HERMIS '94, Volume 2
E. A. Lipitakis (Editor)
1994, Hellenic Mathematical Society
Printed in Hellas

EXPECTED SEEKS IN MIRRORED DISKS

YANNIS MANOLOPOULOS and ATHENA VAKALI
Department of Informatics,
Aristotle University,
Thessaloniki, Greece 54006.

Abstract

Analytic models on the expected seek distances for a set of $k \geq 2$ replicated disks with one read/write head per surface have been reported in the past. The aim of the present paper is twofold. Since previous models are not exact, firstly, we study the performance of such systems for the boundary case of $k = 2$ disks and provide new exact formulae for the expected seek distances traveled. Secondly, we examine the performance of a set of $k \geq 2$ two-headed disks with either independently or dependently moving heads. For both models of the latter case, we derive new exact formulae for the expected seek distances and make performance comparisons to one-headed disks.

1 Introduction

Recently, there has been a considerable interest in replicated parallel disks where data are stored in identical drives. In such systems both fault tolerance and enhanced performance is achieved. In [1,4] analytic models were developed which study the behavior of seek distance. In addition, disks with two movable heads/surface have been commercially available. Previous work on disks with two heads/surface at a fixed distance has covered both data placement and seek time evaluation issues [2,6,9]. All these works conclude that two-headed systems behave more than twice as well as one-headed systems in terms of seeking. Disks with two heads/surface moving independently are studied in [3], whereas in [7] estimates are given for the average seek distance.

In this paper we study the performance of parallel disks with sets of drives having either one or two heads/surface. The structure of the remainder is as follows. In Section 2 we consider a set of 2 identical one-headed disks and derive new exact formulae for the expected seek distance covered by using Markov chains. In Section 3, a set of k identical two-headed disks is analyzed, where the two heads move autonomously all over the disk surface. For all the above models comparisons are made by varying the read/write ratio and/or the number of disks. Finally, conclusions are summarized and future research is suggested.

2 One-headed Mirrored Disks

A parallel system with replicated one-headed disks is considered, where in each drive 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 which will be accessed is made by applying the 'minimun distance' policy, i.e. we access the disk on which the heads are closest to the requested cylinder. Writing new information must be satisfied by all disk drives since they all have to be identical copies. In [1,4] analytic models were developed which study the behavior of seek distance traveled and expressions were derived for the case of reads and writes as functions of the number of disks.

Seek time is approximated by the average number of cylinders traveled by the 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. The model in [1] resulted in specific expressions for the expected seek distances for both read and write requests:

$$\text{E[read]} = C \;/\; (2k+1) \tag{1}$$

$$\text{E[write]} = C\,(1 - I_k) \tag{2}$$

where C is the total number of cylinders, k is the number of disks and $I_k = \frac{2k}{2k+1} I_{k-1}$ ($I_1 = 2/3$). In [4] a more refined model was developed by using Markov chains and taking in consideration the fact that during the first few accesses following a write access, several disks will have their heads positioned on identical cylinders. The result is that the system will behave as if the value of k were reduced. Thus, new formulae were produced for the above measures:

$$\text{E[read]} = \sum_{i=1}^{k} \pi_i\, C \;/\; (2i+1) \tag{3}$$

$$\text{E[write]} = \sum_{i=1}^{k} \pi_i\, C\,(1 - I_k) \tag{4}$$

where π_i is the long-run proportion of the time the process spends in state i, $(i = 1, 2, ..., k)$:

$$\pi_i = \pi_1 \prod_{j=1}^{i} \frac{a_{j-1}}{a_j + w} \qquad \pi_1 = 1 \;/\; \sum_{i=1}^{k} \prod_{j=1}^{i} \frac{a_{j-1}}{a_j + w}$$

where $a_k = 0$, $a_i = r/i$ (for $1 < i < k$) and r (respectively, w) is the percentage of read (respectively, write) requests (where $r + w = 1$).

However, although the model in [4] is more realistic than that of [1], it makes an approximation. More specifically, the probability that an arriving request will refer to a cylinder different than the ones on which the heads currently lie, is assumed to be negligible. The model discussed here is an extension of the analytic model of [4]. For simplicity, we examine a boundary case of two disks using a Markov chain with a state-space {1,2}, where as state of the system we mean the number of distinct head positions over the data band. Thus, 'state 1' (respectively, 'state 2') corresponds to the case that the heads lie at the same cylinder (respectively, different cylinders). All the possible transition functions of the Markov chain are:

1. $p(1,1)$: the two heads lie on top of identical cylinders and either a read request from any of the current cylinders or a write request arrives,
2. $p(1,2)$: the two heads lie on top of identical cylinders and a read request from neither of the current cylinders arrives,
3. $p(2,1)$: the two heads lie on top of different cylinders and a write request arrives, and
4. $p(2,2)$: the two heads lie on top of different cylinders and a read request from neither cylinders arrives.

