The Journal of Systems and Software 55 (2000) 115–128

The Journal of Systems and Software

www.elsevier.com/locate/jss

Data placement schemes in replicated mirrored disk systems

Athena Vakali *, Yannis Manolopoulos

Department of Informatics, Aristotle University, Box 114, 54006 Thessaloniki, Greece

Received 1 March 1999; received in revised form 18 August 1999; accepted 31 October 1999

Abstract

In the present paper, we study the most frequent data placement schemes, i.e., the organ-pipe and camel arrangements, in a mirrored disk system which supports cylinder replication at appropriate locations across the disk surface. Five schemes are proposed to identify the appropriate replica positions for the most frequently accessed cylinders. Three schemes, which are called *Left*, *Right* and *Symmetric* Replication Techniques, are based on an analytical approach. More specifically, given a single replica in the above strategies, estimates are derived for the expected seek distance decrease with respect to a non-replicated mirrored disk system. On the other hand, the remaining two schemes, which are called *Positional* and *Frequency* Replication Techniques, are based on heuristic approaches. We globally compare the performance of these five schemes as a function of the number of replicas. These results are also compared with the corresponding metrics of a conventional (with no replication) mirrored disk system. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Disk management; Mirrored disk; Data placement; Data replication; Seek distance; Performance evaluation

1. Introduction

Storage sub-systems are a vital component of modern computer systems. According to Gibson et al. (1996), the amount of storage sold has been almost doubling each year, and in this context, magnetic disks are the most dominant devices. Typical accesses to disks are much more slower than accesses to the main memory of a computer system. This fact has created the so-called *access gap* which has gained a lot of attention and raised several issues regarding methods to overcome or to minimize the difference between processor speed and disk servicing time.

Given the arrangement of disk surfaces and read/write heads, the time required for a particular disk operation involves mainly the following actions (Ng, 1998; Ruemmler and Wilkes, 1994):

- move the appropriate head to the appropriate cylinder (*seek time*),
- wait for the required sector to rotate around to the location of the head (*latency time*), and
- read the bytes from the disk surface (*block transfer time*).

In the sequel, we concentrate on seeking which is the most important weighting factor of the overall disk service procedure. More specifically, we examine seeking in connection with three important issues which affect the overall storage sub-system performance: data placement, disk mirroring and data replication.

Seek time and seek distances traveled have been evaluated under different disk configurations (Deighton, 1992, 1995; Manolopoulos and Kollias, 1990, Manolopoulos and Vakali, 1991). Additionally, several greedy policies have been proposed, mainly for reducing seek time in anticipation of future requests (Hofri, 1983; King, 1990; Seshardi and Rotem, 1996). By these greedy approaches, there is a head movement at an appropriate position to globally decrease the seek distances traveled by the heads. The appropriate *jockeying* location is $l/3$ cylinders far from the furthest cylinder with respect to the currently request position, where l is the length of the corresponding cylinder range.

The data placement across disk surface affects the overall disk performance. Evidently, if data were stored by taking into consideration their access frequency, then the seek distances would be reduced. In this respect, two policies have been proposed in the literature:

- according to the *organ-pipe arrangement*, data are stored in the disk by placing the most often requested data around the middle of the disk surface. This

* Corresponding author. Tel.: +30-31-99-8415; fax: +30-31-99-6360.

E-mail addresses: avakali@csd.auth.gr (A. Vakali), manolopo@csd.auth.gr (Y. Manolopoulos).

PII: S0164-1212(00)00065-0

method has been proven to minimize the seek distances traveled in conventional single-headed disks (Wong, 1983).

- according to the *camel arrangement*, most often requested data are placed around the middle of the two intervals separated by the middle disk cylinder. The latter method has been proven to be minimal with respect to the seek distances traveled in two-headed disks (Manolopoulos and Kollias, 1990).

Both data arrangement schemes are better described in the next section. The optimal data placement problem has been studied under modern tertiary storage systems such as Tertiary Storage Libraries (Christdoulakis et al., 1997) or Carousel Type Mass Storage Systems (Seshardi et al., 1994).

Redundant inexpensive/independent disk arrays (RAID) have been proposed in order to increase reliability and improve system performance. RAID systems may follow the characteristics of one out of seven levels, numbered 0–6. In particular, according to RAID level 1 (i.e. the *disk mirroring* technique) every disk has an identical corresponding disk storing the same data (Chen et al., 1994). This way, each disk may be viewed as an identical copy of another. In such systems, an immediate backup service is supported, while data are accessible whenever at least one disk is available. Thus, fault tolerance and enhanced performance are achieved at the expense of storage space. Reading data is satisfied by accessing any of the two disks since they store exactly the same data. The choice of the disk to be accessed is made by applying the *nearer-server rule*, i.e., the disk on which the appropriate head is closest to the requested cylinder is chosen. Writing new information must be satisfied by both disks since they have to be identical copies. In (Bitton and Gray, 1988; Bitton, 1989; Lo and Matloff, 1992; Vakali and Manolopoulos, 1997) analytic models have been developed which study the performance behavior of seeking in such systems.

Adaptive disk rearrangement and block reorganization have been studied on single-disk systems to achieve a reduction in expected seek distance by considering request probability distributions (Carson and Reynolds, 1989; Vongsathorn and Carson, 1989). Reliability and performance are improved, also, by taking in consideration adaptive techniques and keeping multiple storage of identical information at free disk locations under various heuristics (Akyurek and Salem, 1995, 1997). The rearrangement techniques have been based on trace driven simulations, and improvement in seek time and seek distance traveled has been reported for conventional disk configurations.

Here, we examine replication schemes in conjunction with certain data placement techniques for a mirrored system, with two disks having one head per surface. This way, we attempt to increase storage device parallelism and availability of most frequently accessed data. The structure of the remainder of the paper is as follows. In Section 2 the data placement schemes in a single disk are presented and their characteristics are described. In Section 3 we assume a set of two mirrored disks and introduce specific replication policies to place replicas in a set of two mirrored disks. There are two general replication categories, where the first one is based on an analytical method, whereas the second approach applies some heuristics in the replication process. In Section 4 we elaborate on the previous replication policies, describe the general replication algorithm used and derive expressions for the expected seek distances traveled. In Section 5 we present experimental results comparing the performance of all the aforementioned policies in terms of expected seeks as a function of the replication degree. In addition, we compare the performance of these policies with that of non-replicated mirrored disks. Finally, future work areas are suggested in Section 6.

2. Data placement schemes

Assume that we are given a conventional disk with one head per surface, each surface consisting of C cylinders. As mentioned in the previous, the organ-pipe arrangement has been proposed as a scheme for efficient data placement in such systems. When placing data with this method, the most frequently accessed data is placed in the central (middle) cylinder (i.e. the $C/2$-th cylinder), whereas the less frequently accessed data are ordered descendingly with respect to their access probabilities and placed alternatively to the left and right of the central cylinder. Fig. 1 depicts the organ-pipe data placement scheme on a conventional disk with a surface of $C = 100$ cylinders, by assuming that the cylinder access probabilities follow a normal distribution function. Since normal distribution has infinite support and disks have finite limits, a truncated normal distribution is used to support the data placement on disks. As shown in the latter figure, there are two mirrored images of the access probabilities with respect to the central cylinder.

The organ-pipe arrangement is a very popular technique from the theoretical and practical point of view. It has been proven that this arrangement is optimal in terms of seek distances traveled (Wong, 1983). More specifically, the organ-pipe arrangement is the scheme which minimizes the expression:

$$\sum_{a=1}^{C} \sum_{b=1}^{C} p(a)\,p(b)\,|a-b|,$$

where a, b refer to all the possible positions of two successively hit cylinders, and $p(a), p(b)$ are their cylinder access probabilities, respectively $(1 \leqslant a, b \leqslant C)$.

