ELSEVIER

Information Processing Letters 61 (1997) 323-329

Information Processing Letters

An exact analysis on expected seeks in shadowed disks

Athena Vakali [1], Yannis Manolopoulos *

Department of Informatics, Aristotle University, 54006 Thessaloniki, Greece

Received 26 January 1996
Communicated by F. Dehne

Abstract

Shadowed disk systems store identical data. Therefore, reads are satisfied by using the "nearer server rule" to choose the appropriate disk, whereas writes are satisfied by all disks. Such systems have been studied in order to derive approximate formulae for the expected seek distances traveled for reads and writes. In the present paper, earlier analytic models are studied again, under a different perspective which takes into account the disk scheduling policy applied and the total number of cylinders per disk. Our analysis is exact and results in new formulae for the expected seek distances traveled for reads and writes. The deviation of the earlier formulae with respect to the new exact formulae decrements with decreasing number of shadowed disks, increasing number of disk cylinders, as well as decreasing ration of reads vs. writes. Thus, it is verified that earlier results could be used as good approximations in the asymptotic case. © 1997 Elsevier Science B.V.

Keywords: Single-headed disks; Shadowed disk systems; Replicated disk; Seek distance; Seek time; Performance evaluation

1. Shadowed disk systems

A shadowed disk system with replicated conventional disks is considered, where in each disk identical data are stored. Reading data is satisfied by accessing any of the disks, since they all store exactly the same data. The choice of the disk to be accessed is made by applying the "nearer server rule", i.e. we access the disk on which the read/write heads are closest to the requested cylinder. Writing new information must be satisfied by all disks since they all have to be identical copies. In [1] and [2,4] analytic models have been developed in order to study the behavior of seek distances traveled, and expressions have been derived for the expected read and write seek distances as functions of the number of disks. Thus, since the disk choice is optimized, there is a certain reduction in expected seeks for reads, whereas seek performance for writes will be at most the maximum of seek distances instead of being their sum. It is, also, noted that the use of such shadowed disk systems provides both reliability and fault tolerance. In addition, an immediate backup service is supported, while data are accessible whenever at least one disk is available.

Seek time may be approximated by the average number of cylinders traveled by the r/w heads when the arm moves from the current cylinder to the requested one. In general, a uniform distribution of requested cylinders is assumed. Although this does not happen in practice, it serves as a good approximation. By assuming independence between successive disk operations, the model in [1] resulted in specific expressions for the expected seek distances traveled for

* Corresponding author. Email: manolopo@athena.auth.gr.

[1] Email: avakali@athena.auth.gr.

PII S0020-0190(97)00030-6

both read and write requests:

$$\mathrm{E}[\mathit{read}] = \frac{C}{2k+1}$$

and

$$\mathrm{E}[\mathit{write}] = C(1 - I_k),$$

where

$$I_k = \begin{cases} \dfrac{2k}{2k+1} I_{k-1} & \text{if } k > 1, \\ 2/3 & \text{if } k = 1, \end{cases}$$

C is the total number of cylinders per disk, and k is the number of shadowed disks.

In [2,4] a more refined model has been developed by using Markov chains and by taking in consideration the fact that during the first few accesses following a write access, several disks will have their r/w heads positioned on identical cylinders. The result is that the system will behave as if the value of k was reduced. A Markov chain is introduced with state-space $\{1, 2, \ldots, k\}$ and transition function p,

$$p(i,1) = w \quad \text{for } 1 \leqslant i \leqslant k, \tag{1}$$

$$p(i,i) = \frac{r(i-1)}{i} \quad \text{for } 1 < i < k, \tag{2}$$

$$p(i,i+1) = \frac{r}{i} \quad \text{for } 1 < i < k, \tag{3}$$

where r (respectively, w) is the percentage of read requests (respectively, writes). Evidently, the following relation holds:

$$r + w = 1.$$

In addition, the boundary conditions have to be defined as:

$$p(1,1) = w, \tag{4}$$

$$p(1,2) = r, \tag{5}$$

$$p(k,k) = r,$$

$$p(k,k+1) = 0.$$

Fig. 1 depicts the respective Markov chain for $k = 5$ disks.

