IT 540 Operating Systems ECE519 Advanced Operating Systems

Prof. Dr. Hasan Hüseyin BALIK

(8th Week)

(Advanced) Operating Systems

8. Virtual Memory

8. Outline

- Hardware and Control Structures
- Operating System Software

Virtual memory	A storage allocation scheme in which secondary memory can be addressed as though it were part of main memory. The addresses a program may use to reference memory are distinguished from the addresses the memory system uses to identify physical storage sites, and program-generated addresses are translated automatically to the corresponding machine addresses.The size of virtual storage is limited by the addressing scheme of the computer system and by the amount of secondary memory available and not by the actual number of main storage locations.
Virtual address	The address assigned to a location in virtual memory to allow that location to be accessed as though it were part of main memory.
Virtual address space	The virtual storage assigned to a process.
Address space	The range of memory addresses available to a process.
Real address	The address of a storage location in main memory.

Virtual Memory Terminology

Hardware and Control Structures

- Two characteristics fundamental to memory management:
    1) all memory references are logical addresses that are dynamically translated into physical addresses at run time
    2) a process may be broken up into a number of pieces that don't need to be contiguously located in main memory during execution
- If these two characteristics are present, it is not necessary that all of the pages or segments of a process be in main memory during execution

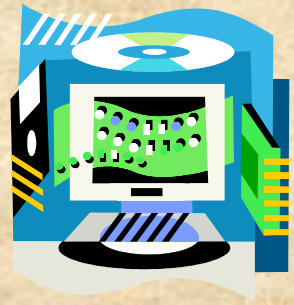
Execution of a Process

- Operating system brings into main memory a few pieces of the program
- Resident set
    - portion of process that is in main memory
- An interrupt is generated when an address is needed that is not in main memory
- Operating system places the process in a blocking state

Continued . . .

Execution of a Process

- Piece of process that contains the logical address is brought into main memory
  - operating system issues a disk I/O Read request
  - another process is dispatched to run while the disk I/O takes place
  - an interrupt is issued when disk I/O is complete, which causes the operating system to place the affected process in the Ready state

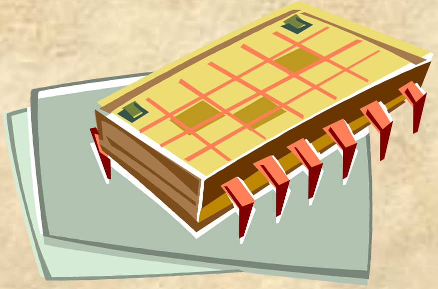


Implications

- More processes may be maintained in main memory
  - only load in some of the pieces of each process
  - with so many processes in main memory, it is very likely a process will be in the Ready state at any particular time
- A process may be larger than all of main memory

Real and Virtual Memory

- **Real memory**
  - main memory, the actual RAM
- **Virtual memory**
  - memory on disk
  - allows for effective multiprogramming and relieves the user of tight constraints of main memory



Thrashing

A state in which the system spends most of its time swapping process pieces rather than executing instructions

To avoid this, the operating system tries to guess, based on recent history, which pieces are least likely to be used in the near future

Principle of Locality

- Program and data references within a process tend to cluster
- Only a few pieces of a process will be needed over a short period of time
- Therefore it is possible to make intelligent guesses about which pieces will be needed in the future
- Avoids thrashing

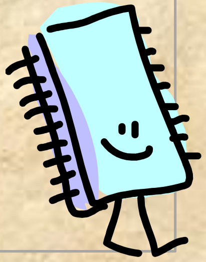
Support Needed for Virtual Memory

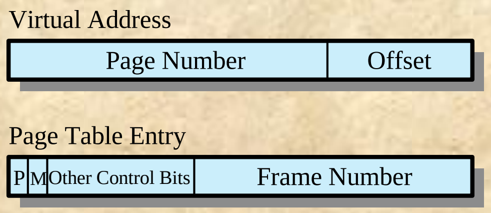
For virtual memory to be practical and effective:

- hardware must support paging and segmentation
- operating system must include software for managing the movement of pages and/or segments between secondary memory and main memory

