$$

and a page with concentration $c > c_0$ and frequency p in the original trace will have a frequency

$$p_t = \frac{p}{Q_0}, \tag{2.13}$$

in the truncated trace. Hence, from (2.12), (2.13) and (2.2) it is seen that a page with concentration c in the original trace has concentration c_t in the truncated trace, with c_t given by

$$c_t = 0 \qquad \text{if } c \leqq c_0,$$
$$= \frac{\xi_0}{Q_0} c \text{ if } c > c_0. \tag{2.14}$$

Consequently

$$\Pr\left\{\frac{Q_0}{\xi_0} c_t > c\right\} = 1 \qquad \text{for } c \leqq c_0, \tag{2.15}$$
$$= \Pr\{c > c \mid c > c_0\} = \frac{1 - F(c)}{1 - F(c_0)} \text{ for } c > c_0.$$

Obviously

$$E(c_t Q_0/\xi_0) = \frac{1}{1 - F(c_0)} \int_{c_0}^{\infty} c\,dF(c) = Q_0/\xi_0,$$

so that $E(c_t) = 1$.

Further

$$\begin{aligned} Q_t(c) &= \frac{1}{1 - F(c_0)} \int_{c}^{\infty} z\,dF(Q_0 z/\xi_0) \\ &= \frac{1}{Q_0} \int_{c}^{\infty} Q_0 x/\xi_0\, dF(x) \\ &= \frac{1}{Q_0} Q(Q_0 c/\xi_0) \text{ for } Q_0 c/\xi_0 > c_0. \end{aligned} \tag{2.16}$$

Analogous relations are obtained if the original trace is truncated by deleting all pages with concentration $c > c_1$.

3. Page fault probability

3.1. The replacement procedures

If the next page to be processed by the CPU is not present in the LS, then it has to be brought forward from the MS to the LS and at the same time a page has to leave the LS to free a place in the LS for the page desired, if all b places of the LS are occupied.

The replacement is called:
Random Out (RO): If the page to leave the LS is chosen at random out of the pages present in the LS;
Least Recently Used (LRU): If the page to leave the LS is the one, which has been used least recently of all pages in the LS;
First In First Out (FIFO): If the page to leave the LS is the one whose sojourn time in the LS is largest of all pages present in the LS.

In the next sections we shall analyse these replacement procedures and derive expressions for the page fault probability, i.e. the probability that the next page to be processed is not present in the LS. The analysis is performed for a trace consisting of v separate pages, which are assumed to be numbered from 1 to v, and for a LS with a capacity of b pages. We shall denote by p_i the frequency of occurrence of page i in the trace.

The basic assumption used in the analysis concerns the sequence of pages in the trace, viz. it is assumed that the pages still to be processed are stochastically independent of the pages already processed. This assumption seems to be reasonable as a first approximation, particularly if the length of the string of dependent pages is smaller than b, see also conclusion.

Starting from these assumption the storage process is modelled as a vector Markov chain. It turns out that this chain can be analyzed completely. However, the numerical evaluation of the formulas so obtained is extremely tedious, except in some rather simple cases. Approximation of this vector chain by means of a simpler Markov chain leads to formulas, which are easy to handle and very accurate, as will be shown.

Since the vector Markov chain has a rather complex structure we shall first discuss the approximative Markov chains. The discussion of the former one is delayed to sections 3.6 and 3.7.

3.2. The LRU procedure

We consider a Markov chain with state space consisting of $b + 1$ states $E_1, E_2, \ldots, E_{b+1}$, see Figure 4.

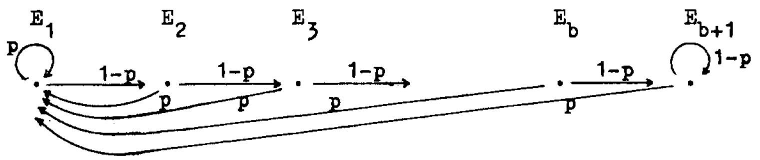

Figure 4.

The one-step transition probabilities p_{ij} are indicated in Figure 4, i.e.

$$\begin{aligned} p_{ij} &= 1 - p,\ j = 1, \qquad & i = 1, \ldots, b + 1; \\ &= 1 - p,\ j = i + 1, \qquad & i = 1, \ldots, b; \\ &= 1 - p,\ j = i = b + 1; & \\ &= 0,\ \text{elsewhere}; & \end{aligned} \tag{3.1}$$

with $0 < p < 1$.

Obviously this Markov chain has a stationary distribution u_j, $j = 1, \ldots, b + 1$ and it is readily found that

$$\begin{aligned} u_j &= p(1 - p)^{j-1},\ j = 1, \ldots, b; \\ &= (1 - p)^b, \qquad j = b + 1. \end{aligned} \tag{3.2}$$

The states $E_1, \ldots, E_b$ correspond with the b positions for pages in the local store. Suppose the next page of the trace, say page k, to be processed has frequency p_k in the trace and is not in the LS, i.e. it is in the MS, which corresponds to state E_{b+1} in the Markov chain; this situation will then correspond to a transition with probability $p = p_k$ from state E_{b+1} to E_1, the entrance state of the LS. If, however, this page k was already in the LS, say at state E_3, then a transition with probability p occurs from state E_3 to E_1.

