Temporally Determinate Disk Access: An Experimental Approach

Mohamed Aboutabl, and Ashok Agrawala
Department of Computer Science
University of Maryland at College Park
aboutabl,agrawala@cs.umd.edu

Jean-Dominique Decotignie
Départment d'Informatique
École Polytechnique Fédérale De Lausanne
decotignie@di.epfl.ch

1 Introduction

The current level of unpredictability associated with the traditional approach for disk service is unacceptable to real-time applications. Consequently, early real-time applications avoided the use of disk drives and operated entirely in main memory.

In order to significantly reduce this variability, we should have a better understanding of its cause(s), and use this knowledge in the design of better device drivers. A typical disk I/O operation incorporates delay components such as: *seek* delay, *rotational* delay, *off-surface transfer* delay, and *host transfer* delay. The last two delay components typically show very small variability. On the other hand, seek and rotational delays have shown variability in the order of 10's of milliseconds.

In order to design a device driver which can accurately predict the service time of a disk I/O request, it is necessary that not only the disk dynamics be known, but that the physical layout of the information on the disk be also reflected in the temporal prediction. The details needed for accurate temporal predictions are not easily available. In this paper, we present an approach for modeling the disk for the purpose of designing a device driver with the ability to accurately predict the service time of a request. This approach is based on a set of carefully synthesized experiments that analyse the temporal behavior of the disk drive in response to I/O requests. The methodology presented here can be applied to any disk drive.

Worthington described, in [Worthington95], a suite of general-purpose techniques and algorithms for acquiring data on the structure and organization of SCSI disks via the ANSI-standard interface. We present techniques that measure more disk parameters such as sector rotation time, track skewness, and boundaries of the recording zones. We also present, where applicable, a subjective study of the extracted disk parameters. Moreover, we use a 3^{rd}-degree piece-wise polynomial to model the seek time, compared to the square-root curve suggested by previous researchers. One of our major contributions is an analytic model to predict the disk I/O service time. This model was proven to predict the disk dynamics and physical geometry to a high degree of accuracy.

2 Design Considerations and the Run-Time System

In this section we discuss various issues affecting the design of our experiments. These experiments were conducted using a 1 GB AC21000 Western Digital © hard disk. We conducted our experiments under the Maruti real-time operating system. Maruti [Marti94] guarantees that the process will run without being preempted, which is helpful when time measurements are taken.

All time measurements were taken with the help of the *maruti_get_current_time* library function which reads the current system time directly from the the clock chip on the motherboard. We conducted several experiments to measure the performance of this function and found out it has an accuracy of ± 3 microseconds.

We modefied the disk device driver that is bundled with Net-BSD such that the interrupt-driven nature of the device driver was replaced by a busy-waiting loop. This decision was mandated by the need for a higher accuracy in our time measurements.

We also observed that when the controller read-ahead cache is enabled, it is difficult to extract any accurate information about the physical characteristics of a disk drive. Therefore, we had to disable the read-ahead controller cache.

3 Measuring The Disk Parameters

In most experiments, we issue two requests to the disk, one immediately after the other. In order to minimize the time before we may issue the next command to the controller, we do not transfer the data accessed by the first command. We define $t(k,l)$ as the time elapsed between the completion of accessing sector k and the completion of accessing sector l, provided that the second request is issued as soon as the first one has been completed (figure 1).

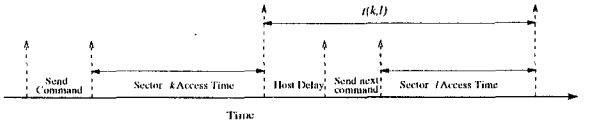

Figure 1: Two successive requests

The first disk parameter we measured was the disk rotation time (DRT), because it was essential in extracting

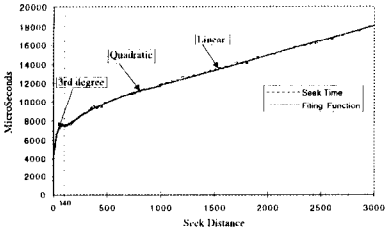

Figure 2: Piece-wise Polynomial Fit for the Seek Curve

all other disk parameters. We measured the time $t(k,k)$ between two successive completions of accessing the same block k. This is equal to one revolution time. We took similar readings for different blocks at scattered locations on the disk. It was observed that the disk rotation time had a mean of 11534 microseconds and a standard deviation of 1.83 microseconds. The experiment was repeated several times measuring the rotation time at the same set of blocks, and the same readings were obtained.