It has been proved that in the case of disks with two heads per surface, the camel arrangement is optimal

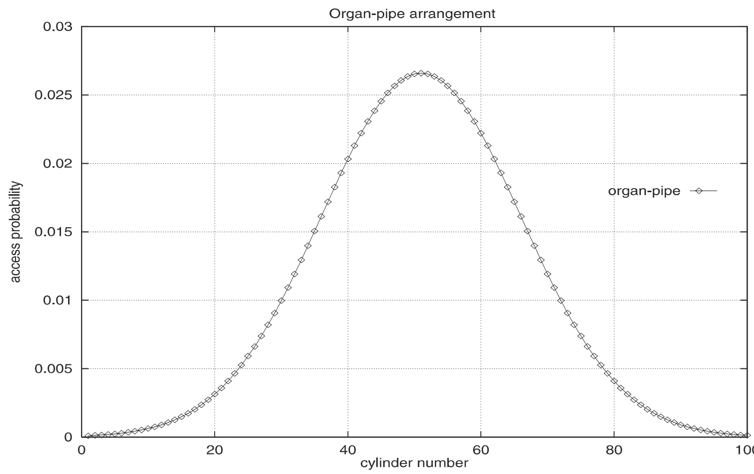

Fig. 1. The organ-pipe data placement scheme.

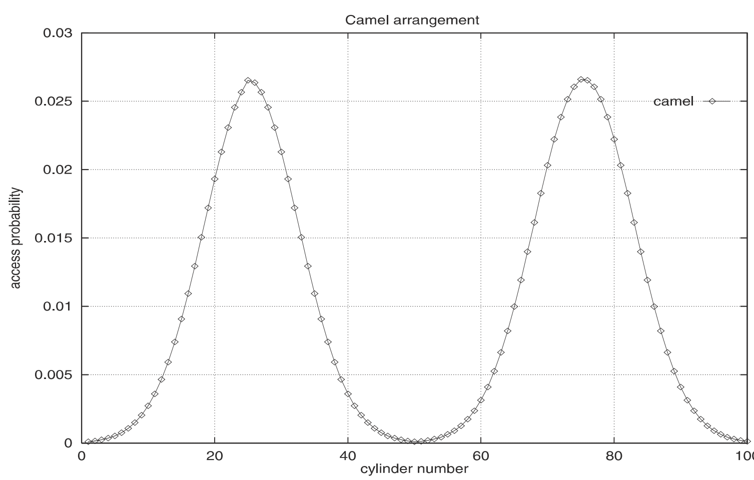

Fig. 2. The camel data placement scheme.

with respect to seek distances traveled (Manolopoulos and Kollias, 1990). In Fig. 2, again we assume that the cylinder access probabilities obey a (truncated) normal distribution function. The camel arrangement might be viewed as two consecutive organ-pipe arrangements centered on cylinder positions $C/4$ and $3C/4$, respectively. Moreover, the two curves are mirrored with respect to the central disk cylinder $C/2$. Therefore, the most frequently accessed cylinders are positioned around two different locations instead of the one location of the central cylinder used in the organ-pipe data placement scheme.

3. The replicated mirrored disk model

In order to increase data availability, a new model is presented here where a set of mirrored disks supports cylinder replication in each disk. Therefore, the replication policy is extended into two different types, an *external* replication of having two identical disks and an *internal* replication of having identical cylinder copies on each disk. As mentioned above, the replication is made by following two different approaches. Both approaches are based on the idea of replicating the most frequently accessed cylinders at appropriate positions on the disk surface. The first approach evaluates replicas positions by maximizing the differences in expected seek distances traveled between a case of a replicated and a case of a non-replicated disk. Following this policy there are three distinct techniques proposed for replication: the *Left*, the *Right* and the *Symmetric Replication Technique*. The second approach applies simple heuristics towards a more straightforward calculation of the replica position. In this respect, two new policies (namely the *Positional* and *Frequency Replication Technique*) are introduced based on either the position of the specific cylinder to be replicated or the cumulative access probabilities between the cylinder to be replicated and the furthest end of the disk surface.

Certain questions do arise concerning the position where to store a replica, the policy under which cylinder replica is chosen, the capacity of the cylinder where the replica will be assigned and so on. In the present paper, we accept that:

- the R most frequent cylinders are selected for replication, R being a parameter,
- for each of the chosen cylinders a single replica is kept in each disk, and
- the cylinder(s) to be replicated are placed on disk cylinders having enough storage space.

The main issue regarding the replication process, is the actual location of each replicated cylinder r. Next, we present three such schemes, where the replication procedure is primarily based on the disk area where the replicas will be placed, i.e., to the left or to the right of the original cylinder c. For each of the three analytical schemes, the specific cylinder location, r, to place the replica is determined so that the expected seek distance traveled difference between the replicated and the non-replicated scheme is maximized. Such an analysis for all the schemes is delivered in the next section. Finally, for each of the two heuristic schemes the replica position is simply identified by taking in consideration either distances on the disk surface or cylinder access probabilities.

3.1. Left Replication Technique

Suppose that a cylinder c is chosen for replication, whereas its replica will be placed in cylinder r at both disks. Fig. 3 represents the model where the replica is placed before the original cylinder position c, i.e., the same information will be stored on both c and r cylinder positions. This approach is called *Left Replication Technique* since the replica is placed at the left-hand side of the original. The middle of the interval between the original and the replica cylinders $(c+r)/2$ is marked in this figure since it is crucial point regarding the choice of the original or the replica cylinder to satisfy the requests.

Fig. 3 shows an example, where there is a random positioning of the two heads over two locations of the mirrored disks, i.e., the head of Disk 1 over cylinder h_1 and the head of Disk 2 over h_2. Each time a request arrives, it can be satisfied by either the original c or the replica r cylinder. Applying the nearer-server rule, the service will be performed by the disk which has its head closer to either c or r. In general the head (e.g. over h_1 or h_2) and the cylinder (e.g. c or r), which will satisfy a request are determined by evaluating the expression: $\min(|h_1-c|,|h_1-r|,|h_2-c|,|h_2-r|)$. More specifically, the following possible cases arise:

$$\begin{aligned}
&\text{if } h_1,\ h_2 < r \quad \text{then service provided by } r,\\
&\text{if } h_1,\ h_2 > c \quad \text{then service provided by } c,\\
&\text{if } r < h_1,\ h_2 < \frac{c+r}{2} \quad \text{then service provided by } r,\\
&\text{if } \frac{c+r}{2} < h_1,\ h_2 < c \quad \text{then service provided by } c,\\
&\text{in all other cases} \quad \text{then evaluate}\\
&\quad \min(|h_1-c|,|h_1-r|,|h_2-c|,|h_2-r|).
\end{aligned}$$

3.2. Right Replication Technique

Under this scheme, the replica (r) is placed at a cylinder after the original cylinder (c). Fig. 4 illustrates a simple example of this approach, which is called *Right Replication Technique*, where the replica is placed in the rth cylinder, after the original cylinder position in both disks. As in the Left Replication Technique, the service is made by applying the nearer-server rule in a similar manner.

Although, these two techniques look identical in nature, however, the results in Section 5 demonstrate a difference in their performance. This is due to the fact that in organ-pipe arrangement the cylinders which are equidistanced from the middle one are not visited with (exactly) equal probability. In reality, the cylinders to the left of the middle cylinder are visited with greater probability than the respective ones to left of the middle cylinder. The same explanation applies to the case of camel arrangement.