Suppose that the next page of the trace to be processed is not page k, so that, if at this moment page k is not located in the LS but in the MS, then it stays in the MS, which corresponds with a transition from E_{b+1} to E_{b+1} in the Markov chain with probability $1 - p$, whereas if page k is in the LS, say at E_2, then it is assumed to pass to state E_3 with probability $1-p$, and if it is at state E_b then it leaves the LS for the MS (E_{b+1}) with probability $1-p$.

To every page of the set of v separate pages we let correspond a Markov chain as described above, with p equal to the frequency of the page in the trace. Due to the structure of this Markov chain, *a* page, being processed, moves to state E_1; it moves one state to the right when it is not processed, and it ultimately leaves E_b and goes to E_{b+1}, i.e. to the MS, and consequently the page least recently used moves with the higher speed through the LS.

Obviously, the present model of v parallel Markov chains for the description of the motion of the pages through the LS is an approximation of the real situation. For the description of the real situation the appropriate model is a vector Markov chain. The main objection to the approximative model is that pages leave the LS, when there is in the actual situation no need to leave the LS; this is due to the fact that the v Markov chains act independently of one another. The influence of this fact on the page fault probability can, however, be eliminated for the greater part, as it will be shown below.

From (3.2) it follows that $(1-p_i)^b$ is for the stationary situation the probability that page number i is not in the LS. Consequently, the probability that the next page to be processed is page i and that it is not in the LS, i.e. page i causes a page fault, is equal to

$$p_i(1-p_i)^b.$$

Hence the page fault probability a is given by

$$a=\sum_{i=1}^{v} p_i(1-p_i)^b. \tag{3.3}$$

Since $1-(1-p_i)^b$ is the probability that page number i is in one of the states $E_1, \ldots, E_b$, it follows that for the present model

$$g=\sum_{i=1}^{v} \{1-(1-p_i)^b\} \tag{3.4}$$

is the average number of pages in the LS.

Since

$$\sum_{i=1}^{v} p_i = 1,$$

and

$$(1-p_i)^b > 1-bp_i, \qquad 1>p_i>0,$$

it is seen from (3.4) that

$$g<b. \tag{3.5}$$

A result, which is intuitively clear, since as already pointed out above, the v Markov chains act independently of one another so that pages may leave the LS although there is no need for.

To compensate this fact we replace the capacity b of the LS by a greater (fictitious) one b_0, with b_0 determined by the condition

$$b = \sum_{i=1}^{v} 1 - (1 - p_i)^{b_0}, \tag{3.6}$$

and take in our model $b_0 + 1$ states in the Markov chain instead of $b + 1$. Hence, with the model so adjusted, the average number of pages in the LS equals b, and the page fault probability α is now determined by

$$\alpha = \sum_{i=1}^{v} p_i(1 - p_i)^{b_0}. \tag{3.7}$$

Obviously

$$\alpha < a. \tag{3.8}$$

Let x_h denote the fraction of the v separate pages, all with the same frequency q_h, $h = 1, \ldots, n$, so that

$$\begin{aligned} &\sum_{h=1}^{n} x_h = 1, \ \sum_{h=1}^{n} v x_h q_h = 1, \\ &c_h = v q_h, \ h = 1, \ldots, n, \quad \Pr\{\boldsymbol{c} = c_h\} = x_h, \ h = 1, \ldots, n. \end{aligned} \tag{3.9}$$

Hence

$$\begin{aligned} a &= \sum_{h=1}^{n} v x_h q_h (1 - q_h)^b = \sum_{h=1}^{n} c_h \left(1 - \frac{c_h}{v}\right)^b \Pr\{\boldsymbol{c} = c_h\}, \\ g &= \sum_{h=1}^{n} v \left\{1 - \left(1 - \frac{c_h}{v}\right)^b\right\} \Pr\{\boldsymbol{c} = c_h\}. \end{aligned} \tag{3.10}$$

Define

$$d = \frac{b}{v}, \ d_0 = \frac{b_0}{v}, \ \gamma = \frac{g}{v}. \tag{3.11}$$

For obvious reasons d will be called the *compression* of the MS–LS combination. For v sufficiently large, which is actually acceptable, we obtain from (3.3)

$$\begin{aligned} a &= \sum_{h=1}^{v} c_h \mathrm{e}^{-dc_h} \Pr\{\boldsymbol{c} = c_h\} = E(\boldsymbol{c}\mathrm{e}^{-d\boldsymbol{c}}), \\ &= \sum_{h=1}^{v} (1 - \mathrm{e}^{-dc_h}) \Pr\{\boldsymbol{c} = c_h\} = 1 - E(\mathrm{e}^{-d\boldsymbol{c}}). \end{aligned} \tag{3.12}$$

In the same way we obtain from (3.6) and (3.7)

$$\alpha = E(\boldsymbol{c}\mathrm{e}^{-d_0\boldsymbol{c}}) = \int_0^\infty c\mathrm{e}^{-d_0 c} dF(c), \tag{3.13}$$

$$d = 1 - E(\mathrm{e}^{-d_0\boldsymbol{c}}) = 1 - \int_0^\infty \mathrm{e}^{-d_0 c} dF(c), \tag{3.14}$$

with

$$\Pr\{\boldsymbol{c} < c\} = F(c) \text{ and } E(\boldsymbol{c}) = \int_0^\infty c dF(c), \tag{3.15}$$

where $F(c)$ is the distribution of the concentration $\boldsymbol{c}$ for the given trace (cf. 2.6).

