

Memory management in unix operating system computer science essay

[Technology](#), [Computer](#)



Introduction to UNIX

According to Leon, 2007, UNIX is an operating system (OS) is software that manages hardware and software resources of a computer. UNIX was first developed in the 1960s and has been constant development ever since.

UNIX is one of the most widely used operating systems in industry, government and education. It is a stable, multi-user, multi-tasking system for servers, desktops and laptops.

UNIX Memory Management

Memory is an important resource in computer. Memory management is the process of managing the computer memory which consists of primary memory and secondary memory. The goal for memory management is to keep track of which parts of memory are in use and which parts are not in use, to allocate memory to processes when they need it and de-allocate it when they are done. UNIX memory management scheme includes swapping and demand paging.

Memory Partitioning

The simplest form of memory management is splitting up the main memory into multiple logical spaces called partition. Each partition is used for separate program. There are 2 types of memory partitioning:-

Single Partition Allocation

Single partition allocation only separates the main memory into operating system and one user process area. Operating system will not able to have

<https://assignbuster.com/memory-management-in-unix-operating-system-computer-science-essay/>

virtual memory using single partition. Using single partition is very ineffective because it only allows one process to run in the memory at one time.

C: UsersEricDesktophssn picsingle partition. jpg

Figure 1. 1: Single Partitioning

Multiple Partition Allocation

Most of the operating system nowadays is using multiple partitions because it is more flexible. Multiple partition allocation enabled multiple programs run in the main memory at once. Each partition is used for one process.

There are two different forms of multiple partition allocation, which is fixed partitioning and variable partitioning. Fixed partitioning divides memory up into many fixed partitions which cannot be change. However, variable partitioning is more flexible because the partitions vary dynamically in the later as processes come and go. Variable partitioning (Variable memory) has been used in UNIX.

C: UsersEricDesktophssn picfix variable partition. jpg

Figure 1. 2: Fixed Partitioning and Variable Partitioning

UNIX Memory Management Strategies

Overlays

Program will be placed into memory during execution. However, a large program will divide into small pieces and load the pieces as they are needed. Overlays will replace the new pieces with the program which is unused. UNIX is using this technique to run a new program by forking the running process which is also known as fork-exec. The overlays technique is illustrated below.

C: UsersEricDesktophssn picoverlay. jpg

Figure 1. 3: Overlay Process

Swapping

Swapping consists of bringing in each process in physical memory entirely and running it. When the process is no longer in use, the process will be terminated or swapped out to disk.

The procedure of swapping is illustrated in figure 1. 3 below.

C: UsersEricDesktophssn picswapping. jpg

Figure 1. 4: Swapping

Initially only process A is in memory. Then process B is swapped into memory from disk. After that, process A terminates or is swapped out to disk. Then process C is swapped into the free space.

External Fragmentation Problem

The size of each process is different, therefore when the processes is been swapped in and out, there will be a multiple holes in the memory because UNIX is using variable partitioning.

C: UsersEricDesktophssn picholes. jpg

Figure 1. 5: External Fragmentation

Solution

There are two techniques to solve this problem, which are memory compaction and fit in the process using algorithms

Compaction

C: UsersEricDesktophssn piccompaction. jpg

Figure 1. 6: Compaction

Memory compaction moves all the processes upward as far as possible, so that all the free memory is placed in one large block. However, it is not a good idea because it requires a lots of CPU time.

Most processes will grow as they run, and the processes data segments can grow, as in many programming languages, the process will grow. If there is a hole is next to the process, it can be allocated and the process is allowed to grow into the hole. Therefore it is good to allocate some extra memory whenever a process is swapped in or out.

Figure 1. 7: Allocating Space for Growing Data (Extracted from <http://www.cs.rpi.edu/~hollingd/opsys-fall2006/notes/Chapter4/Chapter4.pdf> , 22 April 2011)

Algorithms

There are three different types of algorithm can be used to loads the program wherever the memory space is unused, which is first fit, best fit and worst fit.

Algorithms

Descriptions

First Fit

The memory manager scans along the list and allocates the first space to fit the process. First fit is a fast algorithm because it searches as little as possible.

Best Fit

The memory manager scans the whole list and takes the smallest hole that will fit the process. Best fit is slower than first fit because it must search the whole list every time it is called.

Worst Fit

The memory manager scans the whole list and takes the largest available hole, so that the hole broken will be big enough to be useful.

Virtual Memory

UNIX operating system allows user to fully utilize the physical memory installed in the system as well as part of the hard disk called swap space which have been designated for use by the kernel while the physical memory is insufficient to handle the tasks.

Virtual memory managers will create a virtual address space in secondary memory (hard disk) and it will determine the part of address space to be loaded into physical memory at any given time. The benefit of virtual memory relies on separation of logical and physical memory.