3.3. Symmetric Replication Technique

Under this scheme, the replica (r) is placed at a cylinder either before or after the original cylinder (c). Thus, there is a choice of either applying the Left or the

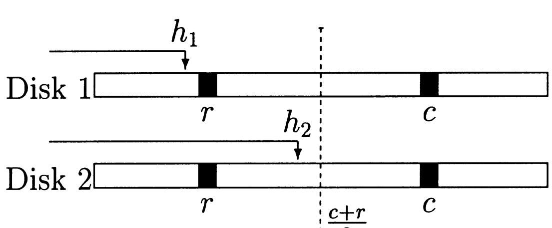

Fig. 3. The Left Replication Technique.

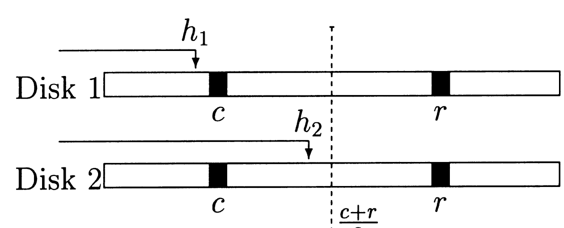

Fig. 4. The Right Replication Technique.

Right Replication Technique depending on the original cylinder position. The central cylinder of the access probability curves (i.e. the $C/2$th cylinder) is the decisive position for the replication scheme to be used. Again, the replica will be placed in cylinder r at both disks. Therefore, the location r is identified by using the following rule:

$$\text{if } c \leqslant \frac{C}{2} \quad \text{then use Right Replication Technique,}$$

$$\text{if } c > \frac{C}{2} \quad \text{then use Left Replication Technique.}$$

This approach is called *Symmetric Replication Technique* since there is a symmetric approach with respect to the replica placement. Again, the service is made by applying the nearer-server rule. This replication scheme is used to gather replicas and originals closer to the central cylinder towards a better seek performance.

3.4. The Positional Replication Technique

The scheme is called *Positional Replication Technique* since it is based on the specific position of the cylinder to be replicated with respect to the two ends of the disk surface. More specifically, this method places the replica $l/3$ cylinder positions far from the inner or the outer disk cylinder depending on the disk end which maximizes l (where $l = \max(C-c, c-1)$). The aim is to allocate between c and r, 2/3 cylinder positions of the total distance between the cylinder c and the most distant cylinder (e.g. the first or the Cth one). Therefore, to identify the location of the replica we apply the following rule:

$$r = \begin{cases} \left\lfloor \dfrac{c}{3} \right\rfloor & \text{when } c \geqslant \dfrac{C}{2}, \\ c + \left\lceil \dfrac{2(C-c+1)}{3} \right\rceil & \text{when } c < \dfrac{C}{2}. \end{cases}$$

Since the nearer-server rule is used to assign the service task to a certain disk, this approach attempts to replicate the most frequent cylinders in such a way that the disk areas accessible to both the original and the replica will be as much as possible balanced in size. This area is expressed by the number of cylinders that will get service from the replica or the original.

3.5. The Frequency Replication Technique

As in the previous scheme, the idea of balancing the number of cylinders that will access the original or the replica is applied to a new replication scheme. The so-called *Frequency Replication Technique* uses the cylinder request frequencies to identify the replica position. Given the cylinder to be replicated, the cumulative access probabilities for cylinders to the left and to the right of the original cylinder are calculated. Having the largest of these two values and their corresponding range of cylinders, this method places the replica in a specific cylinder of the appropriate range, so that this range is divided in two subranges with 1/3 and 2/3 of the corresponding cumulative access probability. The aim is to have between c and r, 2/3 of the total cumulative probability of request frequency accesses. Again, the location of the replica r is identified by applying the following rule:

$$\sum_{i=1}^{r} p(i) \leqslant \frac{1}{3} \sum_{i=1}^{c} p(i) < \sum_{i=1}^{r+1} p(i) \quad \text{when } \sum_{i=1}^{c} p(i) \geqslant \frac{1}{2},$$

$$\sum_{i=1}^{r} p(i) \leqslant \frac{2}{3} \sum_{i=c+1}^{C} p(i) + \sum_{i=1}^{c} p(i) < \sum_{i=1}^{r+1} p(i)$$

$$\text{when } \sum_{i=1}^{c} p(i) < \frac{1}{2}.$$

4. Expected seek distance evaluation

In the present section, we perform the appropriate analysis towards the calculation of expected seek difference between a conventional non-replicated and the replicated mirrored disk system. Table 1 summarizes the parameters of our model.

The choice of the cylinder to be replicated is based on the cylinder access probabilities. Thus, suppose that $p(c)$ represents the probability that a request is for a block placed in cylinder c. In the replicated model, the request might be satisfied by either c or r. The cylinder positions participating in the overall service time, include:

- the cylinder a for the position satisfying the previous request,
- the cylinder c (or r) for the current request, and
- cylinder b which satisfies the next request

The replication might affect the distances traveled from the prior cylinder location a to the requested cylinder c, since the service might be performed by either c or its replica (r). Obviously, the distance from c to the next requested cylinder b is also differentiated by the introduction of r. Without loss of generality, assume that the head of a disk lies on top of the prior request a (i.e. $a = h_2$). We define variable d_{1n} to represent the distance traveled from the prior position to the currently requested cylinder c in a non-replicated mirrored disk system. Therefore, by the nearer-server rule it holds:

$$d_{1n} = \min\{|c-h_2|, |c-h_1|\}.$$

Similarly, variable d_{1r} stands for the distance traveled from the prior positions to the currently requested cylinder c or r in a replicated mirrored disk system, where

$$d_{1r} = \min\{|c-h_2|, |c-h_1|, |r-h_2|, |r-h_1|\}.$$

Table 1
The parameters of our models

Parameter	Explanation
C	Total number of cylinders per disk surface
R	Number of cylinders to be replicated
k	Number of disk drives in a shadowed disk set
l	Cylinder range length
a, c, b	Cylinder locations visited successively, i.e., c: currently requested cylinder, a: cylinder requested prior c, b: cylinder requested after c
r	Replica location of the cylinder c
$p(i)$	Access probability for a cylinder i
h_1	Cylinder location occupied by head in Disk 1
h_2	Cylinder location occupied by head in Disk 2
d_1	Difference between non-replicated and replicated model in the distance traveled from a to c (or its replica)
d_{1n}, d_{1r}	Distances traveled from a to c (or r) under non-replicated and replicated model, respectively
d_2	Difference between non-replicated and replicated model in the distance traveled from c (or its replica) to b
d_{2n}, d_{2r}	Distances traveled from c (or r) to b under non-replicated and replicated model, respectively
E_d	Total difference in expected seek distance between non-replicated and replicated model

Next, we introduce d_1 to be a random variable for the differences in seek distance traveled from cylinder a to c (or r) between the non-replicated and the replicated mirrored disk system. Therefore,

$$d_1 = d_{1n} - d_{1r}.$$

Similarly, we define variables d_{2n} and d_{2r} to represent the distances traveled from the currently requested cylinder to the next request at position b, under a non-replicated and replicated mirrored disk system respectively. In the latter case, d_2 is the random variable for the difference in seek distance traveled from cylinder c (or r) to b between the non-replicated and the replicated mirrored disk system. Evidently, one of the heads is on top of cylinder c (i.e. $h_1 = c$) in the non-replicated model. Also, have in mind that the non-replicated model will result in different d_{2r}, when one head lies on the replica cylinder r (i.e. here $h_1 = r$). Therefore

$$d_2 = d_{2n} - d_{2r},$$

where

$$d_{2n} = \min\{|b - h_2|, |b - c|\}$$

and

$$d_{2r} = \min\{|b - h_2|, |b - r|\}.$$

The quantity d_{2r} applies only when the replica r is used. The overall difference is expected seek distance is expressed by:

$$E_d = E[d_1] + E[d_2]. \tag{1}$$

The calculation of E_d for the replication policies of the previous section is presented in the subsequent two parts.