For given compression d and distribution $F(.)$ of the concentration d_0 is found from (3.14) and then the page fault probability from (3.13).

Example 3.1. (Cf. example 2.2).

For $F(c) = 1 - \mathrm{e}^{-c}$, $c \geqq 0$, it follows that

$$\alpha = \frac{1}{(1+d_0)^2}, \; d = 1 - \frac{1}{1+d_0},$$

so that

$$\alpha = (1-d)^2. \tag{3.16}$$

So for $d = 0.0905$ it follows that $\alpha = 0.82$.

Example 3.2. (Cf. example 2.3).

For $F(c) = 1 - \mathrm{e}^{-\sqrt{2c}}$, $c \geqq 0$, it follows

$$\begin{aligned} \alpha &= \frac{1}{2}\left\{-\frac{1}{d_0^2} + \frac{1}{d_0^2}\sqrt{2d_0}\left(1+\frac{1}{d_0}\right)\mathrm{e}^{1/2d_0}\frac{\sqrt{\pi}}{2}\left(1 - \frac{2}{\sqrt{\pi}}\int_0^{1/\sqrt{2d_0}} \mathrm{e}^{-y^2} dy\right)\right\}. \\ d &= 1 - \frac{2}{\sqrt{2d_0}}\mathrm{e}^{1/2d_0}\frac{\sqrt{\pi}}{2}\left\{1 - \frac{2}{\sqrt{\pi}}\int_0^{1/\sqrt{2d_0}} \mathrm{e}^{-y^2} dy\right\}, \end{aligned} \tag{3.17}$$

so that

$$\alpha = \frac{1}{2}\,\frac{d_0 - d - dd_0}{d_0^2}.$$

For $d_0 = 0.125$ it follows $d = 0.0943$, so that $\alpha = 0.585$.

Since α is the page fault probability it is seen that of the total traffic $E(\boldsymbol{c}) = 1$ offered by the trace, the part

$$\alpha = \int_0^\infty c\mathrm{e}^{-d_0 c} dF(c)$$

is offered to the MS, whereas the LS handles the part

$$1 - \alpha = \int_0^\infty c(1 - \mathrm{e}^{d_0 c}) dF(c).$$

The presence of the LS leads therefore to a *demixing* of the trace. In a more accurate form this may be stated as follows. Due to the presence of the LS the original trace is split up in two traces, that handled by the MS with distribution of $\boldsymbol{c}$ given by

$$F_m(c) = \frac{1}{1-d}\int_0^c \mathrm{e}^{-d_0\gamma} dF(\gamma),$$

and that handled by the LS with distribution of c given by

$$F_l(c) = \frac{1}{d}\int_0^c (1 - e^{-d_0\gamma})\,dF(\gamma).$$

Mixing of these traces with mixing weights $1 - d$ and d, respectively, yields the distribution of c in the original trace,

$$F(c) = (1 - d)F_m(c) + dF_l(c), \tag{3.18}$$

i.e. the mixing weights are proportional to $v - b$ and b.

In the derivation of the relations (3.13) and (3.14) the interaction of the various pages in the LS has been taken into account by introducing the factor d_0. There is, however, another way to take this effect into account.

From the results of the Markov chain model it is seen that the density of the concentration c in the original trace and the trace offered to the MS is reduced by a factor proportional to

$$e^{-dc}. \tag{3.19}$$

The exponential type of this factor is evidently characteristic for the LRU procedure, however, to take account of the interaction of the pages in the LS the factor d in (3.19) has to be replaced by d_0, determined by (3.14). This factor d represents the compression, however, since always b pages should be present in the LS there is no need to reserve room for them in the MS, so that the actual compression is equal to $b/(v - b) = d/(1 - d)$. Hence, on behalf of (3.12), we might determine the page fault probability from

$$\alpha = E(ce^{-dc/(1-d)}). \tag{3.20}$$

Yet another way of reasoning is that, if $\sqrt{\alpha}$ is the page fault probability, only room should be reserved in the MS for $v - (1 - \alpha)v$ pages, so that the actual compression is

$$\frac{b}{\alpha v} = \frac{1}{\alpha} d,$$

and hence of behalf of (3.12)

$$\alpha = E(ce^{-(dc/\alpha)}). \tag{3.21}$$

Example 3.3. If $F(c) = 1 - e^{-c}$ then $\alpha = (1 - d)^2$, according to (3.20), so that we get the same value for the page fault probability as calculated according to (3.13) and (3.14) (cf. example 3.1).

If we apply (3.21) then

$$\alpha = \frac{1}{(1 + d/\alpha)^2},$$

so that

$$\alpha^2 - (1 - 2d)\alpha + d^2 = 0.$$

Therefore

$$\alpha = \frac{1 - 2d + \sqrt{1 - 4d}}{2} = 1 - 2d - d^2 + o(d^3), \quad d \to 0,$$

which differs from $(1 - d)^2$ by $2d^2$, which is small.