Next, we used readings of $t(k,l)$ for pre-selected blocks k and l to determine the exact physical geometry of the disk drive. This included the number of sectors per track, and the number of cylinders per zone.

The time required for one sector, say n, to rotate completely under the read/write head, denoted by $SRT(n)$, was computed as $t(n-1,n) - DRT$. This time increases as we move towards the center of the disk, since the number of sectors per track decreases in that direction.

Other experiments [Aboutabl97] were designed to measure the controller overhead time (COT), the head switching time (HST), as well as other disk parameters. The seek time was also measured as a function of the seek distance and found to be best represented as a 3^{rd}-degree piece-wise polynomial as shown in figure 2.

4 Prediction of the Disk Request Service Time

As shown in figure 3, we assume that the a disk I/O request has been issued at time t. At that instance, the position of the read/write head, denoted by $p(t)$, is on track $k-n$. Let the requested sector s be on track k, which is n tracks away. Let L represent the time for the disk to rotate from the current position $p(t)$ and arrive at the target sector s in track k [1]. The total service time of the request can be

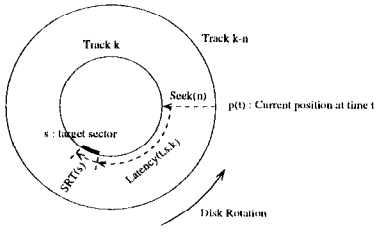

Figure 3: Disk Request Access Time

[1] Refer to [Aboutabl97] for computation details

estimated by the following equations:

$$
\begin{aligned}
Revs &= \left\lfloor \frac{Seek(n)}{DRT} \right\rfloor \\
Threshold &= (\ Seek(n) \bmod DRT\ ) + HST \\
Service\ Time &= SRT(s) + Revs * DRT \\
&\quad + \begin{cases} L & : \text{if L} \geq \text{Threshold} \\ L + DRT & : \text{if L} < \text{Threshold} \end{cases}
\end{aligned}
$$

5 Disk Model Validation

We conducted several experiments in which we issued disk read requests and compared the observed service times with the predictions obtained from our model.

More than 19200 disk blocks were randomly selected from all over the disk. As shown in table 1, our model predicted 96% of the service times within ±200 microseconds of the actually measured values. In 3% of the cases, our prediction deviated from the actual service time by one revolution. According to our model, if our predicted value of the service time misses the requested block by as low as 1 microsecond in the current revolution, then we will have to wait for the block to rotate up to one extra revolution before we can catch it. We believe that the model accuracy can be improved by using a more accurate tool to take time measurements.

Prediction Error	Frequency	Average	Percent
0 to 100	14649	46	76%
100 to 200	3866	148	20%
200 to 300	225	219	1%
One DRT	494	11539	3%
Total	19234		100%

Table 1: Prediction Error (in μSeconds)

References

[Aboutabl97] Mohamed Aboutabl, Ashok K. Agrawala, Jean-Dominique Decotignie. "Temporally Determinate Disk Access: An Experimental Approach." *University of Maryland at College Park* CS technical report CS-TR-3752, 1997.

[Worthington95] Bruce L. Worthington, et. al. "On-Line Extraction of SCSI Disk Drive Parameters." In *Proceedings of the 1995 ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems,* pages 146-156, May 1995.

[Ruemmler94] Chris Ruemmler and John Wilkes. An Introduction to Disk Drive Modeling. *IEEE Computer,* pages 17-28, March 1994.

[Marti94] M. Saksena, J. da Silva and A. K. Agrawala, "Design and Implementation of Maruti-II." *Principles of Real-Time Systems* Sang Son (ed.), 1994. Also available as UMD CS-TR-3181, UMICAS TR-93-122.