According to [2] the explanation of $p(i, i+1)$ in Eq. (3) is as follows.

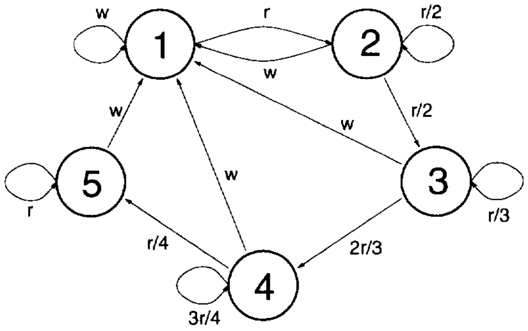

Fig. 1. Markov chain for $k = 5$ according to the Lo–Matloff model.

"In State i, in which there are i distinct positions of the r/w heads, $k - (i - 1)$ heads will be at identical positions, while $i - 1$ heads will be at singleton positions, i.e. positions which are non-duplicates. Suppose the next request is a read. Although this request has k physical drives from which to choose, it has only i different head positions from which to choose, and by symmetry, each of these i head positions is equally likely to be the one closest to the requested cylinder. Thus, there is probability $1/i$ that one of the $k-(i-1)$ same-position heads is the one which is chosen to service this request. Inspection shows that this scenario is the only one which can produce an $i \rightarrow i + 1$ transition."

Thus, new formulae were produced for expected read and write distances traveled by using the above transition function:

$$\mathrm{E}[\mathit{read}] = \sum_{i=1}^{k} \pi_i \frac{C}{2i+1} \tag{6}$$

and

$$\mathrm{E}[\mathit{write}] = \sum_{i=1}^{k} \pi_i C(1 - I_i). \tag{7}$$

The quantity π_i is the long-run proportion of the time the process spends in state i (for $i = 1, 2, \ldots, k$),

$$\pi_i = \pi_1 \prod_{j=1}^{i} \frac{f_{j-1}}{f_j + w},$$

where

$$\pi_1 = \left(\sum_{i=1}^{k} \prod_{j=1}^{i} \frac{f_{j-1}}{f_j + w} \right)^{-1}$$

Table 1
Expected read and write seeks according to the Lo–Matloff model (percentages of C)

	E[*read*]			E[*write*]		
k	$r = 0.95$	$r = 0.5$	$r = 0.05$	$r = 0.95$	$r = 0.5$	$r = 0.05$
2	0.2066	0.2667	0.3267	0.46	0.4	0.34
4	0.134	0.2558	0.3266	0.5632	0.4148	0.3401
6	0.1148	0.2556	0.3266	0.5996	0.4152	0.3401
8	0.1089	0.2556	0.3266	0.6131	0.4152	0.3401
10	0.1071	0.2556	0.3266	0.6178	0.4152	0.3401

and

$$f_i = \begin{cases} 1 & \text{if } i = 1, \\ r/i & \text{if } 1 < i < k, \\ 0 & \text{if } i = k. \end{cases}$$

Table 1 represents the values of both expected read and write seeks as a percentage of the total number of cylinders per disk. These percentages remain the same, no matter what the value of C is, since the parameter C disappears after simplification.

2. The new analysis

The explanation of Eq. (3), as quoted in the previous section, has two drawbacks:

(1) It does not take into consideration the fact that, in order to decide which r/w head should move to satisfy the request, we apply the "nearer server rule".

(2) The boundary conditions $p(1,1) = w$ and $p(1,2) = r$ (of Eqs. (4) and (5), respectively) are simplifications. Both conditions are inexact and the reason is the same: the calculation of these probabilities should take into account the case of read request from the cylinder where all the r/w heads reside immediately after a write request. Thus, in reality $p(1,1) > w$ and $p(1,2) < r$.

In the sequel, we will develop a new analysis, which is based on a different point of view. In simple words, our analysis introduces a new parameter, i.e. the total number of disk cylinders per disk (C). This way, it will become possible to base our analysis on the "nearer server rule" and derive the exact probabilities $p(1,1)$ and $p(1,2)$ in a unified way (and not as exceptions).