Thus, the following relations hold:

$$p(1,1) = s_1\, r + w_1\, w \qquad p(1,2) = n_1\, r \qquad p(2,1) = w_2\, w \qquad p(2,2) = s_2\, r$$

Using the Markov chain properties it holds:

$$(s_1 + s_2 + n_1)\, r + (w_1 + w_2)\, w = 1$$

The calculation of the different variables in this equation is based on the number of appearances of each of the s_1, s_2, n_1, w_1, w_2 in the total number of possible combinations of r/w requests. This number is C^3 since the two heads (one in each disk) might lie in any of the C cylinders and the request might hit any of the C cylinders.

Thus, we have:

$$s_1 = \frac{C}{C^3} = \frac{1}{C^2} \qquad s_2 = \frac{(C-1)\,C^2}{C^3} = \frac{C-1}{C}$$

$$n_1 = \frac{(C-1)\,C}{C^3} = \frac{C-1}{C^2} \qquad w_1 = \frac{C^2}{C^3} = \frac{1}{C} \qquad w_2 = \frac{(C-1)\,C^2}{C^3} = \frac{C-1}{C}$$

Calculation of π_1, π_2 follows by taking in consideration that:

$$\pi_2 = \pi_1\, p(1,2) + \pi_2\, p(2,2) \quad \text{and} \quad \pi_1 + \pi_2 = 1 \Rightarrow$$

$$\pi_1 = \frac{1 - s_2\, r}{1 - r\,(s_2 - n_1)} \quad \text{and} \quad \pi_2 = \frac{n_1\, r}{1 - r\,(s_2 - n_1)}$$

Using the above expressions for π_1, π_2 and the expressions for E[read] and E[write] in [4] we derive that:

$$E[read] = \sum_{i=1}^{2} \pi_i \, C \, / \, (2i+1) = \frac{C}{15} \frac{5 - r\,(5s_2 - 3n_1)}{1 - r\,(s_2 - n_1)} \quad (5)$$

$$E[write] = \sum_{i=1}^{2} \pi_i \, C \, (1 - I_i) = \frac{C}{15} \frac{5 - r\,(5s_2 - 7n_1)}{1 - r\,(s_2 - n_1)} \quad (6)$$

Comparisons of the corresponding expressions for E[read] found in [1,4] are listed in Table 1. The new model, which takes into account the two roles of the transition probability $p(1,1)$ is proven to be not as optimistic as the previous models. Note that the results are expressed as a percentage of the data band. The results derived here as well as in [1,4] for the expected seek for writes are compared in Table 2. In case of writes we have an important performance improvement of 28% when comparing to the model in [1], and 4% to 25% (for $0.1 \leq r \leq 0.9$) when comparing to the model in [4].

r	Formula (1)	(3)	(5)	(7)	(9)
0.9	0.2	0.213	0.328	0.111	0.12
0.7	0.2	0.224	0.332	0.111	0.138
0.5	0.2	0.267	0.333	0.111	0.156
0.3	0.2	0.293	0.333	0.111	0.173
0.1	0.2	0.32	0.333	0.111	0.191

Table 1: Expected seeks for reads (percentage of $C = 200$).

r	Formula (2)	(4)	(6)	(8)	(10)
0.9	0.467	0.453	0.339	0.289	0.28
0.7	0.467	0.427	0.335	0.289	0.262
0.5	0.467	0.4	0.334	0.289	0.244
0.3	0.467	0.373	0.334	0.289	0.227
0.1	0.467	0.347	0.333	0.289	0.209

Table 2: Expected seeks for writes (percentage of $C = 200$).

3 Replicated k Two-headed Disks

Consider a system of k disks having two independent heads/surface. Thus, all the heads lie on at most $2 * k$ different cylinders in the k disks. Since each request refers to a certain cylinder, the requested cylinder imposes the use of $2 * k$ independent variables with the same distribution: k independent variables $(a_1, ..., a_k)$ for the distances of the head A from the requested cylinder and k independent variables $(b_1, ..., b_k)$ of the head B from the requested cylinder in each disk. In total, we have C cylinders per disk, therefore there are C^2 unique seeks (C of size 0, $2 * (C - i)$ of size i=1,2,...,$C - 1$). Thus, the following relations hold:

$$P(a = i) = \frac{2\,(C - i)}{C^2} \qquad P(b = i) = \frac{2\,(C - i)}{C^2}$$

$$P(a \geq i) = \frac{(C - i)(C - i + 1)}{C^2} \qquad P(b \geq i) = \frac{(C - i)(C - i + 1)}{C^2}$$

$$P(a > i) = \frac{(C - i)(C - i - 1)}{C^2} \qquad P(b > i) = \frac{(C - i)(C - i - 1)}{C^2}$$