Obviously the relation (3.20) is the easier one for calculation. Similarly, as before, we can determine the distributions of the concentration of the traces, which are handled by the MS and by the LS, for the case, which led to relation (3.20), as well as for the case, which yielded (3.21), and consider the original trace as a mixing of these traces. However, the mixing weights do not have such an intuitive meaning as before. Moreover, the relations (3.13) and (3.14) merit preference over (3.20) and (3.21), since for

$$\begin{aligned} F(c) &= 0 \text{ for } c \leqq 1, \\ &= 1 \text{ for } c > 1, \end{aligned}$$

it follows from (3.13) and (3.14) that

$$\alpha = 1 - d,$$

which is evidently the true value of $\sqrt{\alpha}$, as it follows from symmetry. From (3.20) and also from (3.21), however, we obtain

$$\alpha = 1 - d + o(d^2) \text{ for } d \to 0.$$

Finally it is noted that (3.13) and (3.14) show that the derivative of d with respect to d_0 is equal to α, a property which is useful in obtaining numerical results.

3.3. *The FIFO procedure*

As in Section 3.2 we consider a Markov chain with state space consisting of $b + 1$ states $E_1, \ldots, E_{b+1}$, see Figure 5.

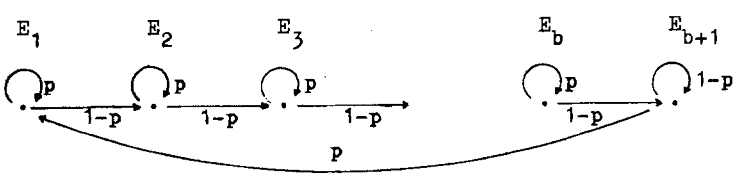

Figure 5.

The one-step transition probabilities p_{ij} are given by

$$\begin{aligned} p_{ij} &= p, & i &= j, & i &= 1, \ldots, b; \\ &= 1 - p, & i &= j = b + 1; \\ &= 1 - p, & j &= i + 1, & i &= 1, \ldots, b; \\ &= p, & i &= b + 1, & j &= 1; \\ &= 0, & &\text{elsewhere}; \end{aligned} \tag{3.22}$$

with $0 < p < 1$. Evidently, this Markov chain has a stationary distribution u_j, $j = 1, \ldots, b+1$, which is found to be

$$\begin{aligned} u_j &= \frac{p}{1+(b-1)p}, \quad j = 1, \ldots, b; \\ &= \frac{1-p}{1+(b-1)p}, \quad j = b+1. \end{aligned} \tag{3.23}$$

As in Section 3.2 the states $E_1, \ldots, E_b$ correspond with the b positions for pages in the LS. From that section and the description of the FIFO procedure in Section 3.1 it is readily seen that this Markov chain may be used to describe in an approximative sense the motion of a page through the LS in the case of the FIFO procedure. For this model it is seen from (3.23) that the probability that the next page to be processed is page i and that page i causes a page fault is given by

$$\frac{p_i(1-p_i)}{1+(b-1)p_i},$$

and hence the page fault probability is given by

$$a = \sum_{i=1}^{v} \frac{p_i(1-p_i)}{1+(b-1)p_i}. \tag{3.24}$$

The average number g of pages in the LS is given by

$$g = \sum_{i=1}^{v} \left(1 - \frac{1-p_i}{1+(b-1)p_i}\right) = \sum_{i=1}^{v} \frac{bp_i}{1+(b-1)p_i}; \tag{3.25}$$

obviously

$$g < b. \tag{3.26}$$

To compensate the fact that $g < b$, we introduce as before a fictitious LS-capacity b_0 defined by

$$b = \sum_{i=1}^{v} b_0 \frac{p_i}{1+(b_0-1)p_i}, \tag{3.27}$$

and calculate the page fault probability from

$$\alpha = \sum_{i=1}^{v} \frac{p_i(1-p_i)}{1+(b_0-1)p_i}. \tag{3.28}$$

With x_h, c_h, q_h, d and d_0 as defined before (cf. (3.9)) we obtain

$$\begin{aligned} \alpha &= \sum_{h=1}^{n} x_h c_h \frac{(1-c_h/v)}{1+(b_0-1)c_h/v}, \\ d &= \sum_{h=1}^{n} x_h \frac{c_h}{1+(b_0-1)c_h/v}. \end{aligned} \tag{3.29}$$

For $v \to \infty$ it follows

$$\alpha = E\left(\frac{\mathbf{c}}{1 + d_0 \mathbf{c}}\right), \tag{3.30}$$

$$d = d_0 E\left(\frac{\mathbf{c}}{1 + d_0 \mathbf{c}}\right) = 1 - E\left(\frac{1}{1 + d_0 \mathbf{c}}\right) \tag{3.31}$$

$$= \alpha d_0,$$

from which it follows

$$\alpha = E\left(\frac{\mathbf{c}}{1 + d\mathbf{c}/\xi}\right). \tag{3.32}$$

The traffic offered to the MS is given by

$$\alpha = \int_0^\infty \frac{c}{1 + d_0 c}\, dF(c), \tag{3.33}$$

and that handled by the LS is

$$1 - \alpha = \int_0^\infty c \frac{d_0 c}{1 + d_0 c}\, dF(c) = d_0 \alpha. \tag{3.34}$$