4.1. Calculation of $E[d_1]$

The calculation of $E[d_1]$ is based on the cylinder positions of the replica r and the original copy c, as well as on the positions h_1 and h_2 where the head lies at each disk. Suppose that a request arrives demanding data stored at cylinders c and r, whereas the previous request was for a block residing at h_1. In this case $a = h_1$ since one head lies on top of the previously requested cylinder by the servicing of this disk, whereas the other head lies on top of h_2 after the service of another prior request. As mentioned earlier, the choice between either the replica or the original cylinder is based on the nearer-server rule, i.e., the minimum distance between h_1 or h_2 and r or c determines the outcome. Then, the general formula for the calculation of $E[d_1]$ is

$$E[d_1] = \sum_{h_1=1}^{C} \sum_{h_2=1}^{C} p(h_1)p(h_2)d_1. \tag{2}$$

In order to calculate $E[d_1]$, all the possible cases of h_1 and h_2 locations related to the locations of the replicas r and the originals c have to be considered. If both h_1 and h_2 are on top of cylinders located to the right-hand side of cylinder $(r + c)/2$, the service will be performed by the head (either h_1 or h_2) which is nearer to the original cylinder location c. Thus, the service is performed by the original cylinder, and there is no difference between this seek distance and the non-replicated seek. Appendixes A.1 and A.3 present the evaluation of $E[d_1]$ for the Left and Right Replication Techniques, respectively. We emphasize the fact that this analysis holds for the case when the previous request was a read request, i.e., heads in each disk lie on top of two distinct cylinders h_1 and h_2.

4.2. Calculation of $E[d_2]$

The calculation of $E[d_2]$ depends on whether the original copy c or the replica r was used to service the previous request. Since there is no change in the seek distance if request was served by c, $E[d_2]$ will be measured by considering all possible cases of serving the request from the replica:

$$E[d_2] = p(r \text{ is used})\ E[d_2/r \text{ is used}]. \tag{3}$$

The probability to use r depends on the relative position of r with respect to c, and is given by the formula:

$$p(r \text{ is used}) = \begin{cases} \sum_{i=1}^{\lceil((r+c)/2)-1\rceil} p(i) & \text{when } c \geqslant r, \\ (1 - \sum_{i=1}^{\lfloor(r+c)/2\rfloor} p(i)) & \text{when } c < r. \end{cases}$$

Suppose that b is the location of the block referenced after the service of the current request by the replica position. Therefore, one head (in one of the disks) is on top of the r location (suppose head h_2, i.e., $r = h_2$), whereas the other head (in the other disk) lies on top of another position suppose h_1. The positions of the heads compared to the requested location, specify the value of the $E[d_2/r$ is used], which (as for the calculation of $E[d_1]$) will be given by the following general formula:

$$E[d_2/r \text{ is used}] = \sum_{b=1}^{C} \sum_{h_1=1}^{C} p(b)p(h_1)d_2.$$

Again, $E[d_2]$ is calculated by examining all possible cases of b and h_1 locations, whereas the partial results found in these cases are combined in the overall Expression (3). Appendixes A.2 and A.4 present the evaluation of $E[d_2]$ for the Left and Right Replication techniques, respectively.

4.3. The general replication algorithm

The following algorithm is used for the evaluation of the expected seek distance for each replication scheme:

1. Apply a data placement scheme (organ-pipe or camel arrangement) where data obey a Normal distribution (as described in Section 2).
2. Choose the R most frequently requested cylinders:
   Under the organ-pipe arrangement:
   cylinders from $C/2-(\lceil R/2\rceil-1)$ to $C/2+\lfloor R/2\rfloor$.
   Under the camel arrangement:
   cylinders from $C/4-(\lceil R/4\rceil-1)$ to $C/4-(\lfloor R/4\rfloor)$, and
   cylinders from $3C/4-(\lceil\lfloor R/2\rfloor/2\rceil-1)$ to $3C/4+\lfloor R/4\rfloor$.
3. For each cylinder out of the R chosen ones
   Under the analytical policies:
   find cylinder r which maximizes E_d by using Eqs. (1)–(3).
   Under the heuristic policies:
   find cylinder r by either Positional or Frequency heuristics.
   Re-estimate the access probabilities for cylinders c and r (see next paragraph).
4. Evaluate expected seek distance differences (Eq. (1)).

4.4. Access probability re-evaluation

After replication the probabilities of the original cylinder (c) and the replica cylinder (r) are updated (see Step 3 of the above algorithm), since the distribution is affected by the introduction of the replica. The access probability for the replica position is increased by an amount q, whereas at the same time the probability of the original cylinder position is decreased by the same amount q. This amount is different depending on the type of replication.

- *Left Replication Technique*: The amount to be added/subtracted to the original probabilities is estimated by the probability of using the replica:

$$q = p(c) \sum_{i=1}^{\lceil((r+c)/2)-1\rceil} p(i). \tag{4}$$

- *Right Replication Technique*: The amount to be added/subtracted to the original probabilities is estimated by the probability of using the replica:

$$q = p(c)\left(1 - \sum_{i=1}^{\lfloor(r+c)/2\rfloor} p(i)\right). \tag{5}$$

- *Other Replication Technique*: In all other replication policies, one of the above calculations is used according to the replica position with respect to the original. In case that the replica is at the left-hand side of the original, the quantity q is evaluated by Eq. (4), whereas if the replica lies at the right-hand side of the original, q is evaluated by Eq. (5).

5. Experimental results

We have run the general replication algorithm of the previous section, which applies to all five replication schemes, for a disk surface of $C = 100$ cylinders totally. In all models the expected seek distance is evaluated as a function of the number of replicated cylinders (R), which varies from 5% to 40% of the total disk surface. Data are stored either by the organ-pipe or the camel arrangement, where the cylinder access probabilities obey either a normal distribution (with $\mu = 50$ and $\sigma = 15$) or a boundary normal distribution converging to a uniform distribution (with $\mu = 50$ and $\sigma = 50$). Here, the uniform distribution for data access probabilities serves as a boundary case.

5.1. Normal probability distribution function

Figs. 5 and 6 depict the expected seek distance for organ-pipe and camel arrangements, under Left, Right, Symmetric, Positional and Frequency Replication Techniques. Organ-pipe shows better behavior than the camel arrangement for almost all replication techniques when we replicate only a few cylinders. This fact was expected since the organ- pipe placement scheme was proven to be the most effective scheme for non-replicated disks. organ-pipe arrangement demonstrates its

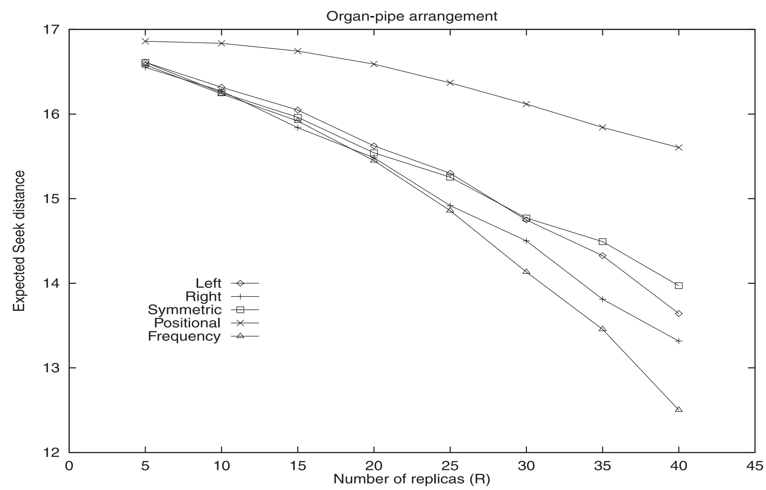

Fig. 5. Expected seek distance for organ-pipe arrangement.

