

Replication in Mirrored Disk Systems

Athena Vakali and Yannis Manolopoulos

Department of Informatics, Aristotle University
54006 Thessaloniki, Greece
{avakali,manolopo}@athena.auth.gr

Abstract. In this paper we study data replication in a mirrored disk system. Free disk space is exploited by keeping replicas of specific cylinders at appropriate disk locations. Assuming an organ-pipe arrangement we calculate the expected seek distance by varying the probability cylinder access under different distributions. Also, analytic formulae are derived for the expected seek distance under replication and comparison with the conventional (without replication) mirrored disk system is performed.

1 Introduction

The “access gap”, i.e. the fact that processor and disk speeds differ by three orders of magnitude, has attracted attention towards minimizing this effect by developing efficient storage subsystems. Seeking is the most important factor in disk operations, therefore we focus on this issue in the sequel. A technique to minimize seeking is the *organ-pipe* arrangement, which places the most frequent data in the central cylinder, whereas the less frequent data are stored alternatively in decreasing order to the left and right of the latter cylinder. It has been proven that this scheme is optimal with respect to seeking [12].

Recently, there has been a considerable interest in shadowed/mirrored disks, where all disks are identical and store the same data. In such systems, enhanced fault tolerance and disk performance are achieved at the expense of storage space. In [3,4,8,10] analytic models have been developed to study the performance of seeking. Also, in [6] the average seek time is estimated when multiple data access streams from different disks are merged into a target disk.

Disk rearrangement and adaptive block reorganization have been studied for single disks, in order to reduce seeking either by considering request probability distributions [5,11] or by applying data replication in free disk space [1,2]. The rearrangement techniques have been based on trace driven simulations and seek improvement has been reported for conventional disk configurations.

In this paper, a mirrored disk system (i.e. with two identical disks) is studied for specific replication schemes. We assume that single requests arrive under a distribution which supports the organ-pipe scheme. The structure of the remainder of the paper is as follows. In Section 2 the model and the system variables are described. In Section 3 we perform the analysis for three replication policies in a mirrored disk system, and derive estimates for the expected seek. In Section 4 we present the algorithm for the evaluation of the expected seek, we

run several models for each of the replication strategies and compare the expected seek found with the corresponding expected seek distance derived for the non-replicated mirrored disk model. Comparisons between all the aforementioned disk configurations are discussed. Finally, future work areas are suggested in Section 5.

2 The Replicated Mirrored Disk Model

As mentioned in the previous, the organ-pipe arrangement is an efficient data placement technique. Figure 1 depicts the cylinder access probabilities of a disk with $N=100$ cylinders under the organ-pipe arrangement, where probabilities obey a normal distribution with variances $\sigma=15$ and $\sigma=40$. Such an arrangement is used in our replicated mirrored disk model.

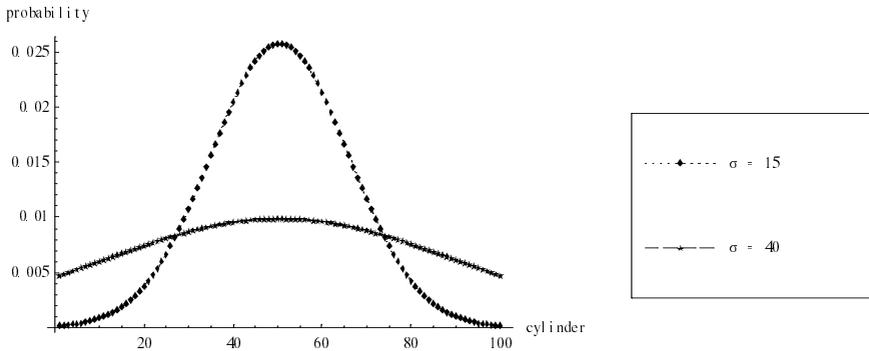


Fig. 1. Organ-pipe data placement scheme.

In a mirrored disk system identical data are stored in both disks. This way, an immediate backup service is supported, while data are accessible whenever at least one disk is available. The choice of the disk to satisfy a read request is made by applying the *nearer-server* rule, i.e. the disk on which the appropriate r/w head is closest to the required cylinder will perform the service.