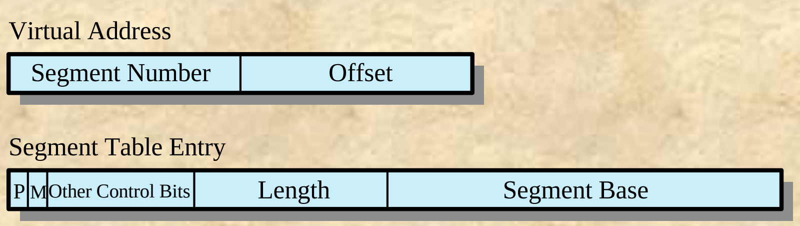
Paging

- The term *virtual memory* is usually associated with systems that employ paging
- Use of paging to achieve virtual memory was first reported for the Atlas computer (1992, one of the world's first supercomputers, joint development between the University of Manchester, Ferranti, and Plessey)
- Each process has its own page table
  - each page table entry contains the frame number of the corresponding page in main memory

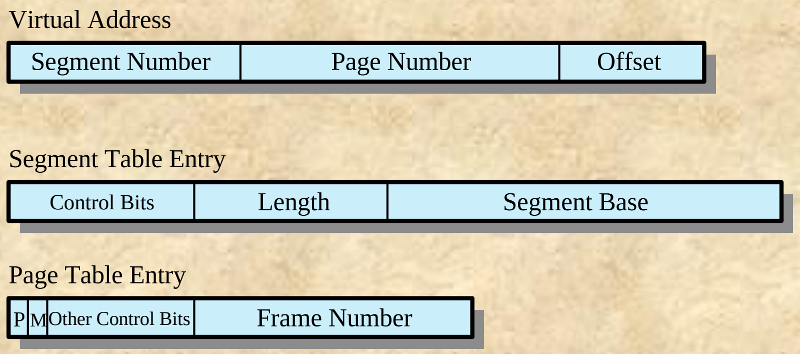




(a) Paging only



(b) Segmentation only



P= present bit
M = Modified bit

(c) Combined segmentation and paging