With $F_m(c)$ and $F_l(c)$ as defined in the previous section we have

$$F_m(c) = \frac{1}{1 - d} \int_0^c \frac{1}{1 + d_0 \gamma}\, dF(\gamma), \tag{3.35}$$

$$F_l(c) = \frac{1}{d} \int_0^c \frac{d_0 \gamma}{1 + d_0 \gamma}\, dF(\gamma). \tag{3.36}$$

Mixing of the trace offered to the MS and that handled by the LS with mixing weights $1 - d$ and d, respectively, yields again the distribution $F(c)$ of the concentration $\mathbf{c}$ of the original trace

$$F(c) = (1 - d)F_m(c) + dF_l(c). \tag{3.37}$$

Just like in the preceding section, where the formulas (3.20) and (3.21) have been derived to take into account the interaction of the pages in the LS, we obtain here for the FIFO procedure

$$\alpha = E\left(\frac{\mathbf{c}}{1 + d\mathbf{c}/(\alpha - 1)}\right), \tag{3.38}$$

and

$$\alpha = E\left(\frac{\mathbf{c}}{1 + d\mathbf{c}/\alpha}\right). \tag{3.39}$$

Note that the last relation is identical with (3.32).

3.4. *The RO procedure*

Again we consider a Markov chain with state space consisting of $b + 1$ states $E_1, \ldots, E_{b+1}$, see Figure 6.

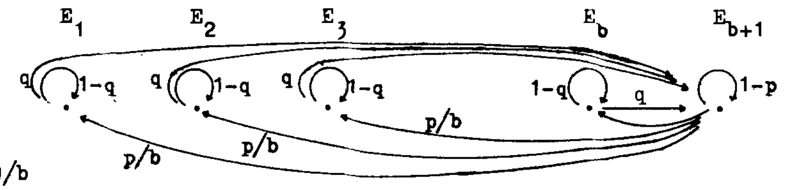

Figure 6.

The one-step transition probabilities are given by

$$\begin{aligned} p_{ij} &= 1 - (1-p)/b, \quad i = j, && i = 1, \ldots, b; \\ &= 1 - p, \quad i = j = b + 1; && \\ &= (1-p)/b, \quad j = b + 1, && i = 1, \ldots, b; \\ &= p/b, && j = 1, \ldots, b; \\ &= 0, \quad \text{elsewhere}; && \end{aligned}$$

with $0 < p < 1$. This Markov chain has a stationary distribution u_j, $j = 1, \ldots, b + 1$, which is given by

$$\begin{aligned} u_j &= \frac{p}{1 + (b-1)p}, \quad j = 1, \ldots, b, \\ &= \frac{1-p}{1 + (b-1)p}, \quad j = b + 1. \end{aligned}$$

Obviously this stationary distribution is identical with that of the Markov chain for the FIFO procedure. Consequently, all results obtained for the FIFO procedure in the preceding section apply also for the Random Out procedure.

3.5. *The vector Markov chain*

Consider a b-component vector Markov chain $\tilde{x}_n$, $n = 1, 2, \ldots$ with stationary transition probabilities. The state space $S^{(i)}$ of every component $x_n^{(i)}$ of the vector

$$\tilde{x}_n \stackrel{\text{def}}{=} (x_n^{(1)}, \ldots, x_n^{(b)})$$

is the set of integers $1, \ldots, v$ for $i = 1, \ldots, b$. The state space S of $\tilde{x}_n$ is the subset of the product space of $S^{(1)}, \ldots, S^{(b)}$, such that no two components of $\tilde{x}_n$ are equal. Hence S contains $v!/(v - b)!$ different states.

The one-step transition probabilities are defined

$$p_{\tilde{\imath}\tilde{\jmath}} \stackrel{\text{def}}{=} \Pr\{\tilde{x}_{n+1} = \tilde{\jmath} \mid \tilde{x}_n = \tilde{\imath}\}, \text{ for any } \tilde{\imath} \text{ and } \tilde{\jmath} \in S.$$

This Markov chain is used to describe the motion of pages in the storage hierarchy as follows. The content of the LS is described by the vector $\tilde{\imath} = (i^{(1)}, \ldots, i^{(b)})$, where the k-th component denotes the number of the page that occupies place k in the LS.