Here, we propose a mirrored disk model supporting cylinder replication in each disk, in order to increase data availability. Our model is based on the following heuristics:

- the R most frequent cylinders are chosen for replication, R being a parameter in our model,
- a single replica is kept for each of the chosen cylinders, and
- the replica cylinder position has enough storage space in order to host the data of the original cylinder.

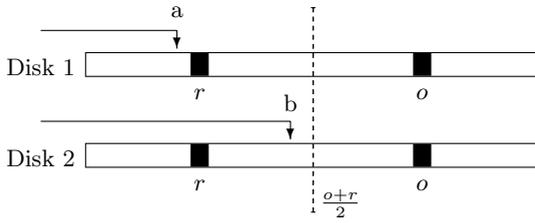


Fig. 2. Nearer-server rule for request servicing - Left Replication Technique.

Suppose that cylinder o is chosen for replication, and its replica is placed on cylinder r at both disks. Thus, if a request refers to the o -th cylinder, it can be satisfied by the r -th one as well. Since there are two disks, we have two heads (namely a and b) available for servicing. According to the nearer-server rule the head that is closer to the requested cylinder will perform the service. Figure 2 represents the model where the replica is placed before the original cylinder position o . The middle of the interval between the original and the replica cylinders $\frac{o+r}{2}$ is a crucial point regarding the choice of the original or the replica cylinder while servicing a read request.

The main issue regarding the replication process, is the choice of the actual cylinder location of the replica r . A first approach is to place the replica r so that the difference in expected seek between the replicated and the non-replicated scheme is maximized. In the next section, we deliver the analysis for the expected seek differences between the conventional non-replicated disk system and the introduced here replicated model.

The selection of the cylinder to be replicated is based on the cylinder access probabilities. Suppose that $p(x)$ is the probability that a request refers to a block of cylinder x . We introduce b_p to be a random variable (in a non-replicated mirrored disk system) for the seek distance traveled from the cylinder where the nearer head lies towards the current requested cylinder. Similarly, b_r is the corresponding random variable in a replicated mirrored disk system. Then, the difference between the non-replicated and the replicated disk system is $d_1 = b_p - b_r$. Similarly, we define n_p to be a random variable in a non-replicated mirrored disk system, referring to the distance traveled from the current cylinder towards the next requested cylinder. Again, n_r is the corresponding random variable in our replicated scheme. As before, the corresponding difference is: $d_2 = n_p - n_r$. Thus, the overall difference in expected seek is:

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

The derivation of $E[d]$ depends on the various replication strategies and is presented in the following section.

3 Replication Strategies

Three replication strategies are presented in the next subsections. The first strategy called *Left Replication*, places the replica in a cylinder to the left of the original cylinder, the second strategy (called *Right Replication*) places the replica to the right of the original cylinder, whereas the third strategy applies either Left or Right Replication policy depending on the position of original cylinder compared to the central cylinder. The expected seek differences are evaluated for each of these strategies. In each case, we need to first evaluate both $E[d_1]$ and $E[d_2]$.

3.1 Left Replication Technique

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 o , as well as on the positions a and b where the r/w head lies at each disk. Suppose that an arriving request refers to data stored at cylinders o and r , whereas the previous request was for the block residing at a (since one head lies on top of the previous requested cylinder). The other head lies on top of b (by the service of another prior request).

The general formula for the calculation of $E[d_1]$ is:

$$E[d_1] = \sum_{a=a_l}^{a_u} \sum_{b=b_l}^{b_u} p(a) p(b) [d_p - d_r] \tag{2}$$

where, a_l, a_u represent the lower and the upper limit respectively of the range where a lies, and b_l, b_u represent the lower and the upper limit respectively of the range where b lies. Variables d_p and d_r are the distances traveled when there is no replication and with replication. In order to calculate $E[d_1]$ all the possible cases of a and b locations related to the locations of the replicas r and the originals o are considered. Obviously, we deal with all cases which contribute to the seek difference derived by the replica's use.