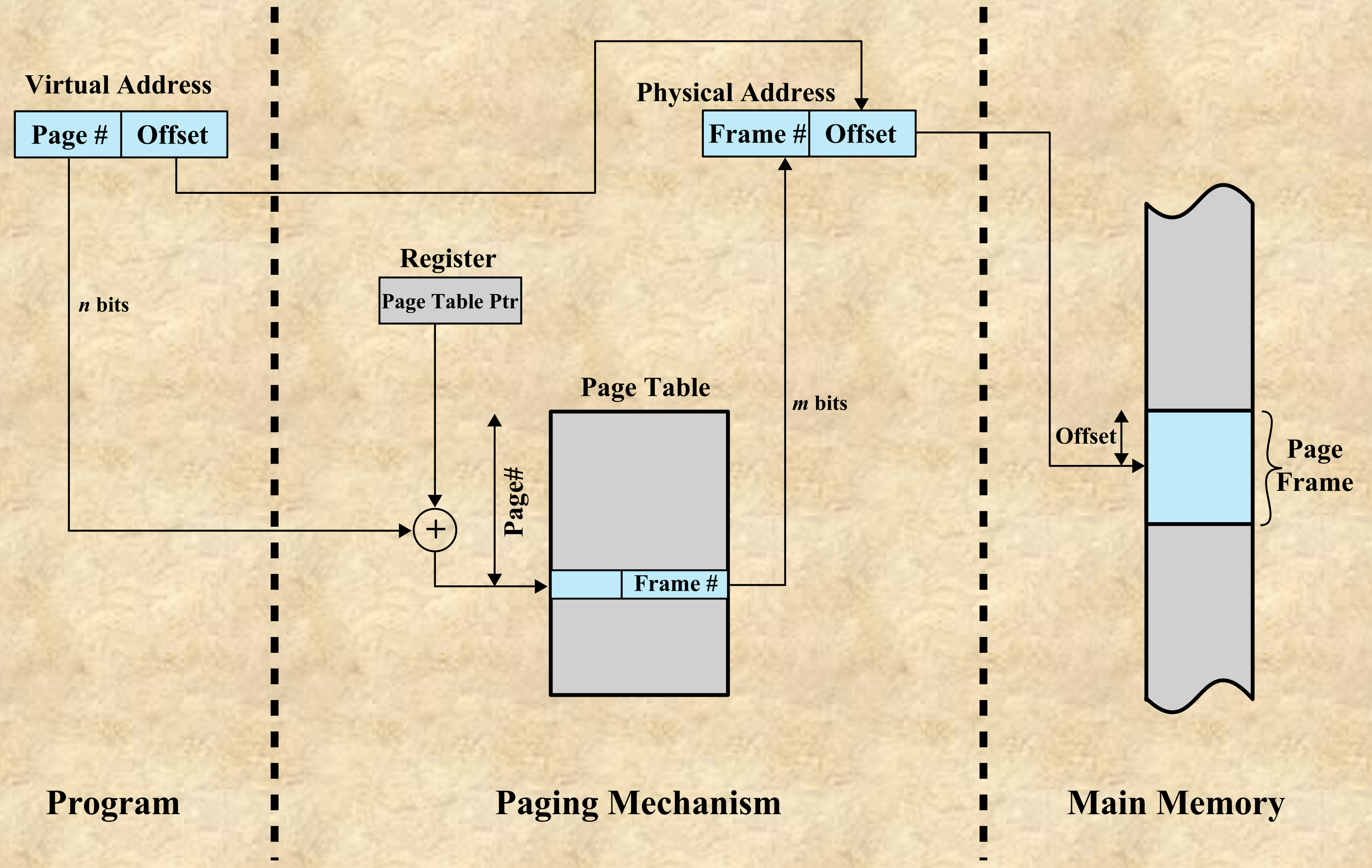


Figure 8.2 Address Translation in a Paging System

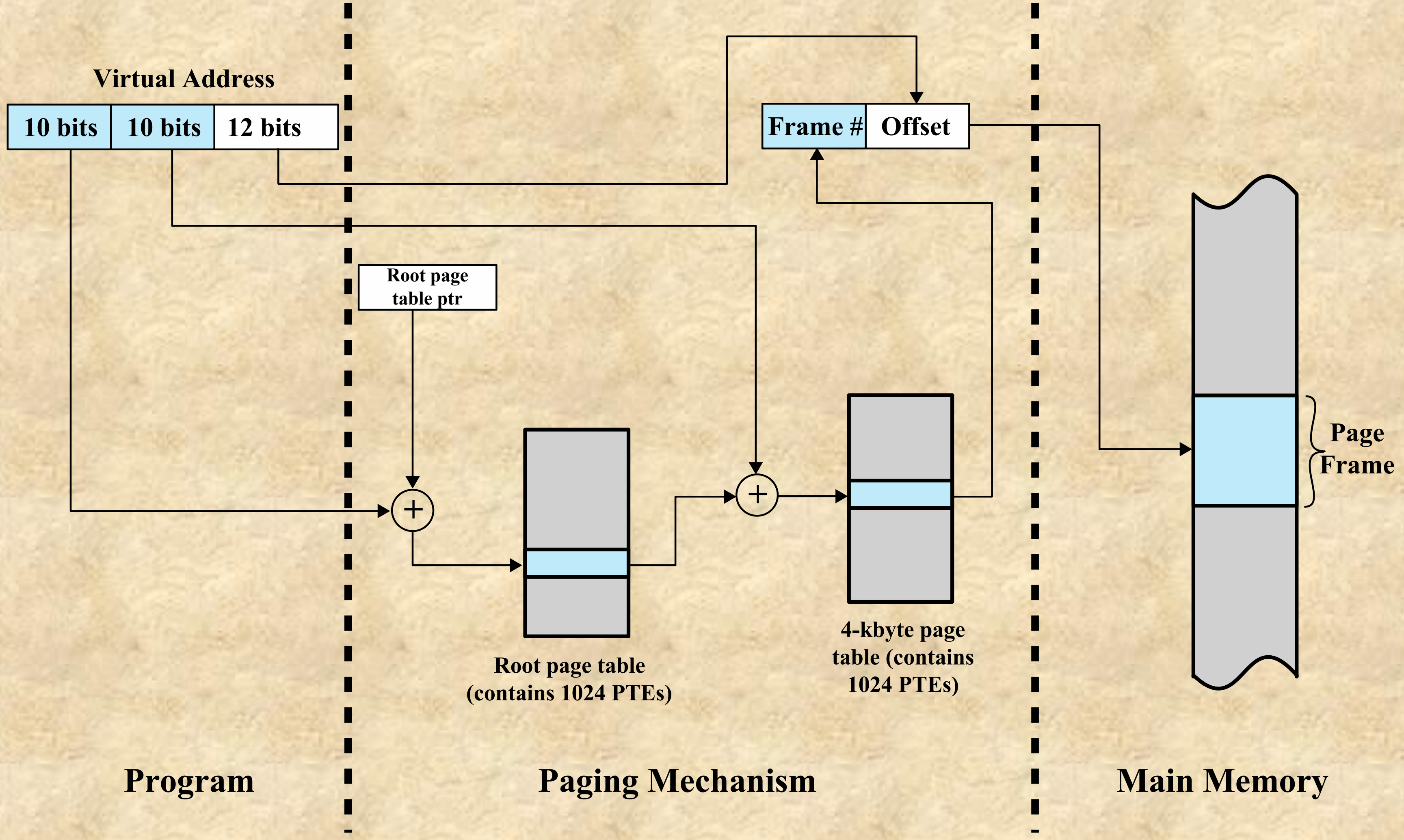


Figure 8.4 Address Translation in a Two-Level Paging System

Inverted Page Table

- Page number portion of a virtual address is mapped into a hash value
  - hash value points to inverted page table
- Fixed proportion of real memory is required for the tables regardless of the number of processes or virtual pages supported
- Structure is called *inverted* because it indexes page table entries by frame number rather than by virtual page number
- This approach are used on the PowerPC, UltraSPARC, and the IA-64 architecture

Inverted Page Table

Each entry in the page table includes:

Page number

Process identifier

- the process that owns this page

Control bits

- includes flags and protection and locking information

Chain pointer

- the index value of the next entry in the chain

Translation Lookaside Buffer (TLB)

- Each virtual memory reference can cause two physical memory accesses:
    - one to fetch the page table entry
    - one to fetch the data
- To overcome the effect of doubling the memory access time, most virtual memory schemes make use of a special high-speed cache called a ***translation lookaside buffer***