A page demand corresponds with a state transition. The matrix of transition probabilities therefore depends on the replacement procedure of the storage hierarchy and will be specified below for the LRU, FIFO and RO procedures.

i. The LRU procedure.

$$\Pr\{\tilde{x}_{n+1} = (i^{(k)}, i^{(1)}, \ldots, i^{(k-1)}, i^{(k+1)}, \ldots, i^{(b)}) \mid \tilde{x}_n = \tilde{\imath}\} = p_{i^{(k)}} \text{ for } k = 1, \ldots, b; \quad (3.40)$$

$$\Pr\{\tilde{x}_{n+1} = (i^{(k)}, i^{(1)}, \ldots, i^{(b-1)}) \mid \tilde{x}_n = \tilde{\imath}\} = p_{i^{(k)}} \text{ for } k = b+1, \ldots, v; \quad (3.41)$$

$p_{\tilde{\imath}\tilde{\jmath}} = 0$ for other $\tilde{\imath}$ and $\tilde{\jmath} \in S$.

ii. The FIFO procedure.

$$\Pr\{\tilde{x}_{n+1} = \tilde{\imath} \mid \tilde{x}_n = \tilde{\imath}\} = p_{i^{(k)}} \text{ for } k = 1, \ldots, b; \quad (3.42)$$

$$\Pr\{\tilde{x}_{n+1} = (i^{(k)}, i^{(1)}, \ldots, i^{(b-1)}) \mid \tilde{x}_n = \tilde{\imath}\} = p_{i^{(k)}} \text{ for } k = b+1, \ldots, v; \quad (3.43)$$

$p_{\tilde{\imath}\tilde{\jmath}} = 0$ for all other $\tilde{\imath}$ and $\tilde{\jmath} \in S$.

iii. The RO procedure.

$$\Pr\{\tilde{x}_{n+1} = \tilde{\imath} \mid \tilde{x}_n = \tilde{\imath}\} = p_{i^{(k)}} \text{ for } k = 1, \ldots, b; \quad (3.44)$$

$$\Pr\{\tilde{x}_{n+1} = (i^{(k)}, i^{(1)}, \ldots, i^{(j-1)}, i^{(j+1)}, \ldots, i^{(b)}) \mid \tilde{x}_n = \tilde{\imath}\} = \frac{1}{b} p_{i^{(k)}} \quad (3.45)$$

for $k = b+1, \ldots, v$ and $j = 1, \ldots, b$;

$p_{\tilde{\imath}\tilde{\jmath}} = 0$ for other $\tilde{\imath}$ and $\tilde{\jmath} \in S$.

Applying this model we take the actual value of the transition probability $p_{i^{(k)}}$ equal to the frequency of occurrence of page $i^{(k)}$ in the program trace.

As the state space S is finite and the Markov chain is irreducible and aperiodic it is an ergodic chain. Hence, there exists a unique stationary distribution, which we denote in the following way:

$$u(\tilde{\imath}) \stackrel{\text{def}}{=} \lim_{n \to \infty} \Pr\{\tilde{x}_n = \tilde{\imath}\}, \quad (3.46)$$

where again $\tilde{\imath} = (i^{(1)}, \ldots, i^{(b)})$.

The norming condition reads

$$u(\tilde{\imath}) = 1.$$

3.6. The exact formulas for the page fault probability

By means of the one-step transition probabilities for the vector Markov chain the equilibrium equations of state can easily be constructed. Of course they are different for the various replacement procedures.

i. The LRU procedure.
Expressed in terms of (3.46) the equations of state read

$$((1 - p_{i^{(1)}})u(\tilde{\imath}) = p_{i^{(1)}} \sum_{j=2}^{b-1} u(i^{(2)}, \ldots, i^{(j)}, i^{(1)}, i^{(j+1)}, \ldots, i^{(b)}) + \\ + \sum_{j=b+1}^{v} u(i^{(2)}, \ldots, i^{(b)}, i^{(j)}) \text{ for every } \tilde{\imath} \in S. \tag{3.47}$$

The solution of (3.47) for $\tilde{\imath} \in S$ reads

$$u(\tilde{\imath}) = \frac{p_{i^{(1)}}}{1 - p_{i^{(1)}}} \frac{p_{i^{(2)}}}{1 - p_{i^{(1)}} - p_{i^{(2)}}} \cdots \frac{p_{i^{(b-1)}}}{1 - p_{i^{(1)}} - \ldots - p_{i^{(b-1)}}} p_{i^{(b)}}. \tag{3.48}$$

For sake of simplicity a derivation of this solution is deleted, but it is easily verified by substitution that (3.48) satisfies (3.47) and the norming condition. This solution has been also obtained in [5], see also [6].

ii. The FIFO procedure.
In this case the system of equations of state for $\tilde{\imath} \in S$ reads

$$\sum_{j=b+1}^{v} p_{i^{(j)}} u(\tilde{\imath}) = p_{i^{(1)}} \sum_{j=b+1}^{v} u(i^{(2)}, \ldots, i^{(b)}, i^{(j)}). \tag{3.49}$$

The stationary distribution $u(\tilde{\imath})$ is presented by

$$u(\tilde{\imath}) = C \prod_{j=1}^{b} p_{i^{(j)}}, \qquad \tilde{\imath} \in S, \tag{3.50}$$

with C being determined by the norming condition.

iii. The RO procedure.
The equations of state for $\tilde{\imath} \in S$ are

$$u(\tilde{\imath}) = \sum_{j=1}^{b} p_{i^{(j)}} u(\tilde{\imath}) + \frac{1}{b} \sum_{k=b+1}^{v} u(i^{(1)}, \ldots, i^{(j-1)}, i^{(j+1)}, \ldots, i^{(b)}, i^{(k)}). \tag{3.51}$$

The solution is the same as that for the FIFO procedure.