1. $a < b \leq r < o$. The introduction of the replica reduces seeking since the position r is closer to the heads location b . The difference is:

$$\sum_{a=1}^r \sum_{b=a+1}^r p(a)p(b)[(o - b) - (r - b)] = (o - r) \sum_{a=1}^r \sum_{b=a+1}^r p(a)p(b) \tag{3}$$

2. $a < r < b < \frac{r+o}{2} < o$. Contribution to the expected seek may result from either head a or b , depending on which results in the minimum distance to be traveled. Thus, if $x = \min(b - r, r - a)$ the difference is:

$$\sum_{a=1}^r \sum_{b=r}^{\lfloor \frac{r+o}{2} \rfloor} p(a) p(b) [(o - b) - x] \tag{4}$$

3. $a < r < \frac{r+o}{2} < b < o$. In this case, there is contribution from the replica only when $o-b > r-a$. Thus, the difference is:

$$\sum_{a=1}^r \sum_{b=\lceil \frac{r+o}{2} \rceil}^o p(a) p(b) [(o-b) - (r-a)] \tag{5}$$

4. $a < r < o < b$. There is contribution from the replica only when $b-o > r-a$. In this case the difference is:

$$\sum_{a=1}^r \sum_{b=o}^N p(a) p(b) [(b-o) - (r-a)] \tag{6}$$

5. $r < a < b < \frac{r+o}{2} < o$. The contribution to the expected difference is:

$$\sum_{a=r}^{\lfloor \frac{r+o}{2} \rfloor} \sum_{b=a+1}^{\lfloor \frac{r+o}{2} \rfloor} p(a) p(b) [(o-b) - (a-r)] \tag{7}$$

6. $r < a < \frac{r+o}{2} < b < o$. There is contribution from the replica only when $o-b > a-r$. In this case the difference is:

$$\sum_{a=r}^{\lfloor \frac{r+o}{2} \rfloor} \sum_{b=\lceil \frac{r+o}{2} \rceil}^o p(a) p(b) [(o-b) - (a-r)] \tag{8}$$

7. $r < a < \frac{r+o}{2} < o < b$. There is contribution from the replica only when $b-o > a-r$. In this case the difference is:

$$\sum_{a=r}^{\lfloor \frac{r+o}{2} \rfloor} \sum_{b=o}^N p(a) p(b) [(b-o) - (a-r)] \tag{9}$$

If both a and b are on top of cylinders located after the cylinder $\frac{r+o}{2}$, the service will be done by the head (either a or b) which is nearer to the original cylinder location o . Thus, the service is performed by the original cylinder, and there is no difference between this case and a non-replicated system. We emphasize the fact that the above analysis holds when heads lie on top of two distinct cylinders a and b . Equation 2 if derived by adding the above Expressions 3-9.

Calculation of $E[d_2]$

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

$$E[d_2] = p(r \text{ is used}) E[d_2/r \text{ is used}] \tag{10}$$

where $p(r \text{ is used}) = \sum_{i=1}^{\lfloor \frac{r+o}{2} \rfloor} p(i)$. Suppose that n is the location of the block referenced after the service of the current request by the replica's position. Thus,

one head is on top of the r -th cylinder, whereas the other head lies on top of the b -th cylinder. The positions of the heads compared to the requested location, specify the value of the $E[d_2 / r \text{ is used}]$ which is given by the following general formula:

$$E[d_2/r \text{ is used}] = \sum_{n=n_l}^{n_u} \sum_{b=b_l}^{b_u} p(n) p(b) [d_p - d_r]$$

where n_l, n_u represent the lower and the upper limit of the n 's range, b_l, b_u the lower and the upper limit of the b 's range and d_p, d_r the previous and the new distance traveled in a non replicated disk system and in our replicated model, respectively. The following cases represent the combinations of positions resulting in new contribution to seeking and there is special reference to the negative contributions derived in some of the cases:

1. $b < n < r < o$. There is contribution from the replica when $n-b > r-n$. In the non-replicated scheme, the seek distance traveled is $d_n = \min(o-n, n-b)$ whereas in our replicated mirrored disk system the seek distance traveled is $d_r = \min(r-n, n-b)$. Thus, the expected seek difference is:

$$\sum_{b=1}^r \sum_{n=b}^r p(b) p(n) (d_n - d_r) \tag{11}$$

2. $b < r < n < \frac{r+o}{2} < o$. There is contribution from the replica only when $o-n > n-r$. The difference in expected seek is:

$$\sum_{b=1}^r \sum_{n=r}^{\lfloor \frac{r+o}{2} \rfloor} p(b)p(n)[(o-n) - (n-r)] = (o-r) \sum_{b=1}^r \sum_{n=r}^{\lfloor \frac{r+o}{2} \rfloor} p(b)p(n) \tag{12}$$