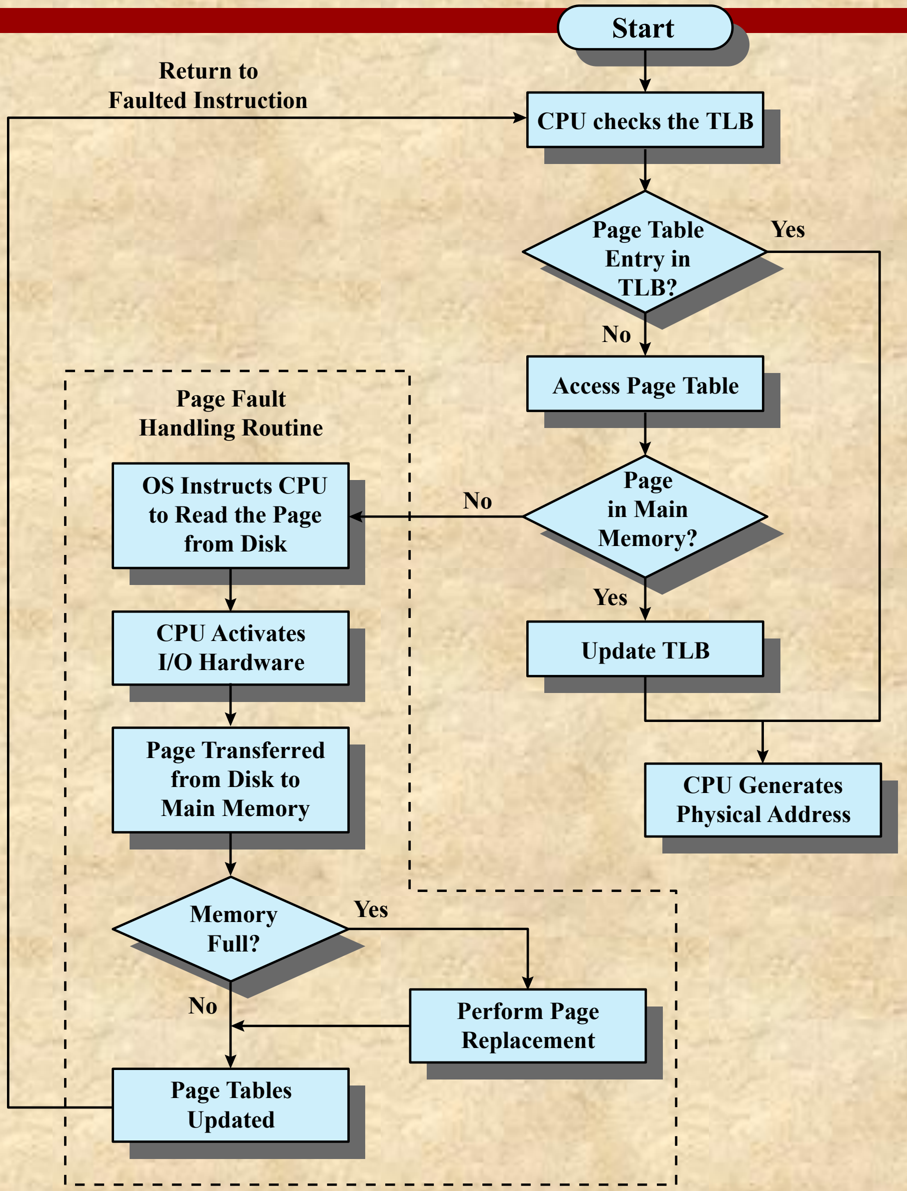

Start
CPU checks the TLB
Return to Faulted Instruction
Page Table Entry in TLB?
Yes
No
Access Page Table
Page in Main Memory?
No
Yes
Update TLB
CPU Generates Physical Address
Page Fault Handling Routine
OS Instructs CPU to Read the Page from Disk
CPU Activates I/O Hardware
Page Transferred from Disk to Main Memory
Memory Full?
Yes
No
Perform Page Replacement
Page Tables Updated

Figure 8.7 Operation of Paging and Translation Lookaside Buffer (TLB) [FURH87]

Page Size

- The smaller the page size, the lesser the amount of internal fragmentation
    - however, more pages are required per process
    - more pages per process means larger page tables
    - for large programs in a heavily multiprogrammed environment some portion of the page tables of active processes must be in virtual memory instead of main memory
    - the physical characteristics of most secondary-memory devices favor a larger page size for more efficient block transfer of data

Example Page Sizes

Computer	Page Size
Atlas	512 48-bit words
Honeywell-Multics	1024 36-bit words
IBM 370/XA and 370/ESA	4 Kbytes
VAX family	512 bytes
IBM AS/400	512 bytes
DEC Alpha	8 Kbytes
MIPS	4 Kbytes to 16 Mbytes
UltraSPARC	8 Kbytes to 4 Mbytes
Pentium	4 Kbytes or 4 Mbytes
IBM POWER	4 Kbytes
Itanium	4 Kbytes to 256 Mbytes

Page Size

the design issue of page size is related to the size of physical main memory and program size

→ main memory is getting larger and address space used by applications is also growing

↓ most obvious on personal computers where applications are becoming increasingly complex

- Contemporary programming techniques used in large programs tend to decrease the locality of references within a process