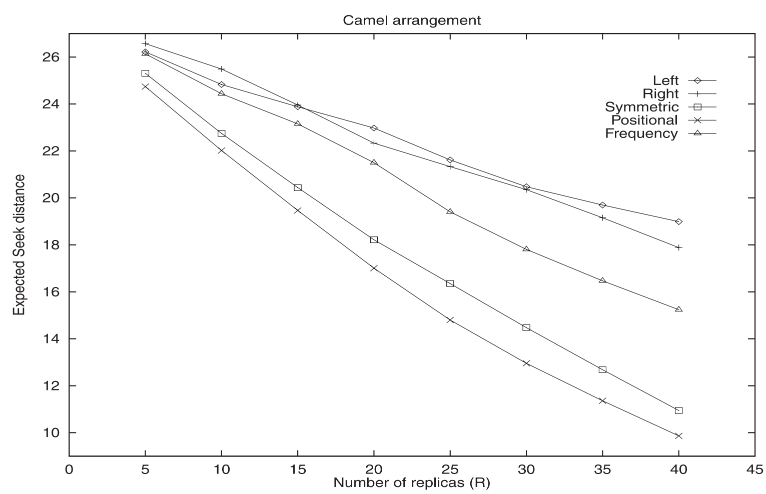

Fig. 6. Expected seek distance for camel arrangement.

best expected seek performance under Frequency Replication Technique, while its worst results appear under the Positional Replication Technique. This is explained by the difference of the replication policy in each of the Frequency and Positional Replication Techniques. More specifically, the Frequency Replication Technique is based on request frequencies for performing replication, such that the most frequent data will be accumulated according to access probabilities. On the other hand, the Positional Replication Technique places replicas at cylinders residing at a certain distance of the original. The organ-pipe placement places the most frequent data near to the middle cylinder so it is expected to perform its best under Frequency Replication (since replicated data will be accumulated), and its worst under Positional Replication (since replicated data will be more scattered across the disks area).

On the other hand, camel arrangement seems to benefit the most by the Positional Replication Technique as the number of replicated cylinders increases. This is explained by the fact that positional heuristic rearranges the two camel curves into more convergent curves. Interestingly, the expected seek distance curve for the camel arrangement (see Fig. 6) under Positional Replication Technique decreases more drastically, whereas for more than 25 replicated cylinders it shows better performance than the organ-pipe arrangement. Therefore, Positional Replication Technique becomes a very useful scheme towards the improvement in seek distance traveled while servicing requests on data stored under the camel arrangement.

6. Uniform probability distribution function

Figs. 7 and 8 depict, also, the expected seek distance for organ-pipe and camel arrangements, under Left, Right, Symmetric, Positional and Frequency Replica-

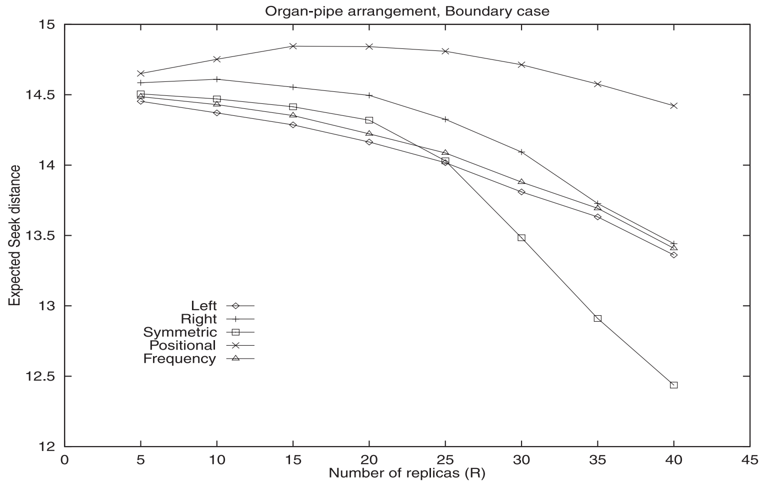

Fig. 7. Expected seek distance for organ-pipe arrangement, boundary case.

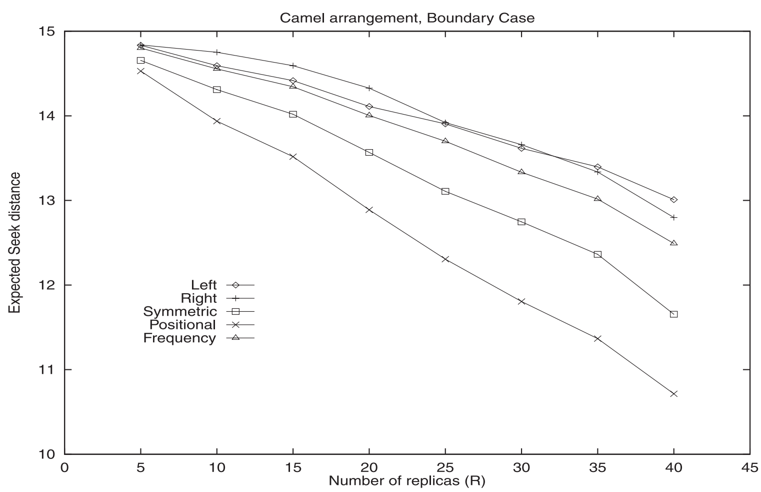

Fig. 8. Expected seek distance for camel arrangement, boundary case.

tion Techniques. However, in these figures data are modeled by a degenerated normal distribution resembling the uniform distribution. Symmetric Replication Technique becomes the best policy for the organ-pipe arrangement as the number of replicas increases (e.g. $R > 25$), whereas Frequency Replication Technique is the best choice for fewer replicated cylinders (e.g. $R < 25$). Positional Technique is the worst scheme under organ-pipe arrangement with uniformly distributed cylinder access probabilities as in the case of organ-pipe with normally distributed access probabilities. This is due to the fact that it places replicas by affecting the original organ-pipe arrangement, which is proven to behave better than the other replication policies. For camel arrangement the Positional Replication Technique is the best scheme, whereas the Left and Right Replication Techniques are the worst, when the cylinder access probabilities obey a uniform distribution function.

7. Comparison of replicated vs. non-replicated models

By assuming a uniform probability distribution function to model the cylinder access probabilities, the expected seek distance is evaluated for the non-replicated shadowed disk system by:

$$E_d = \frac{C}{2k+1}, \qquad (6)$$

where k is the number of shadowed disk, and C is the total number of cylinders (Bitton and Gray, 1988; Lo and Matloff, 1992; Vakali and Manolopoulos, 1997). Therefore, from Eq. (6) it is derived that the expected seek distance for a non-replicated mirrored disk system with $k = 2$ and $C = 100$ will be 20 cylinders. This expected seek distance is used for comparison purposes with our results derived for the case of replicated mirrored disks for all five replication policies.