3. $b < r < o < n$. There is always negative contribution from the replica's use. This negative difference is given by:

$$\sum_{b=1}^r \sum_{n=o}^N p(b) p(n) [(n-o) - (n-r)] = (r-o) \sum_{b=1}^r \sum_{n=o}^N p(b) p(n) \tag{13}$$

4. $b < r < \frac{r+o}{2} < n < o$. The difference in expected seek is:

$$\sum_{b=1}^r \sum_{n=\lceil \frac{r+o}{2} \rceil}^{n=o} p(b) p(n) [(o-n) - (n-r)] \tag{14}$$

There is always negative contribution from this replica use since $o-n < n-r$.

5. $n < r < b < o$.

$$\sum_{b=r}^o \sum_{n=1}^r p(b) p(n) [(b-n) - (r-n)] = \sum_{b=r}^o \sum_{n=1}^r p(b) p(n) (b-r) \tag{15}$$

6. $n < r < o < b$.

$$\sum_{b=o}^N \sum_{n=1}^r p(b) p(n) [(o-n) - (r-n)] = (o-r) \sum_{b=o}^N \sum_{n=1}^r p(b) p(n) \quad (16)$$

7. $r < n < b < o$.

$$\sum_{b=r}^o \sum_{n=r}^b p(b) p(n) [(b-n) - (n-r)] = \sum_{b=r}^o \sum_{n=r}^b p(b) p(n) (b+r) \quad (17)$$

8. $r < n < o < b$.

$$\sum_{b=o}^N \sum_{n=r}^o p(b) p(n) [(o-n) - (n-r)] = (o+r) \sum_{b=o}^N \sum_{n=r}^o p(b) p(n) \quad (18)$$

9. $r < b < o < n$.

$$\sum_{b=r}^o \sum_{n=o}^N p(b) p(n) [(n-o) - (n-b)] = (b-o) \sum_{b=r}^o \sum_{n=o}^N p(b) p(n) \quad (19)$$

10. $r < o < n < b$. The seek distance for the non-replicated scheme is: $d_n = \min(n-o, b-n)$. Thus, there is negative contribution from the replica's use when $n-o > n-r$. The difference is:

$$\sum_{b=o}^N \sum_{n=o}^b p(b) p(n) [d_n - (n-r)] \quad (20)$$

Equation 10 if derived by adding the Expressions 11-20.

3.2 Right Replication Technique

The Right Replication technique places the replica to the right of the original cylinder, i.e. from cylinder locations $o+1$ to N . As before, the appropriate replica location is found by the value maximizing the expected seek expressed by Formula 1.

Calculation of $E[d_1]$

$E[d_1]$ is expressed by Equation 2, by summing the corresponding quantities. Similarly to the previous subsection, we end up with the expression:

$$\begin{aligned} E[d_1] &= (r-o) \sum_{a=r}^N \sum_{b=a+1}^N p(a) p(b) + \sum_{a=\lceil \frac{r+o}{2} \rceil}^r \sum_{b=r}^N p(a) p(b) [(a-o) - x] \\ &+ \sum_{a=o}^{\lfloor \frac{r+o}{2} \rfloor} \sum_{b=r}^N p(a) p(b) [(a-o) - (b-r)] \end{aligned}$$

$$\begin{aligned}
 &+ \sum_{a=1}^o \sum_{b=r}^N p(a) p(b) [\min(b - o, o - a) - (b - r)] \\
 &+ \sum_{a=\lceil \frac{r+o}{2} \rceil}^r \sum_{b=a+1}^r p(a) p(b) [(a - o) - (r - b)] \\
 &+ \sum_{a=o}^{\lfloor \frac{r+o}{2} \rfloor} \sum_{b=\lceil \frac{r+o}{2} \rceil}^r p(a) p(b) [(a - o) - (r - b)] \\
 &+ \sum_{a=1}^o \sum_{b=o}^r p(a) p(b) [\min(b - o, o - a) - (r - b)].
 \end{aligned}$$