A Markov chain is used again to describe the process. In the same manner, as a state we define the number of distinct values of cylinders where r/w heads lie along the k disks, i.e. the state space is $\{1, 2, \ldots, k\}$. Being at state i means that there are exactly i distinct cylinder positions occupied by the k r/w heads. Therefore, at state i there are $C - i$ cylinder positions non-occupied by any r/w head along the k drives. In this state, also, there is a non-singleton position with $k - (i - 1)$ r/w heads, whereas the $i - 1$ r/w heads lie on top of $i - 1$ distinct positions.

It is clear, that if a write request arrives, then all k heads will move to an identical position. Thus, easily we derive that Eq. (1) holds in our model too:

$$p(i, 1) = w \quad \text{for } 1 \leqslant i \leqslant k. \tag{8}$$

However, the question arising is: "when do we move from an i state to an $i+1$ state?" Since C is the number of disk cylinders and each cylinder position is equally likely to be accessed next, there is an i/C probability that one of the occupied cylinders will be requested next. In such a case, evidently the state remains the same. In case that a non-occupied cylinder is hit (under probability $(C - i)/C$), there are chances that we do move to an $i + 1$ state. An exact formula for the probability $p(i, i + 1)$ is going to be derived next.

Let us consider a simple instance of a system being at a state $i = 3$, with $k = 4$ disks, each disk having $C = 4$ cylinders. Evidently, there are 2 $(= k - i + 1)$ heads positioned on top of the same cylinder. In the left column of Table 2, we show all the possible initial head arrangements of the 4 $(= 2 + 1 + 1)$ heads on top of the 4 cylinders. Consider, now, that a read request arrives hitting any cylinder with equal probability. The rest of the entries in each line of Table 2 depict all the possible head placements which will be produced

Table 2
Initial and final placements of head positions for $i = 3$, $k = 4$ and $C = 4$

Initial head placements	Final head placements			
2110	2110	2110	2110	2101
2101	2101	2101	2110	2101
2011	2011	1111	2011	2011
0211	1111	0211	0211	0211
1210	1210	1210	1210	1201
1201	1201	1201	1111	1201
1120	1120	1120	1120	1111
1021	1021	1111	1021	1021
0121	1021	0121	0121	0121
1102	1102	1102	1111	1102
1012	1012	1102	1012	1012
0112	1012	0112	0112	0112

after a read request arrives on each specific cylinder. As in [2,4], it is assumed that these head placements are produced equiprobably. In each of these cases, the r/w head which will move depends on the relative positions between the occupied and the hit cylinder. According to the data of this table, we observe that $p(3,4) = 6/48 = 1/8$, whereas according to Eq. (3) of the Lo–Matloff model this probability is $1/3$.

In order to formulate the following analysis, we define the notion of the *subinterval* as the number of cylinders between any two successively occupied cylinders, or between the beginning of the disk and the first occupied cylinder, or finally, between the last occupied cylinder and the end of the disk. It is known that the subinterval length, denoted by *sub*, obeys the probability distribution function [3]:

$$P(C, sub, i) = \binom{C - sub - 1}{i - 1} \Big/ \binom{C}{i}.$$

Following the reasoning of [2,4], we accept that the non-singleton position can take any place before, among or after the singleton positions with equal probability (see Table 2). In the previous example, the non-singleton position can equiprobably be either the first, the second or, finally, the third occupied disk cylinder. Thus, in general, we have to consider three distinct cases.

(1) The non-singleton position is the first occupied cylinder. Since all cylinders are hit equiprobably, any of the cylinders of the subinterval to the left of the non-singleton position may be hit by a read under probability sub/C. Thus, one head out of the $k - i + 1$ ones will have to move, and an $(i, i + 1)$ state will be produced. This case will happen with probability:

$$P_1 = \sum_{sub=1}^{C-i} \frac{sub}{C} P(C, sub, i) = \frac{1}{C} \frac{\binom{C}{i+1}}{\binom{C}{i}} = \frac{1}{C} \frac{C - i}{i + 1}.$$