3.1 Expected Seek for Independent Seeks

In this subsection we adopt the assumption of [1]. In case of a read, one of the chosen disk's heads satisfies the 'minimum distance' property and is sufficient for servicing the request. In other words, $\min(a_1, ..., a_k, b_1, ...b_k)$ is the expected seek distance for read requests. Thus, the calculation of the independent variable:

$$\begin{aligned} v_r &= P(\min(a_1, ..., a_k, b_1, ..., b_k) \geq i) \\ &= P(a_1 \geq i) \;...\; P(a_k \geq i)\; P(b_1 \geq i) \;...\; P(b_k \geq i) \end{aligned}$$

will yield the resulting expected seek distances for reads:

$$\begin{aligned} \text{E[read]} &= \sum_{i=1}^{C-1} v_r = \sum_{i=1}^{C-1} \left(\frac{(C - i)(C - i + 1)}{C^2} \right)^{2k} \\ &\approx \frac{1}{C^{4k}} \sum_{i=1}^{C-1} (C - i)^{4k} = C \sum_{i=1}^{C-1} \frac{1}{C} (1 - \frac{i}{C})^{4k} \end{aligned}$$

The sum of the latter expression is the Riemann's sum for the integral: $\int_0^1 (1 - \frac{i}{C})^{4k} = \frac{1}{4k+1}$. Thus, by replacing this result to the latter expression the expected seek for reads is:

$$\text{E[read]} = C \;/\; (4k + 1) \qquad (7)$$

In case of a write, all the disks will execute the request. Thus, the expected seek distance for writes is: $\max(\min(a_1, b_1)...\min(a_k, b_k))$. A new independent variable, v_w, is considered. Similarly:

$$\begin{aligned} v_w &= P(\max(\min(a_1,b_1) \ldots \min(a_k,b_k)) \geq i) \\ &= 1 - P(\max(\min(a_1,b_1) \ldots \min(a_k,b_k)) < i) \\ &= 1 - P(\min(a_1,b_1) < i) \ldots P(\min(a_k,b_k) < i) \\ &= 1 - (1 - P(\min(a_1,b_1) \geq i)) \ldots (1 - P(\min(a_k,b_k) \geq i)) \\ &= 1 - \left(1 - \frac{(C-i)^2\,(C-i+1)^2}{C^4}\right) \ldots \left(1 - \frac{(C-i)^2\,(C-i+1)^2}{C^4}\right) \end{aligned}$$

Thus:

$$\begin{aligned} \text{E[write]} &= \sum_{i=1}^{C-1} v_w = \sum_{i=1}^{C-1} 1 - \left(1 - \frac{(C-i)^2\,(C-i-1)^2}{C^4}\right)^k \\ &\approx \sum_{i=1}^{C-1} 1 - \left(1 - \frac{(C-i)^4}{C^4}\right)^k = \sum_{i=1}^{C-1} 1 - \left(1 - (1-\frac{i}{C})^4\right)^k \end{aligned}$$

This sum is elaborated by using again the Riemann's sum for the integral: $I_k = \int_0^1 x^k\,(2-x)^k\,(2-2x+x^2)^k\,dx$ and by replacing $u = 1-x$ and $u = sinv$. Thus, finally we derive that the expected seek for writes is:

$$\text{E[write]} = C\,(1 - I_k) \qquad (8)$$

where $I_k = \frac{4k}{4k+1} I_{k-1}$ $(I_1 = 4/5)$.

The fifth column in Table 1 (respectively, Table 2) refers to read (respectively, write) performance of the present model. The values are constant as in the case of the model in [1] (not a function or r). The seek distance gain due to the usage of two-headed disks is an almost 45% (respectively, 38%) decrease for reads (respectively, writes).