Calculation of $E[d_2]$

Similarly, $E[d_2]$ is expressed by Equation 10 by adding the following summations:

$$\begin{aligned}
 E[d_2] = &\sum_{b=1}^o \sum_{n=b+1}^o p(b) p(n) [\min(o - n, n - b) - \min(r - n, n - b)] \\
 &+ \sum_{b=o}^r \sum_{n=b}^r p(b) p(n) [(n - b) - \min(r - n, n - b)] \\
 &+ \sum_{b=1}^o \sum_{n=r}^N p(b) p(n) (r - o) + \sum_{b=1}^o \sum_{n=o}^r p(b) p(n) [(n - o) - (r - n)] \\
 &+ \sum_{b=o}^r \sum_{n=1}^o p(b) p(n) (o - b) + \sum_{b=r}^N \sum_{n=1}^o p(b) p(n) (o - r) \\
 &+ \sum_{b=n}^r \sum_{n=o}^r p(b) p(n) [\min(b - n, n - o) - (b - n)] \\
 &+ \sum_{b=r}^N \sum_{n=o}^r p(b) p(n) [\min(b - n, n - o) - (r - n)] \\
 &+ \sum_{b=n}^N \sum_{n=r}^N p(b) p(n) [\min(b - n, n - o) - \min(n - r, b - n)].
 \end{aligned}$$

Again, the total seek $E[d]$ equals the sum of quantities $E[d_1]$ and $E[d_2]$.

3.3 Symmetric Replication Technique

This technique applies either Left or Right Replication technique depending on the original cylinder location compared to the middle of the disk. Thus, Symmetric Replication is governed by the following rule:

if original $\leq \frac{N}{2}$ then use Right Replication

if original $> \frac{N}{2}$ then use Left Replication

Evaluation of the expected seek is based on the adopted policy, i.e. in case of Left Replication we use the formulae derived in Subsection 3.1, whereas in case of Right Replication we use the formulae derived in Subsection 3.2.

4 Expected Seek Distance

The algorithm used for the evaluation of the expected seek for each of the replication schemes, in our replicated mirrored disk model, follows:

1. Use normal distribution for the data placement
2. Choose the R most frequent requested cylinders
3. For each of the R chosen evaluate their $E[d]$ and find the cylinder r which maximizes $E[d]$
4. Re-estimate the probabilities for cylinders o and r
5. Evaluate expected seek distance

After replication the probabilities of the original cylinder (o) and the replica cylinder (r) are updated since the distribution is affected by the introduction of the replica. The probability of the replica position is increased by an amount pr , whereas the probability of the original cylinder position is decreased by the same amount (pr). We have run the above algorithm for a model consisting of $N=100$ cylinders. The cylinder access probability obeys the normal distribution, whereas the organ-pipe placement scheme has been adopted. Two variations of the normal distribution are applied ($\sigma=15$ and $\sigma=40$). The expected seek is evaluated for each of these placement schemes by the algorithm described above. Figures 3 and

