Mass Storage Structure

Practice Exercises

11.1 Is disk scheduling, other than FCFS scheduling, useful in a single-user environment? Explain your answer.

Answer: In a single-user environment, the I/O queue usually is empty. Requests generally arrive from a single process for one block or for a sequence of consecutive blocks. In these cases, FCFS is an economical method of disk scheduling. But LOOK is nearly as easy to program and will give much better performance when multiple processes are performing concurrent I/O, such as when a Web browser retrieves data in the background while the operating system is paging and another application is active in the foreground.

11.2 Explain why SSTF scheduling tends to favor middle cylinders over the innermost and outermost cylinders.

Answer: The center of the disk is the location having the smallest average distance to all other tracks. Thus the disk head tends to move away from the edges of the disk. Here is another way to think of it. The current location of the head divides the cylinders into two groups. If the head is not in the center of the disk and a new request arrives, the new request is more likely to be in the group that includes the center of the disk; thus, the head is more likely to move in that direction.

11.3 Why is rotational latency usually not considered in disk scheduling? How would you modify SSTF, SCAN, and C-SCAN to include latency optimization?

Answer: Most disks do not export their rotational position information to the host. Even if they did, the time for this information to reach the scheduler would be subject to imprecision and the time consumed by the scheduler is variable, so the rotational position information would become incorrect. Further, the disk requests are usually given in terms of logical block numbers, and the mapping between logical blocks and physical locations is very complex.

11.4 How would use of a RAM disk affect your selection of a disk-scheduling algorithm? What factors would you need to consider? Do the same considerations apply to hard-disk scheduling, given that the file system stores recently used blocks in a buffer cache in main memory?

Answer: Disk scheduling attempts to reduce the overhead time of disk head positioning. Since a RAM disk has uniform access times, scheduling is largely unnecessary. The comparison between RAM disk and the main memory disk-cache has no implications for hard-disk scheduling because we schedule only the buffer cache misses, not the requests that find their data in main memory.

11.5 Why is it important to balance file system I/O among the disks and controllers on a system in a multitasking environment?

Answer: A system can perform only at the speed of its slowest bottleneck. Disks or disk controllers are frequently the bottleneck in modern systems as their individual performance cannot keep up with that of the CPU and system bus. By balancing I/O among disks and controllers, neither an individual disk nor a controller is overwhelmed, so that bottleneck is avoided.

11.6 What are the tradeoffs involved in rereading code pages from the file system versus using swap space to store them?

Answer: If code pages are stored in swap space, they can be transferred more quickly to main memory (because swap space allocation is tuned for faster performance than general file system allocation). Using swap space can require startup time if the pages are copied there at process invocation rather than just being paged out to swap space on demand. Also, more swap space must be allocated if it is used for both code and data pages.

11.7 Is there any way to implement truly stable storage? Explain your answer.

Answer: Truly stable storage would never lose data. The fundamental technique for stable storage is to maintain multiple copies of the data, so that if one copy is destroyed, some other copy is still available for use. But for any scheme, we can imagine a large enough disaster that all copies are destroyed.

11.8 The term "Fast Wide SCSI-II" denotes a SCSI bus that operates at a data rate of 20 megabytes per second when it moves a packet of bytes between the host and a device. Suppose that a Fast Wide SCSI-II disk drive spins at 7,200 RPM, has a sector size of 512 bytes, and holds 160 sectors per track.

a. Estimate the sustained transfer rate of this drive in megabytes per second.

b. Suppose that the drive has 7,000 cylinders, 20 tracks per cylinder, a head-switch time (from one platter to another) of 0.5 millisecond, and an adjacent-cylinder seek time of 2 milliseconds. Use this additional information to give an accurate estimate of the sustained transfer rate for a huge transfer.

c. Suppose that the average seek time for the drive is 8 milliseconds. Estimate the I/O operations per second and the effective transfer rate for a random-access workload that reads individual sectors that are scattered across the disk.

d. Calculate the random-access I/O operations per second and transfer rate for I/O sizes of 4 kilobytes, 8 kilobytes, and 64 kilobytes.

e. If multiple requests are in the queue, a scheduling algorithm such as SCAN should be able to reduce the average seek distance. Suppose that a random-access workload is reading 8-kilobyte pages, the average queue length is 10, and the scheduling algorithm reduces the average seek time to 3 milliseconds. Now calculate the I/O operations per second and the effective transfer rate of the drive.