Also, if a read request arrives at the subinterval to the right of the non-singleton position, then we will have a transition $i \longrightarrow i + 1$ with probability

$$P_2 = \sum_{sub=1}^{C-i} \left\lceil \frac{sub}{2} \right\rceil / C P(C, sub, i)$$

$$= \frac{1}{C} \frac{1}{\binom{C}{i}} \sum_{sub=1}^{C-i} \left\lceil \frac{sub}{2} \right\rceil \binom{C - sub - 1}{i - 1}.$$

This expression is explained by the fact that, if the subinterval has length equal to *sub*, then one of the $k - i + 1$ heads will move if the read request arrives to the $\lceil sub/2 \rceil$ closest cylinders out of the *sub* ones, whereas the neighbor singleton head will move to service the read request if the latter refers to the closest $\lfloor sub/2 \rfloor$ cylinders. We note that, on purpose, we move one of the non-singleton heads to service the read request referring to the median subinterval cylinder (the $\lceil sub/2 \rceil$th one), in order to better/evenly distribute the heads along the data band.

Therefore, if the non-singleton position is the first occupied cylinder, then we will have an $i \longrightarrow i + 1$ transition under probability $P_1 + P_2$.

Table 3
Representative $p(i, i+1)$ values under the Lo-Matloff and the new models as functions of i and C

state i	model	new model		
	Lo-Matloff	$C = 10$	$C = 100$	$C = 1000$
1	1	0.9	0.99	0.999
2	0.5	0.4222	0.4925	0.4993
3	0.3333	0.2611	0.3266	0.3327
4	0.25	0.1786	0.2437	0.2494
5	0.2	0.1273	0.1939	0.1994
6	0.1667	0.0913	0.1607	0.1661
7	0.1429	0.0636	0.137	0.1423
8	0.125	0.0406	0.1191	0.1244
9	0.1111	0.02	0.1054	0.1106

(2) The non-singleton position is neither the first nor the last occupied cylinder. With the same reasoning as in the previous case, it can be derived that we will have an $i \rightarrow i+1$ transition with probability $P_2 + P_2$.

(3) The non-singleton position is the last occupied cylinder. This case is symmetric to the first one. Therefore, we will have an $i \rightarrow i+1$ transition with probability $P_2 + P_1$, too.

Since the second case may arise $i - 2$ times, we conclude that the probability to have an $i \rightarrow i+1$ transition after a read request is

$$p(i, i+1) = r\frac{(P_1 + P_2) + (i-2)(P_2 + P_2) + (P_2 + P_1)}{i}$$
$$= \frac{2r}{iC}\left(\frac{C-i}{i+1} + \frac{i-1}{\binom{C}{i}}\sum_{sub=1}^{C-i}\left\lceil\frac{sub}{2}\right\rceil\binom{C-sub-1}{i-1}\right). \tag{9}$$

Given a read request, we will not have an $i \rightarrow i+1$ transition if one of the i occupied cylinders is hit, or one of the singleton heads will be nearer to the request. Therefore, we conclude that

$$p(i, i) = 1 - p(i, i+1). \tag{10}$$

According to our analysis the probabilities $p(1, 1)$ and $p(1, 2)$ are not regarded as a special boundary conditions (see Eqs. (4) and (5) respectively), since they are easily derived from the above formulae. Thus we have

$$p(1, 1) = w + \frac{r}{C}, \qquad p(1, 2) = \frac{r(C-1)}{C}.$$

Thus, the only boundary cases which have to be explicitly defined are the following two,

$$p(k, k) = r, \qquad p(k, k+1) = 0,$$

as in the model of [2,4].

As depicted in Table 3, the new values for the transition probability $p(i, i+1)$ are always smaller than the ones calculated in the previous model of [2,4]. It is interesting to emphasize the following two observations with respect to the contents of this table:

(1) The transition probability values of the earlier model converge to the new exact model's values as the number of cylinders C increases. For example, for $i = 1$ there is an 11.1% deviation when $C = 10$, 1.1% deviation when $C = 100$, and 0.1% deviation when $C = 1000$, respectively.

(2) The transition probability values of the earlier model converge to new exact model's values as the number of state i (and evidently, the number of shadowed disks k) decreases. More specifically, when $C = 10$, the earlier model deviates from the new one by 11.1% for $i = 1$, 57.1% for $i = 5$, and 455.6% for $i = 9$, respectively.