There is a considerable gain in expected seek distances when a replication technique is used under the normal distribution. As shown in Figs. 5 and 6 the expected seek distance under the organ-pipe arrangement is always less than 17 cylinders for all replication techniques. The improvement rates for all replication techniques vary in the range of 15–48%, when compared to the expected seek distance of the non-replication scheme (20 cylinders). In case of the camel arrangement under all five replication techniques there is a more skewed expected seek distance curve which is beneficial as the number of replicas increase. More specifically, the replication schemes results in expected seek distances of less than 20 cylinders when there is a replication of over than 12% of the disk cylinders area. The improvement in cases of over this amount of replication reaches a rate of almost 50% (40 replicas under Positional Replication Technique) as compared to the non-replicated scheme.

Figs. 9 and 10 depict the improvement rates (%) in expected seek distance for organ-pipe and camel arrangements respectively, under all five replication policies under the uniform distribution over a non-replicated mirrored disk system. As shown in Fig. 9, the performance of organ-pipe arrangement is improved substantially compared to that of the corresponding non-replicated model. More specifically, there is an improvement rate of almost 26% (e.g. for $R = 5$ under the Positional Replication Technique) reaching a highest rate of almost 38% (e.g. for $R = 40$ under Symmetric Replication Technique). Similarly, the camel arrangement (see Fig. 10) has an improvement rate varying between almost 26% (e.g. for $R = 5$ under Right Replication Technique) reaching a highest rate of almost

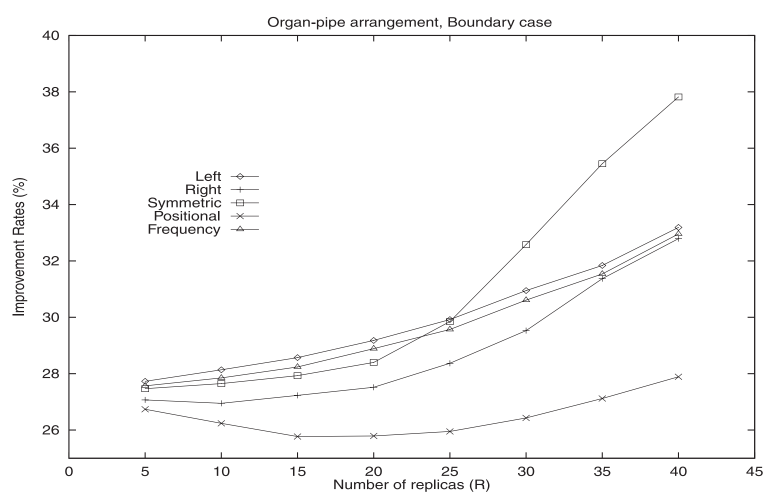

Fig. 9. Improvement rate (%) for organ-pipe arrangement and uniform distribution.

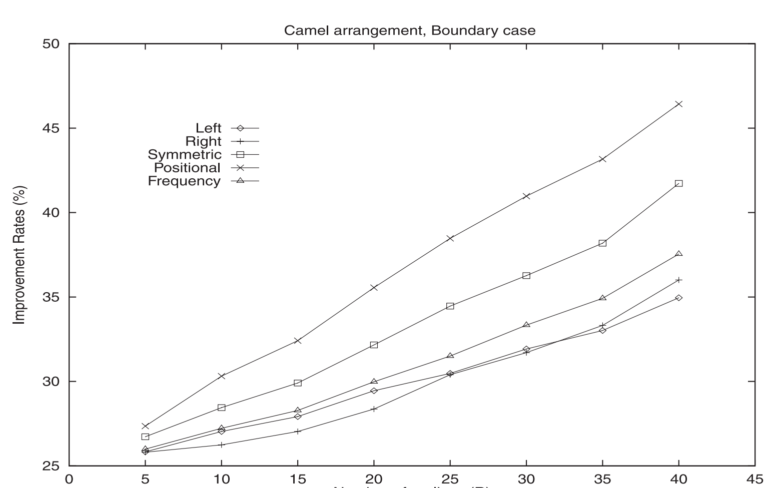

Fig. 10. Improvement rates (%) for camel arrangement and uniform distribution.

47% (e.g. for $R = 40$ under the Positional Replication Technique).

8. Conclusion – further research

Mirrored disk systems are studied under specific data placement schemes on which several replication policies are applied. The replication aims at exploitation of the free disk space and performance improvement as well. The initial simple replication approach is extended to two heuristic approaches towards the minimization of expected seek distance. Several analytic models are presented under different parameters and the replication shows significant improvement rates in the seek distance traveled. Furthermore, the heuristic approaches refine the seek performance under the organ-pipe and camel arrangements for uniformly distributed cylinder access probabilities. Compared to the usual mirrored disks with no replication, our model shows significant improvement in the expected seek distance by supporting analytical and/or heuristic replication.

A measure for determining the optimal level of replication is under consideration. The number of cylinders to be replicated should be specified by considering the following issues:

number of non-occupied cylinders. Each time there are some free cylinders, which could be used for replication. This number is in the range from 1 to C_o cylinders, where o is the number of occupied cylinders. Thus, there is a storage cost to be considered.

replication cost. There is a cost for performing the initial replication. Evidently, the more replication copies, the more scanning will be needed on the disk surface. This cost is determined by seek time that is involved at locating each replication position across the disk cylinder area.

read/write activity. In the present study, we have examine only the case of disk accessing to perform read operations. However, in the general case one has to take into consideration write operations as well. Having replicas leads to increased disk activity when writes are performed.

All the above parameters should contribute to a general formula by means of weighting factors.

Further research should extend the replication schemes presented here, by introducing adaptive block replication in mirrored disks. The analysis could be supported by simulation analysis. The idea of creating several areas on disk where the most frequent data will reside seems quite promising. Thus, examining the effect of placing replicas at high priority disk areas could extend the presented models. The extension of the study to a more general model with more than two disks (shadowed disk set) is underway. Furthermore, the methods used in the present paper could be implemented for a storage subsystem supporting non-identical disk types, and examine the proposed schemes in a more complicated setting. Finally, anticipatory head movement toward optimal positioning is another possible research direction.

Appendix A

A.1. Left Replication Technique – evaluation of $E[d_1]$

The analysis for the evaluation of $E[d_1]$ under Left Replication Technique is presented in Table 2. Eq. (2) if derived by adding the quantities given in the last column entitled "Contributions".

A.2. Left Replication Technique – evaluation of $E[d_2]$

The analysis for the evaluation of $E[d_2]$ under Left Replication Technique is presented in Table 3. The following cases represent the combinations of positions resulting in new contribution to the seek distance and there is special reference to the negative contributions derived in some of the cases.

A.3. Right Replication Technique – evaluation of $E[d_1]$

The summation of all contributions results in the evaluation of Eq. (2) for the Right Replication Technique (see Table 4).

A.4. Right Replication Technique – evaluation of $E[d_2]$

$E[d_2]$ will be measured by Eq. (3) for the Right Replication Technique. Again, the partial quantities found in the following contributions column, are summed in the overall formula (see Table 5).