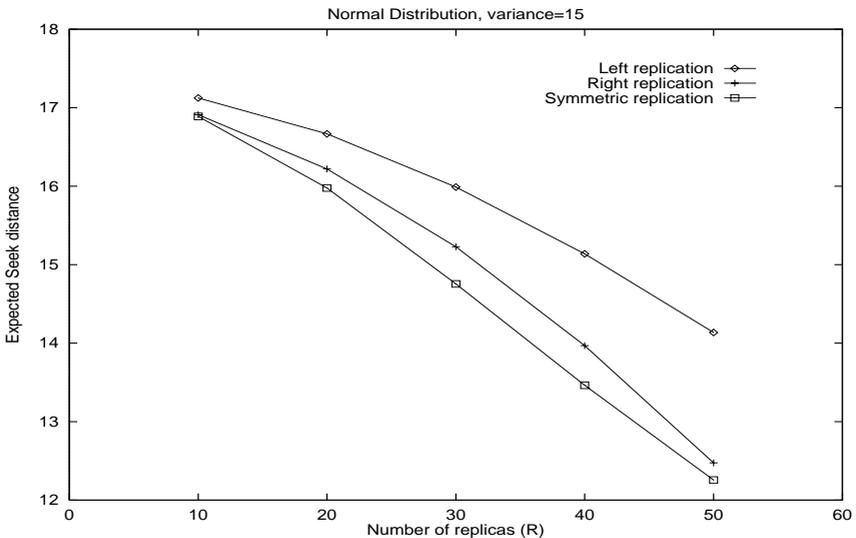


Fig. 3. Expected seek distance as a function of the number of replicas (for $\sigma=15$).

4 represent the expected seek distance metric for Organ-pipe arrangement which is expressed by Normal Distribution with variance $\sigma=15$ and $\sigma=40$ respectively. These results are compared with the corresponding expected seek derived for the non-replicated mirrored disk model. In [4] the expected seek was found to be $\frac{N}{5}$ for a mirrored disk system with no replication. For our deterministic model of $N=100$ cylinders the non-replicated model will result in an expected seek of 20 cylinders. This seek distance is used for comparisons with our results.

Left Replication

Here, the amount to be added/subtracted to the original probabilities is estimated by the probability of using the replica i.e.

$$pr = p(r) \sum_{i=1}^{\lceil \frac{r+\sigma}{2} - 1 \rceil} p(i)$$

Compared to the mirrored disk models with no replication, our model shows significant seek improvement. When using normal distribution with variance $\sigma=40$, we have a range of expected seeks from 18.5 cylinders for $R=10$ replicas to 16.8 cylinders for $R=50$. Thus, there is an improvement rate from 8% to 16% approximately. When using normal distribution with variance $\sigma=15$, we have a range of expected seeks from 17.1 cylinders for $R=10$ to 14.1 cylinders for $R=50$. Thus, there is an improvement rate from 14% to 30% approximately.

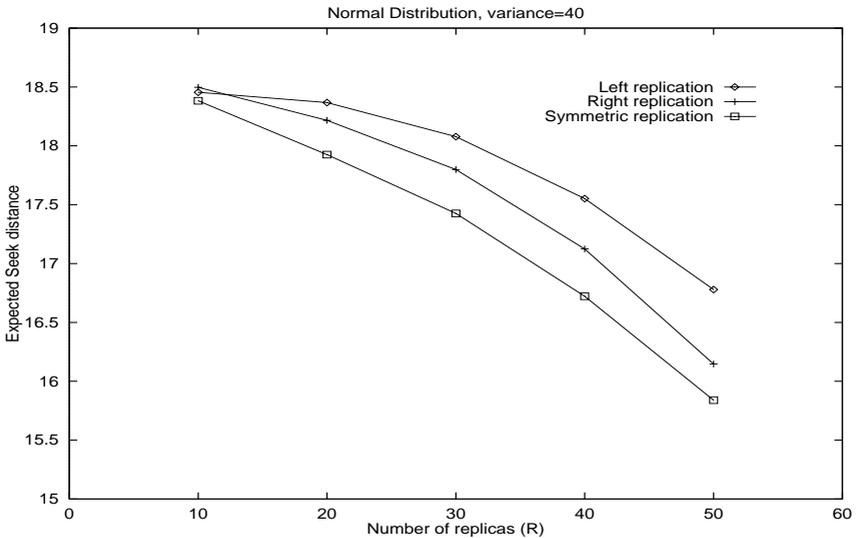


Fig. 4. Expected seek distance as a function of the number of replicas (for $\sigma=40$).

Right Replication

Here, the amount to be added/subtracted to the original probabilities is:

$$pr = p(r) \left(1 - \sum_{i=1}^{\lfloor \frac{r+\sigma}{2} \rfloor} p(i) \right)$$

When using normal distribution with variance $\sigma=40$, we have a range of expected seeks from 18.5 cylinders for $R=10$ to 16.2 cylinders for $R=50$. Thus there is an improvement rate from 8% to 19% approximately. When using normal distribution with variance $\sigma=15$, we have a range of expected seeks from 16.9 cylinders for $R=10$ to 12.5 cylinders for $R=50$. In this case, the improvement rate ranges from 15% to 38% approximately.