Answer:

a. The disk spins 120 times per second, and each spin transfers a track of 80 KB. Thus, the sustained transfer rate can be approximated as 9600 KB/s.

b. Suppose that 100 cylinders is a huge transfer. The transfer rate is total bytes divided by total time. Bytes: 100 cyl * 20 trk/cyl * 80 KB/trk, i.e., 160,000 KB. Time: rotation time + track switch time + cylinder switch time. Rotation time is 2000 trks/120 trks_per_sec, i.e., 16.667 s. Track switch time is 19 switch_per_cyl * 100 cyl * 0.5 ms, i.e., 950 ms. Cylinder switch time is 99 * 2 ms, i.e., 198 ms. Thus, the total time is 16.667 + 0.950 + 0.198, i.e., 17.815 s. (We are ignoring any initial seek and rotational latency, which might add about 12 ms to the schedule, i.e. 0.1%.) Thus the transfer rate is 8981.2 KB/s. The overhead of track and cylinder switching is about 6.5%.

c. The time per transfer is 8 ms to seek + 4.167 ms average rotational latency + 0.052 ms (calculated from 1/(120 trk_per_second * 160 sector_per_trk)) to rotate one sector past the disk head during reading. We calculate the transfers per second as 1/(0.012219), i.e., 81.8. Since each transfer is 0.5 KB, the transfer rate is 40.9 KB/s.

d. We ignore track and cylinder crossings for simplicity. For reads of size 4 KB, 8 KB, and 64 KB, the corresponding I/Os per second are calculated from the seek, rotational latency, and rotational transfer time as in the previous item, giving (respectively) 1/(0.0126), 1/(0.013), and 1/(0.019). Thus we get 79.4, 76.9, and 52.6 transfers per second, respectively. Transfer rates are obtained from 4, 8, and 64 times these I/O rates, giving 318 KB/s, 615 KB/s, and 3366 KB/s, respectively.

e. From 1/(3+4.167+0.83) we obtain 125 I/Os per second. From 8 KB per I/O we obtain 1000 KB/s.

11.9 More than one disk drive can be attached to a SCSI bus. In particular, a Fast Wide SCSI-II bus (see Exercise 11.8) can be connected to at most 15 disk drives. Recall that this bus has a bandwidth of 20 megabytes per second. At any time, only one packet can be transferred on the bus between some disk's internal cache and the host. However, a disk can be moving its disk arm while some other disk is transferring a packet on the bus. Also, a disk can be transferring data between its magnetic platters and its internal cache while some other disk is transferring a packet on the bus. Considering the transfer rates that you calculated for the various workloads in Exercise 11.8, discuss how many disks can be used effectively by one Fast Wide SCSI-II bus.

Answer: For 8 KB random I/Os on a lightly loaded disk, where the random access time is calculated to be about 13 ms (see Exercise 11.8), the effective transfer rate is about 615 MB/s. In this case, 15 disks would have an aggregate transfer rate of less than 10 MB/s, which should not saturate the bus. For 64 KB random reads to a lightly loaded disk, the transfer rate is about 3.4 MB/s, so five or fewer disk drives would saturate the bus. For 8 KB reads with a large enough queue to reduce the average seek to 3 ms, the transfer rate is about 1 MB/s, so the bus bandwidth may be adequate to accommodate 15 disks.