A page fault occurs when a page to be processed is not in the LS. The probability M_k that page k is not in the LS is presented by

$$M_k = \sum_{j=1}^{b} \sum_{i^{(j)}=1, \neq k}^{v} u(\tilde{\imath}). \tag{3.52}$$

The page fault probability α therefore is expressed by

$$\alpha = \sum_{k=1}^{v} p_k M_k. \tag{3.53}$$

The expressions (3.52) and (3.53) are hard to evaluate, even for simple distributions of the page frequency p_i. However these simple cases can be used to evaluate the accuracy of the results of the approximative method by comparing them with those of the exact one.

3.7. *Comparison of exact and approximative methods*

In the following α_e and α_a denote the page fault probabilities determined according to the exact and approximative method respectively.

Two cases are considered; the formulas (3.6), (3.7), (3.27) and (3.28) are used for the evaluation of α_a and the formulas (3.48), (3.50), (3.52) and (3.53) for that of α_e.

i. All v pages have equal page frequency p. Using the fact that $vp = 1$ it follows that

$\alpha_e = \alpha_a = 1 - bp$ for LRU, FIFO and RO.

Hence the approximative method gives exact results in this case.

ii. One page has a frequency p_1 and $v - 1$ pages have a frequency p_2 and $p_1 \gg p_2$. This case is very unfavourable for the approximative method, since many empty places are introduced in the LS (see Section 3.2).

Numerical results for the LRU procedure are listed in the table below.

$v = 100$	$p_1 = 0.5$ $p_2 = 0.00505$		$p_1 = 0.9$ $p_2 = 0.00101$	
b	α_e	α_a	α_e	α_a
2	0.61806	0.58938	0.10790	0.09899
4	0.51484	0.49203	0.09706	0.09697
8	0.46632	0.46467	0.09293	0.09293
16	0.42425	0.42424		
$v = 400$	$p_1 = 0.5$ $p_2 = 0.00125$		$p_1 = 0.9$ $p_2 = 0.00025$	
b	α_e	α_a	α_e	α_a
2	0.62328	0.59423	0.10873	0.09975
4	0.52718	0.50377	0.09934	0.09925
8	0.49311	0.49126	0.09825	0.09825
16	0.48121	0.48120		

It is seen that the relative error of the approximation depends on the value of p_1 as well as of b, being much smaller for the larger values of b and smaller for the smaller values of p_1, as it could be expected. It has already been said that the cases considered in the table are unfavourable for the accuracy of the approximative method. However, the results obtained show that the approximative method leads to extremely good results and it will be even better for less unfavourable cases, like: "10 per cent of the v pages have a cumulative frequency of about 0.9 in the trace", which often occurs in practice.

3.8. Application to an operating system

The method for the determination of the page fault probability has been used for the calculation of the performance of a multiprogramming system (MPS), developed by Philips-Electrologica and the results have been compared with those obtained by measurements.

The operating system uses core memory as LS and a disk as MS. A part of the system is resident in core and does not need to be considered further. A fixed number b of pages in core is used as virtual storage for 200 pages of $1k$ octads of the operating system. Page swaps occur on demand from disk to memory by means of the LRU procedure.

The question is which influence b has on the performance of the system, as this number can be chosen at "system generation time". To answer this question we make use of the page frequencies of the 200 pages, obtained by means of software measurements, see [4]. The pages are divided into five groups and the page frequencies of the pages of each group are taken equal to the average page frequency of that group.

The results are approximately as follows.

Group i	Number of pages	p_i (page frequency)
1	4	0.100
2	4	0.050
3	6	0.025
4	6	0.012
5	180	0.001

By means of the described approximative method the page fault probability has been determined for various values of the memory capacity b. As the MPS was made I/O-bound the run time of a batch of jobs is directly dependent on the page fault probability, viz. the additional run time needed because of the page swaps is equal to the product of α and the access time for disk. Hence, the curve for the run times as a function of b is, apart from a scale transformation, the same as the one for α as a function of b. This observation makes it possible to compare the analytical results with those from measured run times.

In Figure 7 the curve represents results obtained by the analytic method; the crosses are results from measurement, see [4]. It is seen that the analytic method leads to a very good description of the actual performance.