Table 2
Left Replication Technique – evaluation of $E[d_1]$

Interval	Distances		Contributions
	d_{1n}	d_{1r}	
$h_1 < h_2 \leqslant r \leqslant (c+r)/2$	$c-h_2$	$r-h_2$	$(c-r)\sum_{h_1=1}^{r}\sum_{h_2=h_1+1}^{r} p(h_1)p(h_2)$
$h_1 \leqslant r < h_2 < (c+r)/2$	$c-h_2$	$\min(h_2-r, r-h_1)$	$\sum_{h_1=1}^{r}\sum_{h_2=r+1}^{\lfloor (c+r)/2\rfloor} p(h_1)p(h_2)(d_{1n}-d_{1r})$
	If $r \geqslant (h_1+h_2)/2$, then $d_{1r} = h_2 - r$, else $d_{1r} = r - h_1$		
$h_1 \leqslant r < (c+r)/2 < h_2 < c$	$c-h_2$	$r-h_1$	$\sum_{h_1=1}^{r}\sum_{h_2=\lfloor (c+r)/2\rfloor+1}^{c} p(h_1)p(h_2)(c-h_2-r+h_1)$
	This contribution is valid only when $c-h_2 > r-h_1$, otherwise is there no difference		
$h_1 \leqslant r < (c+r)/2 < c < h_2$	h_2-c	$r-h_1$	$\sum_{h_1=1}^{r}\sum_{h_2=c}^{C} p(h_1)p(h_2)(h_2-c-r+h_1)$
	This contribution is valid only when $h_2-c > r-h_1$, otherwise there is no difference		
$r < h_1 < h_2 < (c+r)/2 < c$	$c-h_2$	h_1-r	$\sum_{h_1=r}^{\lfloor (c+r)/2\rfloor}\sum_{h_2=h_1+1}^{\lfloor (c+r)/2\rfloor} p(h_1)p(h_2)(c-h_2-h_1+r)$
$r < h_1 < (c+r)/2 < h_2 < c$	$c-h_2$	h_1-r	$\sum_{h_1=r}^{\lfloor (c+r)/2\rfloor}\sum_{h_2=\lfloor (c+r)/2\rfloor+1}^{c} p(h_1)p(h_2)(c-h_2-h_1+r)$
$r < h_1 < (c+r)/2 < c < h_2$	h_2-c	h_1-r	$\sum_{h_1=r}^{\lfloor (c+r)/2\rfloor}\sum_{h_2=c}^{C} p(h_1)p(h_2)(h_2-c-h_1+r)$

Table 3
Left Replication Technique – evaluation of $E[d_2]$

Interval	Distances		Contributions
	d_{2n}	d_{2r}	
$b \leqslant r \leqslant h_1 < c$	h_1-b	$r-b$	$(h_1-r)\sum_{h_1=r}^{c}\sum_{b=1}^{r} p(b)p(h_1)$
$b \leqslant r < c < h_1$	$c-b$	$r-b$	$(c-r)\sum_{h_1=c}^{C}\sum_{b=1}^{r} p(b)p(h_1)$
	There is no contribution when $b < (r+h_1)/2$		
$h_1 < b < r < c$ if $(r+h_1)/2 \leqslant b < (c+h_1)/2$ else $b \geqslant (r+h_1)/2, (c+h_1)/2$	$b-h_1$	$r-b$	$\sum_{h_1=1}^{r}\sum_{b=h_1}^{r} p(b)p(h_1)(2b-h_1-r)$
	$c-b$	$r-b$	$(c-r)\sum_{h_1=1}^{r}\sum_{b=h_1}^{r} p(b)p(h_1)$
$h_1 \leqslant r \leqslant b \leqslant c$	$\min(c-b, b-h_1)$	$\min(b-r, r-h_1)$	$\sum_{h_1=1}^{r}\sum_{b=r}^{c} p(b)p(h_1)(d_{2n}-d_{2r})$
$h_1 \leqslant r \leqslant c \leqslant b$	$\min(b-c, c-h_1)$	$b-r$	$\sum_{h_1=1}^{r}\sum_{b=c}^{C} p(b)p(h_1))(d_{2n}-b+r)$
$r \leqslant h_1 \leqslant b < c$	$\min(c-b, b-h_1)$	$b-h_1$	$\sum_{h_1=r}^{c}\sum_{b=h_1}^{c} p(b)p(h_1)(d_{2n}-b+h_1)$
	No contribution from replica when $b \leqslant (c+h_1)/2$		
$r \leqslant h_1 \leqslant c < b$	$b-c$	$b-h_1$	$(h_1-c)\sum_{h_1=r}^{c}\sum_{b=c}^{C} p(b)p(h_1)$
	Always negative contribution from the replica use		
$r \leqslant b \leqslant h_1 < c$	h_1-b	$b-r$	$\sum_{h_1=r}^{c}\sum_{b=r}^{h_1} p(b)p(h_1)(h_1+r-2b)$
	No contribution from replica when $b \geqslant (r+h_1)/2$		
$r \leqslant b < c \leqslant h_1$	$c-b$	$\min(h_1-b, b-r)$	$\sum_{h_1=c}^{C}\sum_{b=r}^{c} p(b)p(h_1)(c-b-d2)$
$r \leqslant c < h_1 \leqslant b$	$\min(h_1-b, b-c)$	$\min(h_1-b, b-r)$	$\sum_{h_1=c}^{C}\sum_{b=c}^{h_1} p(b)p(h_1)(d_{2n}-d_{2r})$
	No contribution from the replica when $b \geqslant (c+h_1)/2$		

Table 4
Right Replication Technique – evaluation of $E[d_1]$

Interval	Distances		Contributions
	d_{1n}	d_{1r}	
$c < r \leqslant h_1 \leqslant h_2$	$h_1 - c$	$h_1 - r$	$(r-c)\sum_{h_1=r}^{C}\sum_{h_2=h_1+1}^{C} p(h_1)p(h_2)$
$c < (c+r)/2 \leqslant h_1 < r \leqslant h_2$	$h_1 - c$	$\min(h_2 - r, r - h_1)$	$\sum_{h_1=\lfloor (c+r)/2\rfloor+1}^{r}\sum_{h_2=r+1}^{C} p(h_1)p(h_2)(h_1 - c - d2)$
$c \leqslant h_1 \leqslant (c+r)/2 < r < h_2$	$h_1 - c$	$h_2 - r$	$\sum_{h_1=c}^{\lfloor (c+r)/2\rfloor}\sum_{h_2=r}^{C} p(h_1)p(h_2)(h_1 - c - h_2 + r)$
	This contribution is valid only when $h_2 - r < h_1 - c$, otherwise there is no difference		
$h_1 \leqslant c < r \leqslant h_2$	$\min(h_2 - c, c - h_1)$	$\min(h_2 - r, c - h_1)$	$\sum_{h_1=1}^{c}\sum_{h_2=r}^{C} p(h_1)p(h_2)(d_{1n} - d_{1r})$
	There is contribution when $c > (h_1 + h_2)/2 \wedge h_2 - r < c - h_1$		
$c < (c+r)/2 \leqslant h_1 < h_2 \leqslant r$	$h_1 - c$	$r - h_2$	$\sum_{h_1=\lfloor (c+r)/2\rfloor+1}^{r}\sum_{h_2=h_1+1}^{r} p(h_1)p(h_2)(d_{1n} - d_{1r})$
$c \leqslant h_1 \leqslant (c+r)/2 < h_2 \leqslant r$	$h_1 - c$	$\min(r - h_2, h_1 - c)$	$\sum_{h_1=c}^{\lfloor (c+r)/2\rfloor}\sum_{h_2=\lfloor (c+r)/2\rfloor+1}^{r} p(h_1)p(h_2)(d_{1n} - d_{1r})$
	There is contribution only when $r - h_2 < h_1 - c$, otherwise there is no difference		
$h_1 \leqslant c \leqslant (c+r)/2 < h_2 \leqslant r$	$\min(h_2 - c, c - h_1)$	$\min(r - h_2, c - h_1)$	$\sum_{h_1=1}^{c}\sum_{h_2=\lfloor (c+r)/2\rfloor+1}^{h_2} p(h_1)p(h_2)(d_{1n} - d_{1r})$
	There is no contribution when $r - h_2 > c - h_1 \wedge h_2 - c > c - h_1$		