Segmentation

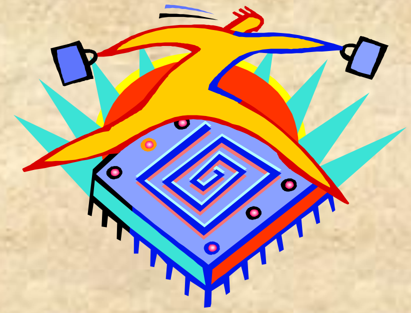
- Segmentation allows the programmer to view memory as consisting of multiple address spaces or segments

Advantages:

- simplifies handling of growing data structures
- allows programs to be altered and recompiled independently
- lends itself to sharing data among processes
- lends itself to protection

Segment Organization

- Each segment table entry contains the starting address of the corresponding segment in main memory and the length of the segment
- A bit is needed to determine if the segment is already in main memory
- Another bit is needed to determine if the segment has been modified since it was loaded in main memory



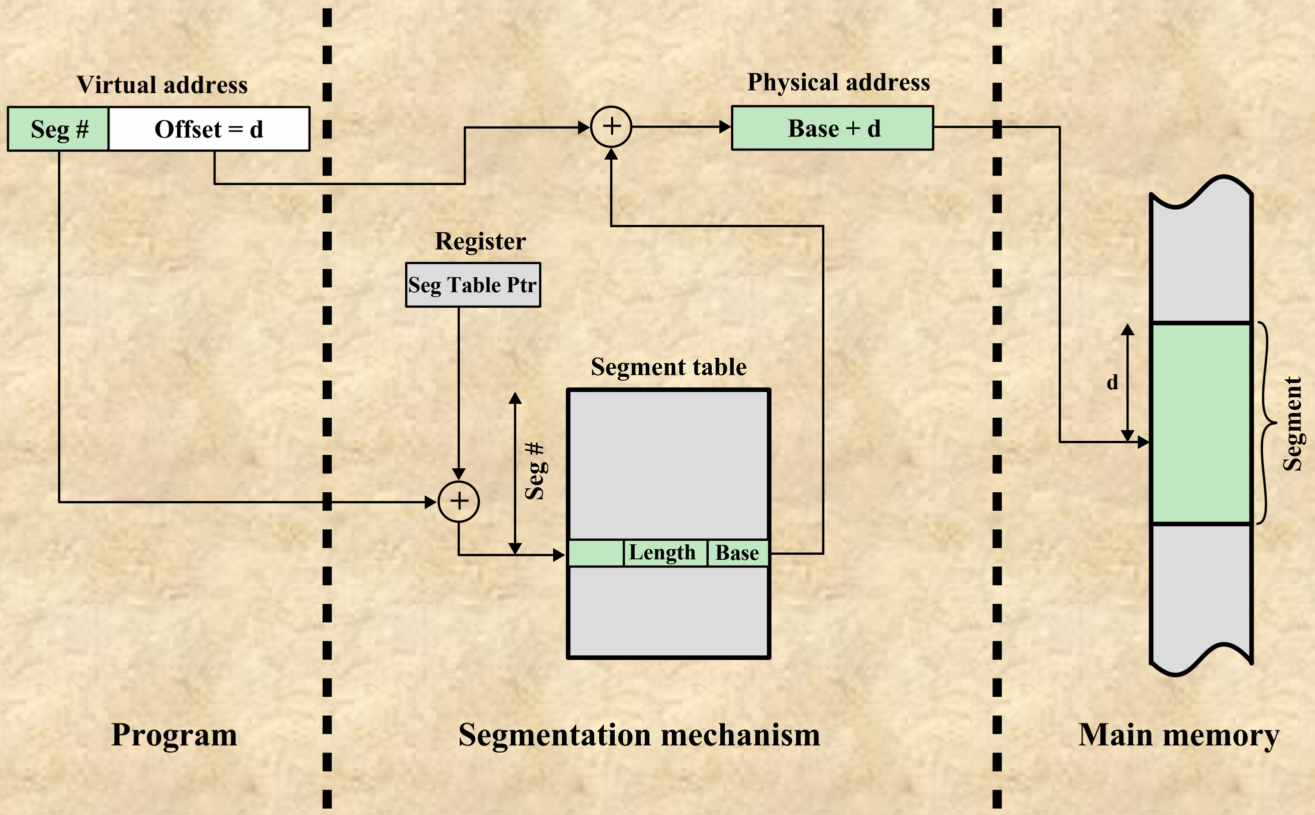


Figure 8.11 Address Translation in a Segmentation System

Combined Paging and Segmentation

In a combined paging/segmentation system a user's address space is broken up into a number of segments. Each segment is broken up into a number of fixed-sized pages which are equal in length to a main memory frame

