

OPERATING SYSTEM

Lecture 9

Input/Output Management

Dr Victoria Mukami

INTRODUCTION

This lecture focuses on input and output management. We will first review the categories of I/O devices, how I/O performance can be improved and the various times that are used during the performance. In addition, we will review the organization of the I/O function.

Learning objectives

By the end of this topic, you should be able to:

1. Describe the categories of I/O devices
2. Understand the organization of the I/O function
3. Understand the various times that are used to calculate I/O performance

OVERVIEW

So far, we have talked about input/output requests as resources that are required by jobs, processes, and threads. These requests are normally fulfilled by the processor on a job-by-job basis. This lecture focuses on input/output management. The device manager oversees managing any peripheral devices required by the system. Device management normally involves four basic functions:

- Device status monitoring
- Policy enforcement
- Device allocation
- Device deallocation

I/O DEVICES

External devices that engage in I/O with the processor can be grouped into three main categories.

Human readable: These are devices that can communicate with the end user. Think output and input devices. Devices such as monitors, keyboards, mice, and printers are good examples of human-readable devices.

Machine-readable: These are the opposite of human-readable as they can communicate with electronic equipment such as USB keys, sensors and controllers [2].

Communication: these devices enable communication with remote devices. These include routers, modems, and digital line drivers [2].

In addition to these categories, devices can be categorized based on their functionality. Devices are categorized as either dedicated, shared or virtual [1].

Dedicated: These are devices that are assigned only one job at a time [1]. They can only work on one job until they complete the job or release it. A printer is a good example of such a device. The printer will only print one job at a time and another job cannot be started until the previous job is completed. The biggest drawback to dedicated devices is that only one user can be allocated a device for the duration of the job. This is a problem as the user may not use the device 100% of the time.

Shared: These devices can be allocated several jobs at the same time. A hard disk is a good example of a shared device as it can work with several processes. This is done through interleaving which has to be handled carefully to prevent conflicts [1].

Virtual: These devices are a combination of dedicated and shared devices. They are dedicated devices that have been transformed into sharing devices [1]. Printers for instance can be shared using spooling or using servers that can hold several jobs. This of course means management is key to avoiding deadlocks during allocation to the devices.

EVOLUTION OF I/O DEVICES

Like the evolution of computers so has there been an increase in the complexity of the I/O functions. The following are the evolution steps [2].

1. The processor directly controlled a peripheral device. This was seen as a microprocessor-controlled device.
2. The second was where the controller or I/O module was added. The processor no longer controlled the peripheral devices directly. The processor uses the I/O module without interrupts.
3. In this step, interrupts were introduced thereby saving the processor time as it did not have to wait for I/O requests to be carried out.
4. The I/O module was given direct control of memory during this stage. The I/O module could move data to and from memory without involving the processor.

5. This stage involved the I/O module by making it a separate processor. The I/O processor can fetch and execute these instructions without processor intervention [2].
6. Finally, in the current stage, the I/O module has a local memory of its own. Several I/O devices are controlled with minimal processor involvement [2].

MANAGEMENT OF I/O REQUESTS

The device manager manages the I/O requests through three main tasks that are handled by a specific component.

I/O Traffic controller:

This controller checks the status of all the i/o devices, control units and channels [1]. The controller [1] must determine if there is a path available to a requested device. If more than one path exists, then the controller must determine which path to select. Finally, the controller must determine which path will be available in the event all paths are busy. This is done through a database that contains the status and connections for each unit within the I/O module.

I/O Scheduler

This oversees the allocation and deallocation of devices [1]. It implements policies already in place for the same. The scheduler behaves the same as the process scheduler. It must decide how to satisfy requests and which requests must be satisfied first. No pre-emption happens with the I/O scheduler as compared to the process scheduler.

I/O device handler

This is the part in charge of the actual transfer of data and processes [1]. It processes the I/O interrupts, handles errors that may arise and provides detailed scheduling algorithms [1]. Each device would have its device handler.

DIRECT ACCESS STORAGE DEVICES

These are devices that can read or write to random areas within the storage. These include magnetic disks, optical disks and solid-state drives. This section looks at these devices and how management occurs within them.