11.10 Remapping bad blocks by sector sparing or sector slipping can influence performance. Suppose that the drive in Exercise 11.8 has a total of 100 bad sectors at random locations and that each bad sector is mapped to a spare that is located on a different track within the same cylinder. Estimate the number of I/O operations per second and the effective transfer rate for a random-access workload consisting of 8-kilobyte reads, assuming a queue length of 1 (that is, the choice of scheduling algorithm is not a factor). What is the effect of a bad sector on performance?

Answer: Since the disk holds 22,400,000 sectors, the probability of requesting one of the 100 remapped sectors is very small. An example of a worst-case event is that we attempt to read, say, an 8 KB page, but one sector from the middle is defective and has been remapped to the worst possible location on another track in that cylinder. In this case, the time for the retrieval could be 8 ms to seek, plus two track switches and two full rotational latencies. It is likely that a modern controller would read all the requested good sectors from the original track before switching to the spare track to retrieve the remapped sector and thus would incur only one track switch and rotational latency. So the time would be 8 ms seek + 4.17 ms average rotational latency + 0.05 ms track switch + 8.3 ms rotational latency + 0.83 ms read time (8 KBis 16 sectors, 1/10 of a track rotation). Thus, the time to service this request would be 21.8 ms, giving an I/O rate of 45.9 requests per second and an effective bandwidth of 367 KB/s. For a severely time-constrained application this might matter, but the overall impact in the weighted average of 100 remapped sectors and 22.4 million good sectors is nil.

11.11 In a disk jukebox, what would be the effect of having more open files than the number of drives in the jukebox?

Answer: Two bad outcomes could result. One possibility is starvation of the applications that issue blocking I/Os to tapes that are not mounted in drives. Another possibility is thrashing, as the jukebox is commanded to switch tapes after every I/O operation.

11.12 If magnetic hard disks eventually have the same cost per gigabyte as do tapes, will tapes become obsolete, or will they still be needed? Explain your answer.

Answer: Tapes are easily removable, so they are useful for off-site backups and for bulk transfer of data (by sending cartridges). As a rule, a magnetic hard disk is not a removable medium.

11.13 It is sometimes said that tape is a sequential-access medium, whereas a magnetic disk is a random-access medium. In fact, the suitability of a storage device for random access depends on the transfer size. The term *streaming transfer rate* denotes the rate for a data transfer that is underway, excluding the effect of access latency. By contrast, the *effective transfer rate* is the ratio of total bytes per total seconds, including overhead time such as access latency.

Suppose that, in a computer, the level-2 cache has an access latency of 8 nanoseconds and a streaming transfer rate of 800 megabytes per second, the main memory has an access latency of 60 nanoseconds and a streaming transfer rate of 80 megabytes per second, the magnetic disk has an access latency of 15 milliseconds and a streaming transfer rate of 5 megabytes per second, and a tape drive has an access latency of 60 seconds and a streaming transfer rate of 2 megabytes per seconds.

a. Random access causes the effective transfer rate of a device to decrease, because no data are transferred during the access time. For the disk described, what is the effective transfer rate if an average access is followed by a streaming transfer of (1) 512 bytes, (2) 8 kilobytes, (3) 1 megabyte, and (4) 16 megabytes?

b. The utilization of a device is the ratio of effective transfer rate to streaming transfer rate. Calculate the utilization of the disk drive for each of the four transfer sizes given in part a.

c. Suppose that a utilization of 25 percent (or higher) is considered acceptable. Using the performance figures given, compute the smallest transfer size for disk that gives acceptable utilization.

d. Complete the following sentence: A disk is a random-access device for transfers larger than _____ bytes and is a sequential-access device for smaller transfers.

e. Compute the minimum transfer sizes that give acceptable utilization for cache, memory, and tape.

f. When is a tape a random-access device, and when is it a sequential-access device?

Answer:

a. For 512 bytes, the effective transfer rate is calculated as follows. ETR = transfer size/transfer time.