Table 5
Right Replication Technique – evaluation of $E[d_2]$

Interval	Distances		Contributions
	d_{2n}	d_{2r}	
$h_1 \leqslant b \leqslant c < r$	$\min(c - b, b - h_1)$	$\min(r - b, b - h_1)$	$\sum_{h_1=1}^{c}\sum_{b=h_1}^{c} p(a)p(h_1)(d_{2n} - d_{2r})$
	There is no contribution from the replica when $b \leqslant (c + h_1)/2$		
$h_1 \leqslant c \leqslant b \leqslant r$	$b - c$	$\min(r - b, b - h_1)$	$\sum_{h_1=1}^{c}\sum_{b=c}^{r} p(a)p(h_1)(b - c - d_{2r})$
$h_1 \leqslant c < r \leqslant b$	$b - c$	$b - r$	$(r-c)\sum_{h_1=1}^{c}\sum_{b=r}^{C} p(a)p(h_1)$
$b \leqslant c \leqslant h_1 \leqslant r$	$c - b$	$h_1 - b$	$\sum_{h_1=c}^{r}\sum_{b=1}^{c} p(a)p(h_1)(c - h_1)$
	There is always negative contribution from the replica use		
$b \leqslant 0 < r \leqslant h_1$	$c - b$	$r - b$	$(c-r)\sum_{h_1=r}^{C}\sum_{b=1}^{c} p(a)p(h_1)$
	There is always negative contribution from the replica use		
$c \leqslant b \leqslant h_1 < r$	$\min(h_1 - b, b - c)$	$h_1 - b$	$\sum_{h_1=c}^{r}\sum_{b=c}^{h_1} p(a)p(h_1)(d_{2n} - h_1 + b)$
	There is no contribution from the replica when $b > (c + h_1)/2$		
$c \leqslant b \leqslant r < h_1$	$\min(h_1 - b, b - c)$	$r - b$	$\sum_{h_1=r}^{C}\sum_{b=c}^{r} p(a)p(h_1)(-r + b)$
$c \leqslant h_1 \leqslant b < r$	$b - h_1$	$r - b$	$\sum_{h_1=c}^{r}\sum_{b=h_1}^{r} p(a)p(h_1))(2b - h_1 - r)$
	This contribution from the replica is valid only when $b \leqslant (r + h_1)/2$, otherwise there is no difference		
$c \leqslant h_1 \leqslant r < b$	$b - h_1$	$b - r$	$\sum_{h_1=c}^{r}\sum_{b=r}^{C} p(a)p(h_1)(r - h_1)$
$c < r \leqslant b < h_1$	$\min(h_1 - b, b - c)$	$\min(h_1 - b, b - r)$	$\sum_{h_1=r}^{C}\sum_{b=r}^{h_1} p(a)p(h_1)(d_{2n} - d_{2r})$
	There is no contribution from the replica when $b < (c + h_1)/2$		

References

Akyurek, S., Salem, K., 1995. Adaptive block rearrangemnt. ACM Transactions on Computer Systems 13 (2), 89–121.

Akyurek, S., Salem, K., 1997. Adaptive block rearrangement under UNIX. Software – Practice and Experience 27 (1), 1–23.

Bitton, D., Gray, J., 1988. Disk Shadowing. In: Proceedings of the 14th VLDB Conference, pp. 331–338.

Bitton, D., 1989. Arm Scheduling in shadowed disks. In: Proceedings of the IEEE COMPCON 89 Conference, pp. 132–136.

Carson, S., Reynolds, P., 1989. Adaptive Disk reorganization, Technical Report. University of Maryland, UMIACS-TR-89-4.

Chen, P., Lee, E., Gibson, G., Katz, R., Paterson, D., 1994. RAID – high performance, reliable secondary storage. ACM Computing Surveys 26 (2), 145–185.

Christodoulakis, S., Triantafillou, P., Zioga, F., 1997. Principles of optimally placing data in tertiary storage Libraries. In: Proceedings of the 23rd VLDB Conference, pp. 236–245.

Deighton, K., 1992. The number of cylinders traveled for a single query on a two seek arm backing storage device. Database Technology Journal 4 (2), 93–96.

Deighton, K., 1995. Average time to travel on two-headed non-linear discs. The Computer Journal 38 (10), 811–817.

Gibson, G.A., Vitter, J.S., Wilkes, J., et al., 1996. Strategic directions in storage I/O issues in large-scale computing. ACM Computing Surveys 28 (4), 763–779.

Hofri, M., 1983. Should the two-headed disk be greedy? – yes it should. Information Processing Letters 16 (2), 83–85.

King, R., 1990. Disk arm movement in anticipation of future requests. ACM Transactions on Computer Systems 8 (3), 214–229.

Lo, R.W., Matloff, N.S., 1992. Probabilistic limit on the virtual size of replicated disc systems. IEEE Transactions on Knowledge and Data Engineering 4 (1), 99–102.

Manolopoulos, Y., Kollias, J.G., 1990. Optimal data placement in two-headed disk systems. BIT 30, 216–219.

Manolopoulos, Y., Vakali, A., 1991. Seek distances in disks with two independent heads per surface. Information Processing Letters 37, 37–42.

Ng, S.W., 1998. Advances in disk technology – performance issues. IEEE Computer 31 (5), 75–81.

Ruemmler, C., Wilkes, J., 1994. An introduction to disk drive modeling. IEEE Computer 27 (3), 17–28.

Seshardi, S., Rotem, D., Segev, A., 1994. Optimal arrangements of cartridges in Carousel type mass storage systems. The Computer Journal 37 (10), 873–887.

Seshardi, S., Rotem, D., 1996. The two headed disk: stochastic dominance of the greedy policy. Information Processing Letters 57, 273–277.

Vakali, A., Manolopoulos, Y., 1997. An exact analysis on expected seeks in mirrored disks. Information Processing Letters 61 (6), 323–329.

Vongsathorn, P., Carson, S., 1989. A system for adaptive disk rearrangement. Software – Practice and Experience 20 (3), 225–242.

Wong, C., 1983. Algorithmic Studies in Mass Storage Systems. Computer Science Press, Rockville, MD.

Athena Vakali received a M.Sc. degree in Computer Science from the Computer Science University, Purdue University, USA (1988) and a Ph.D. Degree in Computer Science from the Department of Informatics, Aristotle University, Greece (1997). She is now a Lecturer at the Department of Informatics in Aristotle university. She has also been a visiting researcher at Purdue University. Her main research interests include: storage subsytems, disk and tapes placement and scheduling, multimedia data placement and caching in distributed systems and in the Web. On these topics she has published several papers in refereed journals and conference proceedings. She is a member of ACM, IEEE and she is a Fulbright fellow.

Yannis Manolopoulos received a 5-year B.Eng. (1981) in Electrical Engineering and a Ph.D. (1986) in Computer Engineering, both from the Aristotle University of Thessaloniki. He has been with the Department of Computer Science of the University of Toronto, the Department of Computer Science of the University of Maryland at College Park and the Department of Electrical and Computer Engineering of the Aristotle University of Thessaloniki. Currently, he is Professor at the Department of Informatics of the latter university. He has published over 90 papers in refereed scientific journals and conference proceedings. He is the co-author of a book entitled "Advanced Database Indexing" and author of two textbooks on data/file structures, which are being taught in most of the computer science/engineering departments in Greece. His research interests include spatial and temporal databases, geographical information systems, www-based information systems, data mining, text databases, and performance evaluation of storage subsystems.