C: UsersEricDesktophssn picVR. jpg

Figure 1. 8: Logical Memory and Physical Memory

Demand Paging

Paging is a memory allocation strategy by transferring a fixed-sized unit of the virtual address space called virtual page whenever the page is needed to execute a program. As the size of frames and pages are the same, any logical page can be placed in any physical frame of memory.

Every processes will be logical divided and allocate in the virtual address space. There is a page table in the virtual memory to allocate and keep tracking of the pages to map into the frames.

C: UsersEricDesktophssn picpaging pic. jpg

Figure 1. 9: The relation between virtual addresses and physical memory addresses is given by the page table

UNIX will perform page swapping only when the program needs a certain page. This procedure is called demand paging. The page will be paged into the memory only when it is needed to execute. The whole process will not be paged into the memory, only the pages needed are swapped in.

Demand paging decreases the paging time and physical memory needed because only the needed pages will be paged and the reading time of the unused pages can be avoided.

C: UsersEricDesktophssn picdeand paging. jpg

Figure 1. 10: Demand paging

As the example in the figure, only page 6, 7, 8 and 9 is needed in Process A. Therefore, only pages 6, 7, 8 and 9 will be paged into the memory.

Page Fault Problem

A page fault occurs when a program try to use a page that is not in the memory, due to demand paging will only paged the pages into the memory when it is needed. For example in figure 1. 9, if the program try to use Page 1 for Process A in memory, the operating system will interrupt occurs as a result of trying access a missing page because Page 1 is not paged in the memory.

Solution for Page Fault

<https://assignbuster.com/memory-management-in-unix-operating-system-computer-science-essay/>

C: UsersEricDesktophssn picPage Fault Handling. jpg

Figure 1. 11: Page Fault Handling (Retrieved from Slides Material: Chapter 17 slides 6)

The diagram above illustrated the steps in handling page fault. When page fault occurs during program execution, the kernel will first locate the missing page on the backing store (disk). After located the page, it will find a free memory frame in the physical memory and copy into it. The page table will be reset after that and the instruction will be restart.

Problem- No Free Frames

When all the frames in the memory is been used, the other problem will occurs. This will cause the pages is unable to paged into the memory.

Solution- Page Replacement Algorithms

Solution for no free frames problem is to find a memory frame that is idle and free the frame using a page replacement algorithm. There are three common types of page replacement algorithm such as First in First out (FIFO), Optimal and Least Recently Used (LRU).

UNIX is using least recently used algorithm for page replacement. The least recently used algorithm replaces the page that has not been used for the longest time, on the assumption that the page will not be needed again. The page table will record every time the page being referenced, and when page

replacement is needed, every page will be checked to find the oldest recorded time.

Conclusion

Every operating system has different memory management. UNIX also has their exclusive memory management strategies to manage the memory resource optimally. UNIX is using multiple and variable partitioning so that the memory can be stored and use more flexible.

UNIX uses overlays and swapping to replace the unused program. However, it is facing external fragmentation problem and solve by loading the program into memory by using best fit algorithm.

Besides, UNIX also fully utilized the virtual memory (physical memory and swap space) by using demand paging. It allows user to store physical memory in the hard disk because the RAM memory was always insufficient.

Frequently Ask Question (FAQ)

What is a Kernel Memory?

Kernel Memory

UNIX owns a (semi-)private memory space called Kernel memory. Kernel uses RAM to keep itself memory resident to ensure that user programs do not overwrite or corrupt the kernel /user's data structures. Strong memory protection is implemented in kernel memory management to keep users from corrupting the system area.

C: UsersEricDesktopkernel. jpg

Figure 1. 12: Kernel memory

What is fork in UNIX?

Fork is a function used to duplicate a process. The newly created process is called “ Child” and the original process called “ Parent”.

What is the difference between Swapping and Paging?

Swapping

Paging

- Bringing in each process in physical memory entirely and running it.
- Only the needed pages to execute the program will be swapped in.
- Not using Virtual Memory.
- Use advantage on Virtual Memory.
- Process size must be equal or less than main memory.
- Process size can be greater than main memory.
- Less flexible in handling memory.
- More flexible in handling memory.

Limitations

Appendices

Layered Architecture of the UNIX system(Extracted from Fundamentals of Computer Technology, 2007 by LEON. A. and LEON. M.)

Text Segment

Initialized Part of Data Segment

Uninitialized Part of Data Segment

Heap Storage

Stack Segment

Environment Variables, etc

Low Address

High Address

The UNIX-style Memory layout(Extracted from Operations Systems by Gary Nutt 1997, Chapter 11: Memory Management)