Therefore, by using this alternative point of view (i.e. by taking into account the "nearer server rule" and introducing the number of cylinders per disk C), we overcome the analytical simplifications of the transition probability function $p(i, i+1)$ in [2,4], and verify the validity of theses earlier results as a good approximation in the asymptotic case.

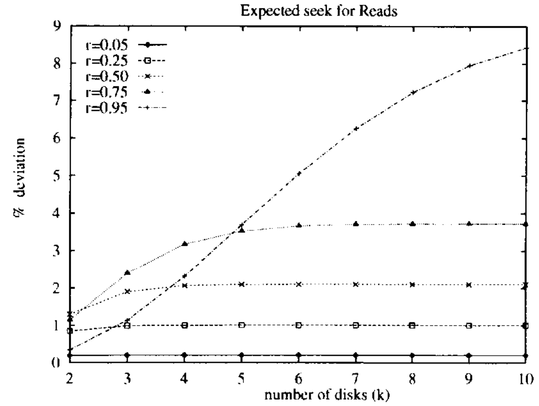

Fig. 2. Deviation percentages of Lo-Matloff model for expected seek of reads, $C = 10$.

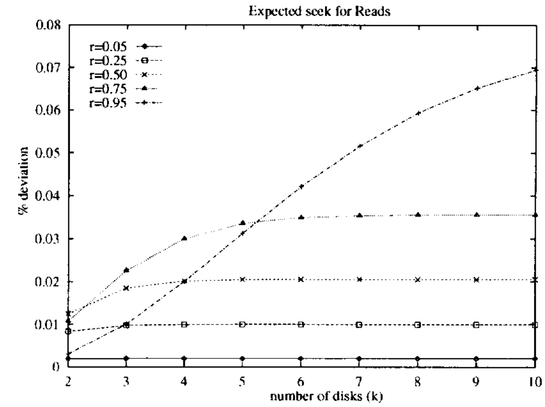

Fig. 4. Deviation percentages of Lo-Matloff model for expected seek of reads, $C = 1000$.

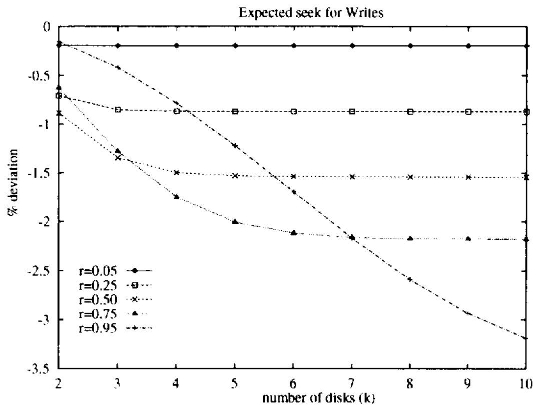

Fig. 3. Deviation percentages of Lo-Matloff model for expected seek of writes, $C = 10$.

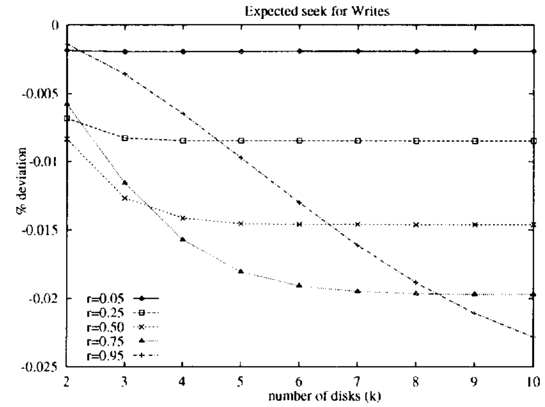

Fig. 5. Deviation percentages of Lo-Matloff model for expected seek of writes, $C = 1000$.