3.2 Expected Seek for Dependent Seek Sistances

In the previous section, we assumed that the seek distances are independent of each other as in [1]. The need for each write request to be served by each disk, imposes a dependency between seeks since after a write a set of k heads will be in identical positions [4]. Thus, in order to serve the request following a write, the choice is made out of at most $k+1$ (and not $2k$) different cylinders. Similarly, for the next request the choice is made out of at most $k+2$ positions etc. All the possible fluctuations of the number of different cylinder are:

1. the number remains the same (e.g. an arriving read request refers to a

cylinder 'occupied' already by a head),
2. the number is increased by 1 (e.g. an arriving read request refers to a cylinder not 'occupied' by r/w heads),
3. the number becomes $k+1$ (e.g. an arriving write request forces k heads to move to identical positions, while the other heads may lie in 1,2,...,k different positions).
Consider a Markov with state space $\{k+1, k+2, ..., 2k\}$ and transition functions: $p(i,i) = s_i$, $p(i,i+1) = n_i$ and $p(i,k+1) = w_i$. The values of $i = k+1, ..., 2k$ are transformed to the state-space $\{1, 2, ..., k\}$. Thus:

$$s_i = \frac{k+i-1}{k+i}\, r \qquad n_i = \frac{1}{k+i}\, r \qquad w_i = w = 1 - r$$

By introducing the long-run proportion of the time the process spends in states $i = 1, 2, ..., k$ we have:

$$p(i,i) + p(i,i+1) + p(i,k+1) = 1 \Rightarrow s_i + n_i + w_i = 1$$

By using Markov chain properties we have:

$$\pi_j = \sum_{i=1}^{k} \pi_i\, p(i,j) \Rightarrow \pi_j = \pi_{j-1} n_{j-1} + \pi_j s_j \Rightarrow \pi_j = \pi_{j-1}\, f_j$$

where $f_j = \frac{n_{j-1}}{n_j + w}$ for $j = 2, ..., k$. The fact that $\sum_{i=1}^{k} \pi_i = 1$ yields:

$$\pi_1 = \frac{1}{\sum_{i=1}^{k} \prod_{j=1}^{i} f_j}$$

By using these formulae and the formulae for expected read and write seek found in Subsection 3.1 the new expected seek for reads and writes will be (respectively):

$$\mathrm{E[read]} = \sum_{i=1}^{k} \pi_i\, C \;/\; (4i+1) \tag{9}$$

$$\mathrm{E[write]} = \sum_{i=1}^{k} \pi_i\, C\, (1 - I_i) \tag{10}$$

The last column in Table 1 (respectively, Table 2) refers to read (respectively, write) performance of the present model. The values are close to those of the fifth column, though they are higher (respectively, lower) for reads (respectively, writes). In comparison with the model in [4] there is again a considerable gain due to the usage of two-headed disks. This gain for reads is in the range of 40% to 5% as r varies from 0.9 to 0.1, whereas for writes lies in the range of 40% to 55% as r varies from 0.9 to 0.1.

4 Conclusions

In mirrored disks data are kept in a number of identical disk copies. In the present report, analysis on the expected seeks traveled for read and write requests was carried out by examining both one- and two-headed disks. Comparisons to previous models were made. Further study can examine different servicing policies which might be applied (e.g. SCAN, SSTF) and/or new data placement policies which might influence the system performance. Two-headed disk systems with heads at a fixed distance would also be examined.

References

[1] D. Bitton and J. Gray, Disk Shadowing, *Proceedings of the 14th VLDB Conference*, pp.331-338, LA, CA, 1988.

[2] A. Calderbank, E. Coffman and L. Flatto, Optimum head separation in a disk with two r/w heads, *Journal of the ACM*, Vol.31, No.4, pp.433-466, 1984.

[3] M. Hofri, Should the two-headed discs be greedy?, *Information Processing Letters*, Vol.16, pp.83-86, 1983.

[4] R.W. Lo and N.S. Matloff, Probabilistic Limit on the Virtual Size of Replicated Disc Systems, IEEE Transactions on Knowledge and Data Engineering, Vol.4, No.1, pp.99-102, 1992.

[5] Y. Manolopoulos, Probability Distributions for Seek Time Evaluation, *Information Sciences*, Vol.60, No.1-2, pp.29-40, 1991.

[6] Y. Manolopoulos and J.G. Kollias, Performance of a Two-headed Disk System when Serving Database Queries under the SCAN Policy, *ACM Transactions on Database Systems*, Vol.14, No.3, pp.425-442, 1989.

[7] Y. Manolopoulos and A. Vakali, Seek Distances in Disks with Two Independent Heads per Surface, *Information Processing Letters*, Vol.37, pp.37-42, 1991.

[8] Y. Manolopoulos, Seek Time Evaluation, *Encyclopedia of Computer Science and Technology*, Marcel Deker, NY, in press.

[9] I.P. Page and R.T. Wood, Empirical Analysis of a Moving Disk Head Model with two Heads Separated by a Fixed Number of Tracks, *The Computer Journal*, Vol.24, No.4, pp.339-341, 1981.