Magnetic Disk Storage

These are storage devices such as hard disks and floppy disks. Magnetic disk drives usually have a stack of platters where the read and write occurs. These are usually held together and rotated using a central spindle. There is space in between the platters that allow read and write heads to read and write to the platter. The platters are two-sided meaning that the read and write heads are found both at the top and bottom of the platter. The arms are held by a controller on the side of the platter and when the top head moves, all the others move too. Figure 1 shows an illustration of the magnetic disk structure.

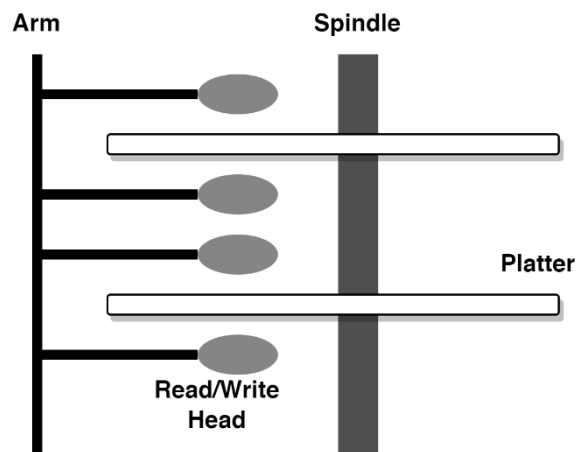


Figure 1: Magnetic disk structure (Adapted [1])

Let us discuss three times that are used when reviewing access to a file.

Seek time: This is the slowest time and is the time required to position the read/ write arm.

Search time: This is the rotational delay, that is, the time it takes to rotate the disk until the requested record is moved under the read/write head.

Transfer time: fastest time. It is the time when data is transferred from the disk to the main memory.

The I/O device handler has a seek strategy which is a predetermined policy that the device handler uses to allocate access to the device among the many processes that may be waiting for it [1]. This strategy determines the order in which the processes get

the device. The goal of this strategy is to keep the seek time at a minimum. Remember the seek time is considered the slowest time among access times. Several strategies exist, First-Come First-Served (FCFS), Shortest Seek Time First (SSTF) and SCAN. All these algorithms aim to minimize arm movement, minimize mean response time, and minimize the variance in the response time [1]. The following is a summary of the strategies.

- FCFS works well with shorter queues of jobs. When the waiting list grows the service time becomes long.
- SSTF is popular and appealing and works well with average queues of jobs. When introduced to heavy or longer queues then it is unable to manage and lower the seek time.
- SCAN is like SSTF regarding throughput and the mean service time. It mainly works well with average queues.

Optical Disk Storage

These disks are somewhat like magnetic disks except for the number of tracks, size of sectors, spinning speed and the surface layer. While a magnetic disk platter has many platters, the optical disc has only one spiral disk. Several types of optical disks exist CDs, DVDs, and Blu-Rays. The difference between the three is in the way that the devices write data to the discs.

Solid State Drives

These drives are fast but costly as they use flash memory. Unlike magnetic drives, the SSD does not have any movable parts therefore seek and search times do not apply to these devices. The difference between the magnetic disks and the SSD is based on access times, power consumption, moving parts, cost, capacity, and noise production. Access times within the hard disk are much slower in comparison to the solid-state drive.

RAID

RAID stands for Redundant Array of Independent Disks. RAID works on the premise that several small-capacity disks are better than having fewer larger-capacity disks. RAID disks result in faster I/O performance and improved data recovery in case a disk

fails [1]. A normal RAID may have up to 5 disks connected to a controller that allows for the transfer of data. RAID has seven levels from Level 0 to 6 with different I/O request rates, different error correction methods and different data transfer rates.

SUMMARY

This lecture was a focus on input and output management. We have reviewed the categories of I/O devices, how I/O performance could be improved and the various times that are used during I/O performance. In addition, we reviewed the organization of the I/O function.

DISCUSSION TOPIC

Conduct small web research and identify three areas that may benefit from the use of RAID drives. Identify how access times would be affected by using RAID drivers.

REFERENCES

[1] McHoes, A., & Flynn, I., Understanding Operating Systems. Boston: Cengage Learning, 2018

[2] Stallings, W., Operating Systems: Internals and Design Principles. Harlow: Pearson Education Limited, 2018.