By using the new probability measures introduced here (i.e. Eqs. (8)-(10)), we derive new values for the π_i as

$$\pi_i = \pi_1 \prod_{j=1}^{i} g_j,$$

where

$$\pi_1 = \Big(\sum_{i=1}^{k} \prod_{j=1}^{i} g_j\Big)^{-1}$$

and

$$g_i = \begin{cases} 1 & \text{if } i = 1, \\ \dfrac{p(i-1,i)r}{p(i,i+1)r + w} & \text{if } 1 < i < k, \\ \dfrac{p(k-1,k)r}{w} & \text{if } i = k. \end{cases}$$

These values are introduced in the formulae given in the previous section for expected seek for both reads and writes (Eqs. (6) and (7)).

Figs. 2 and 3 represent the percentage deviation of the seek performance of the previous model [2,4], when compared to the exact model presented here, for $C = 10$. Similarly, Figs. 4 and 5 represent the same percentage deviation of the seek performance, for C = 1000. From these figures we observe that:

- **reading** according to the earlier model is optimistic. In other words, our exact model results in expected seek values bigger than those of the previous model. For example, when $C = 10$, the biggest deviation (8.5%) occurs for the largest values of both the number of shadowed disks $k = 10$, and read ratio $r = 0.95$. For the same value of C, the deviation percentage decreases for less disks (e.g. for r = 0.95, there is a 5.05% deviation for $k = 6$ and 0.33% deviation for $k = 2$). The lowest percentages appear for the smallest values of both k and r, e.g. 0.19% when $k = 2$ and $r = 0.05$. When C increases, there is a convergence between the earlier and new model, and the deviation percentages appear to be quite low (e.g. for $C = 1000$, we remark a 0.07% deviation for $k = 10$, $r = 0.95$ and 0.002% deviation for k = 2, $r = 0.05$).
- **writing** according to the earlier model is pessimistic. Thus, our model behaves better than the previous model. Again for $C = 10$ the biggest deviation (−3.2%) occurs for the largest values of both $k = 10$ and $r = 0.95$. The percentage decreases as the number of shadowed disks becomes less (e.g. for $r = 0.95$, there is a −1.7% deviation for $k = 6$ and −0.15% deviation for $k = 2$). The lowest deviation percentages appear for the smallest values of both k and r, e.g. −0.19% deviation when $k = 2$ and $r = 0.05$. Models converge again as C increases (e.g. for $C = 1000$, we remark a −0.02% deviation for $k = 10$, $r = 0.95$, and −0.002% deviation for $k = 2$, $r = 0.05$).

In both cases (i.e. reading and writing), we remark that the deviation of the earlier approximate model decreases with increasing number of cylinders C, decreasing number of shadowed disks k, and decreasing read ratio r. We note that in practice, k is not much larger than 2, whereas C is much larger that 100. Thus, again we verify the validity of the earlier results as a good approximation in the asymptotic case.

3. Epilogue

Shadowed disk systems provide both reliability and fault tolerance. In addition, an immediate backup service is supported, while data are accessible whenever at least one disk is available. Analytic models have been developed in [1,2,4] in order to derive expressions for the expected read and write seek distances traveled, as functions of the number of available disks.

In the present paper, we re-examine these earlier approaches and derive new exact formulae for the expected read and write seeks by taking into account the "nearer server rule". It is shown that the earlier expressions in [2,4] for reads (respectively, writes) were optimistic (respectively, pessimistic) in the sense that produced values smaller (respectively, larger) than the exact ones. However, it is shown that this earlier model can be used as a close approximate model for large numbers of disk cylinders per disk C and small numbers of available disks k. Thus, we effectively close the case.

References

[1] D. Bitton and J. Gray, Disk shadowing, in: *Proc. 14th VLDB Conf.*, Los Angeles, CA (1988) 331–338.

[2] R.W. Lo and N.S. Matloff, Probabilistic limit on the virtual size of replicated disk systems, *IEEE Trans. Knowledge Data Engineering* 4 (1) (1992) 99–102.

[3] Y. Manolopoulos and J.G. Kollias, Estimating disk head movement in batched searching, *BIT* 28 (1988) 27–36.

[4] N.S. Matloff and R.W. Lo, A greedy approach to the write problem in shadowed disk systems, in: *Proc. 6th IEEE Data Engineering Conf.* (1990) 553–557.