If X is transfer size, then transfer time is ((X/STR) + latency).
Transfer time is 15ms + (512B/5MB per second) = 15.0097ms.
Effective transfer rate is therefore 512B/15.0097ms = 33.12 KB/sec.
ETR for 8KB = .47MB/sec.
ETR for 1MB = 4.65MB/sec.
ETR for 16MB = 4.98MB/sec.

b. Utilization of the device for 512B = 33.12 KB/sec / 5MB/sec = .0064 = .64
For 8KB = 9.4%.
For 1MB = 93%.
For 16MB = 99.6%.

c. Calculate .25 = ETR/STR, solving for transfer size X.
STR = 5MB, so 1.25MB/S = ETR.
1.25MB/S * ((X/5) + .015) = X.
.25X + .01875 = X.
X = .025MB.

d. A disk is a random-access device for transfers larger than K bytes (where K > disk block size), and is a sequential-access device for smaller transfers.

e. Calculate minimum transfer size for acceptable utilization of cache memory:
STR = 800MB, ETR = 200, latency = $8 * 10^{-9}$.
200 (XMB/800 + 8 X 10^{-9}) = XMB.
.25XMB + 1600 * 10^{-9} = XMB.
X = 2.24 bytes.
Calculate for memory:
STR = 80MB, ETR = 20, L = $60 * 10^{-9}$.
20 (XMB/80 + $60 * 10^{-9}$) = XMB.
.25XMB + $1200 * 10^{-9}$ = XMB.
X = 1.68 bytes.
Calculate for tape:
STR = 2MB, ETR = .5, L = 60s.
.5 (XMB/2 + 60) = XMB.
.25XMB + 30 = XMB.
X = 40MB.

f. It depends upon how it is being used. Assume we are using the tape to restore a backup. In this instance, the tape acts as a sequential-access device where we are sequentially reading the contents of the tape. As another example, assume we are using the tape to access a variety of records stored on the tape. In this instance, access to the tape is arbitrary and hence considered random.

11.14 Suppose that we agree that 1 kilobyte is 1,024 bytes, 1 megabyte is $1{,}024^2$ bytes, and 1 gigabyte is $1{,}024^3$ bytes. This progression continues through terabytes, petabytes, and exabytes ($1{,}024^6$). Several proposed scientific projects plan to record and store a few exabytes of data during the next decade. To answer the following questions, you will need to

make a few reasonable assumptions; state the assumptions that you make.

a. How many disk drives would be required to hold 4 exabytes of data?

b. How many magnetic tapes would be required to hold 4 exabytes of data?

c. How many optical tapes would be required to hold 4 exabytes of data (see Exercise 11.34)?

d. How many holographic storage cartridges would be required to hold 4 exabytes of data (see Exercise 11.33)?

e. How many cubic feet of storage space would each option require?

Answer:

a. Assume that a disk drive holds 10 GB. Then 100 disks hold 1 TB, 100,000 disks hold 1 PB, and 100,000,000 disks hold 1 EB. To store 4 EB would require about 400 million disks.

b. If a magnetic tape holds 40 GB, only 100 million tapes would be required.

c. If an optical tape holds 50 times more data than a magnetic tape, 2 million optical tapes would suffice.

d. If a holographic cartridge can store 180 GB, about 22.2 million cartridges would be required.

e. Storage space:
   i. A 3.5" disk drive is about 1" high, 4" wide, and 6" deep. In feet, this is 1/12 by 1/3 by 1/2, or 1/72 cubic feet. Packed densely, the 400 million disks would occupy 5.6 million cubic feet. If we allow a factor of two for air space and space for power supplies, the required capacity is about 11 million cubic feet.
   ii. A 1/2" tape cartridge is about 1" high and 4.5" square. The volume is about 1/85 cubic feet. For 100 million magnetic tapes packed densely, the volume is about 1.2 million cubic feet.
   iii. For 2 million optical tapes, the volume is 23,400 cubic feet.
   iv. A CD-ROM is 4.75" in diameter and about 1/16" thick. If we assume that a holostore disk is stored in a library slot that is 5" square and 1/8" wide, we calculate the volume of 22.2 million disks to be about 40,000 cubic feet.