Symmetric Replication

In this case, when using normal distribution with variance $\sigma=40$, we have a range of expected seeks from 18.4 cylinders for $R=10$ to 15.8 cylinders for $R=50$. Thus, there is an improvement rate from 8% to 21% approximately compared to the non-replicated mirrored disk system. When using normal distribution with variance $\sigma=15$, we have a range of expected seeks from 16.9 cylinders for $R=10$ to 12.3 cylinders for $R=50$. Thus, there is an improvement rate from 16% to 39% approximately. Thus, the Symmetric Replication scheme shows a better behavior since it takes advantage of both techniques and exploits better the data distribution.

5 Conclusions - Further Research

A mirrored disk system is studied, under specific replication schemes which are proposed in order to exploit the free disk space and improve performance. Several analytic models are evaluated which show significant seek improvement. The improvement rates vary from 8% to 39% approximately. The improvement rate is affected by both the data placement distribution and the number of replicated cylinders. In every case, the performance is improved as the number of replicas increases. The normal distribution with variance $\sigma=40$ shows a weaker performance than the corresponding model with variance $\sigma=15$, as shown in Figures 3-4. This is the case, since Normal distribution with variance $\sigma=15$ produces a curve which is skewer (around the central cylinder) than the corresponding curve of variance $\sigma=40$. So, when $\sigma=15$ data are clustered in a narrower disk area resulting in reduced seek distances. Symmetric Replication shows the best behavior by reaching an improvement rate of 39% approximately. It is concluded that Symmetric Replication which combines both Left and Right Replication reduces seeking significantly.

Further research should extend the replication schemes presented here, by introducing adaptive block replication in mirrored disks. The analysis could be

supported by simulation experiments. Also, issues like the number of replicas, as well as “high” priority positions to place replicas could extend the presented models. The organ-pipe placement could be replaced by other data placement policies in order to compare different methods and possibly come up with a more efficient scheme. The replication idea could be adapted to modern optimal data placement strategies applied to technologically advanced large-scale storage media such as Tertiary Storage Libraries discussed in [7]. Also, the camel-arrangement [9] is a data placement scheme which can be applied to the models studied in the present work since there are two mirrored disks suitable for the two peaks of the camel-arrangement scheme.

References

1. S. Akyurek and K. Salem: Adaptive Block Rearrangement, *ACM Transactions on Computer Systems*, Vol.13, No.2, pp.89-121, 1995.
2. S. Akyurek and K. Salem: Adaptive Block Rearrangement Under UNIX, *Software - Practice and Experience*, Vol.27, No.1, pp.1-23, 1997.
3. D. Bitton: Arm Scheduling in Shadowed Disks, *Proceedings IEEE COMPCON 89 Conference*, pp.132-136, 1989.
4. D. Bitton and J. Gray: Disk Shadowing, *Proceedings 14th VLDB Conference*, pp.331-338, 1988.
5. S. Carson and P. Reynolds: Adaptive Disk Reorganization, Technical Report UM-IACS TR-89-4, University of Maryland at College Park, MD, 1989.
6. C. Chien and N. Bourbakis: Approximation of Seek Time for Merged Disks, *Transactions of the Society for Computer Simulation* Vol.12, No.3, pp.245-260, 1995.
7. S. Christodoulakis, P. Triantafyllou and F. Zioga: Principles of Optimally Placing Data in Tertiary Storage Libraries, *Proceedings 23rd VLDB Conference*, pp.236-245, 1997.
8. 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.
9. Y. Manolopoulos and J.G. Kollias: Optimal Data Placement in Two-headed Disk Systems, *BIT*, Vol.30, pp.216-219, 1990.
10. A. Vakali and Y. Manolopoulos: An Exact Analysis on Expected Seeks in Shadowed Disks, *Information Processing Letters*, Vol.61, pp.323-329, 1997.
11. P. Vongsathorn and S. Carson: A System for Adaptive Disk Rearrangement, *Software - Practice and Experience*, Vol.20, No.3, pp.225-242, 1989.
12. C. Wong: *Algorithmic Studies in Mass Storage Systems*, Computer Science Press, 1983.