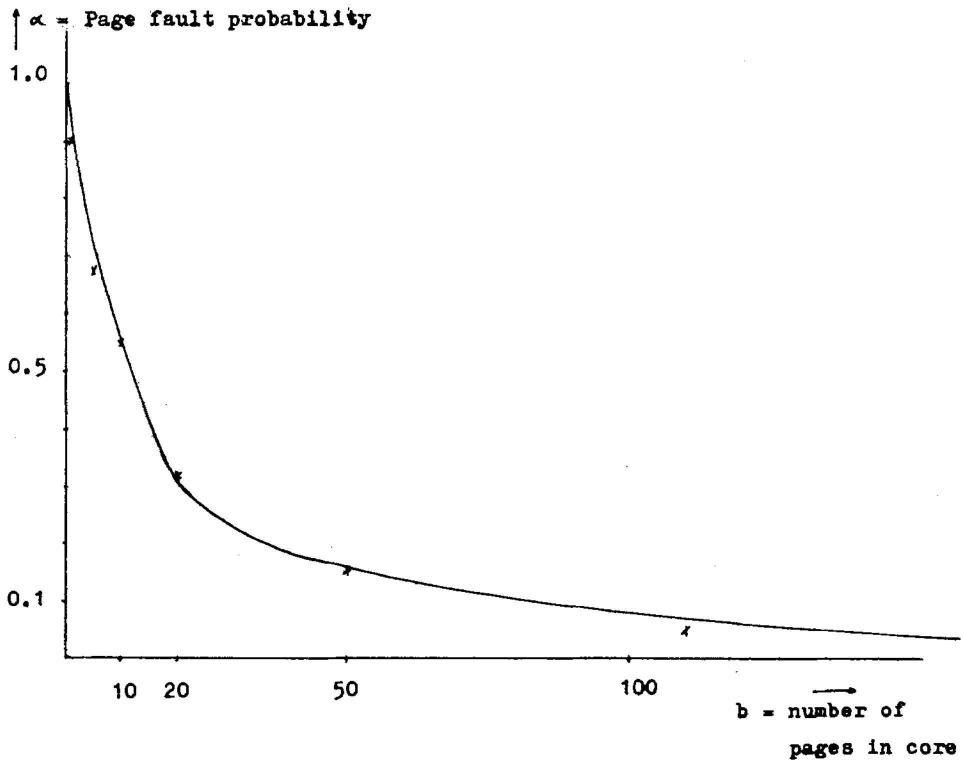

Figure 7.

4. Conclusion

In the preceding sections some models for storage hierarchy have been exposed and investigated. From the results, which are obtained so far, we conclude that the proposed approach of the problem is very useful. However, a number of questions and problems are encountered in practice, which are worthwhile to mention and which need further investigation.

i. The determination of the value v can be rather intricate in practical cases. Presumably, v should be taken equal to the number of separate pages in the trace to be processed. However, for large main storages the „working set" concept of Denning [2] should be applied. The size of v is then a fraction of the total number of pages in the main storage.

ii. The assumption that all separate pages are processed independent of one another is surely not true. The error made due to this assumption is negligeable if the number of dependent separate pages in the trace is smaller than the number of pages in the local store (However, further information and study of this point is desirable).

iii. If instead of a two level hierarchy a three level hierarchy is applied the method developed may be refined to analyse this situation. Some numerical studies for a three level hierarchy are desirable.

To get more insight in the nature of the traces some additional research should be made, e.g.

i. It is extremely desirable to study the page fault probability for various distributions $F(c)$ of the concentration. For instance for

a. $F(c) = 1 - e^{-(\pi/4)c^2}, \quad c > 0;$

b. $F(c) = 1 - e^{-kc^{\frac{1}{4}}}, \quad c > 0;$

with k determined by $\int_0^\infty c\,dF(c) = 1$, i.e. $k = (4!)^{\frac{1}{4}}$;

c. $F(c) = 1 - \dfrac{1}{(1 + \delta c)^\gamma}$, with $\delta(\gamma - 1) = 1,\ \gamma > 1.$

ii. Mixing and demixing should be studied. For instance if two traces with distributions $F_1(c)$ and $F_2(c)$ are given with

$$\alpha_1 = E(\boldsymbol{c}_1 e^{-d_0^{(1)}c_1}),\ d_1 = 1 - E(e^{-d_0^{(1)}c_1}) \text{ for } F_1(c),$$

$$\alpha_2 = E(\boldsymbol{c}_2 e^{-d_0^{(1)}c_2}),\ d_2 = 1 - E(e^{-d_0^{(2)}c_2}) \text{ for } F_2(c),$$

and if these traces are mixed with mixing weights ρ_1 and ρ_2 so that $\rho_1 + \rho_2 = 1$, $1 > \rho_1 > 0$, what can be said about the relation between α on the one hand and α_1, α_2 on the other hand if

$$\alpha = E(\boldsymbol{c}e^{-d_0c}),\ d = 1 - E(e^{-d_0c}) \text{ for } F(c),$$

with

$$F(c) = \rho_1 F_1(c) + \rho_2 F_2(c).$$

This question can be investigated by means of the formulas developed in the preceding sections.

Acknowledgement

The authors are grateful for many helpful discussions with C. P. J. Alewijnse and A. Palmboom.

REFERENCES

[1] L. A. Belady, A study of replacement algorithms for a virtual storage computer. *I.B.M. Systems Journal*, Vol. 5, 2 (1966) 78–101.
[2] P. J. Denning, The working set model for program behaviour. *Comm. A.C.M.* 11, 5 (1968) 323–333.
[3] R. L. Mattson, J. Gecsei, D. R. Slutz and I. L. Traiger, Evaluation techniques for storage hierarchies. *I.B.M. Systems Journal* 2(1970) 78–117.
[4] B. M. Middel, MPS onderzoek (in Dutch). Philips report CDT.TF 5208.
[5] W. J. Hendricks, The stationary distribution of an interesting Markov chain. *Journal of Appl. Prob.* 9 (1972) 231–233.
[6] P. J. Burville and J. F. C. Kingman, On a model for storage and search. *Journal of Appl. Prob.* 10 (1973) 697–701.