Segmentation is visible to the programmer

Paging is transparent to the programmer

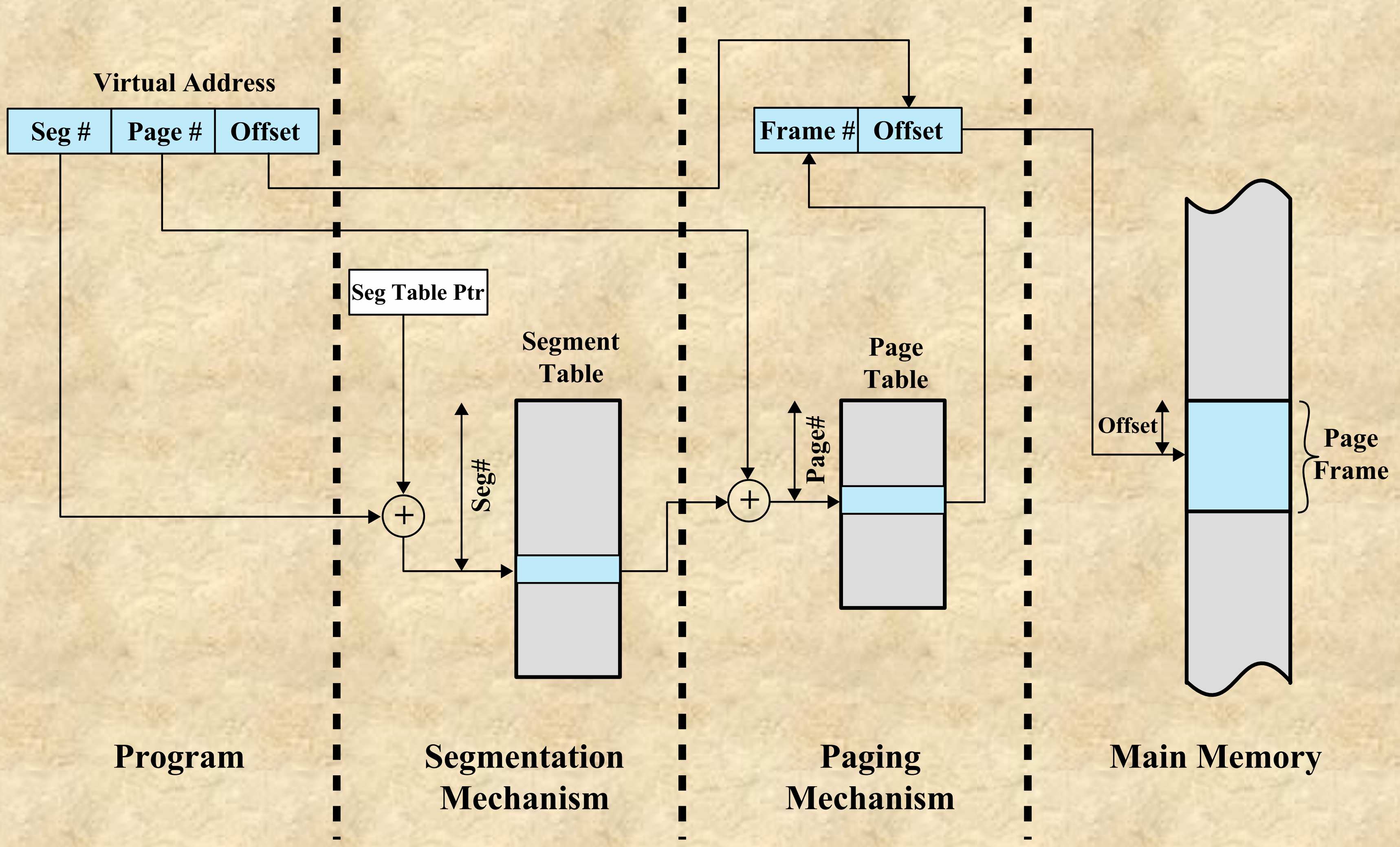
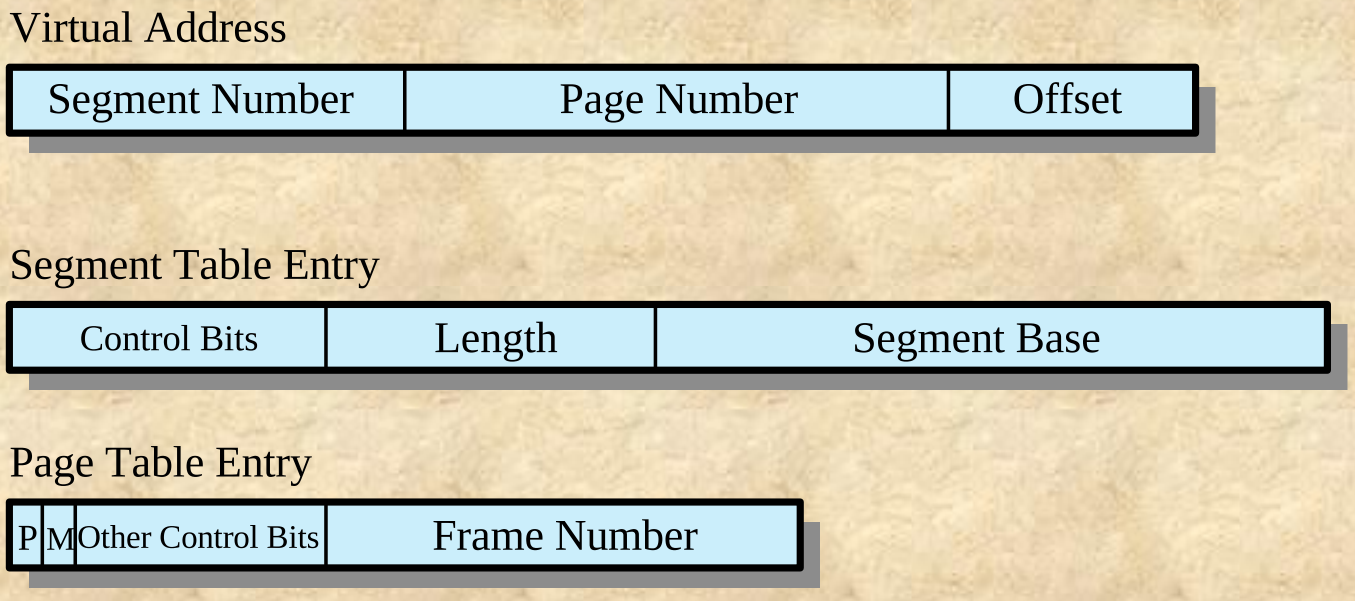


Figure 8.12 Address Translation in a Segmentation/Paging System

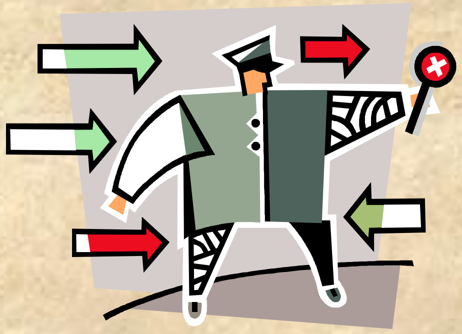


P= present bit
M = Modified bit

(c) Combined segmentation and paging

Protection and Sharing

- Segmentation lends itself to the implementation of protection and sharing policies
- Each entry has a base address and length so inadvertent memory access can be controlled
- Sharing can be achieved by segments referencing multiple processes



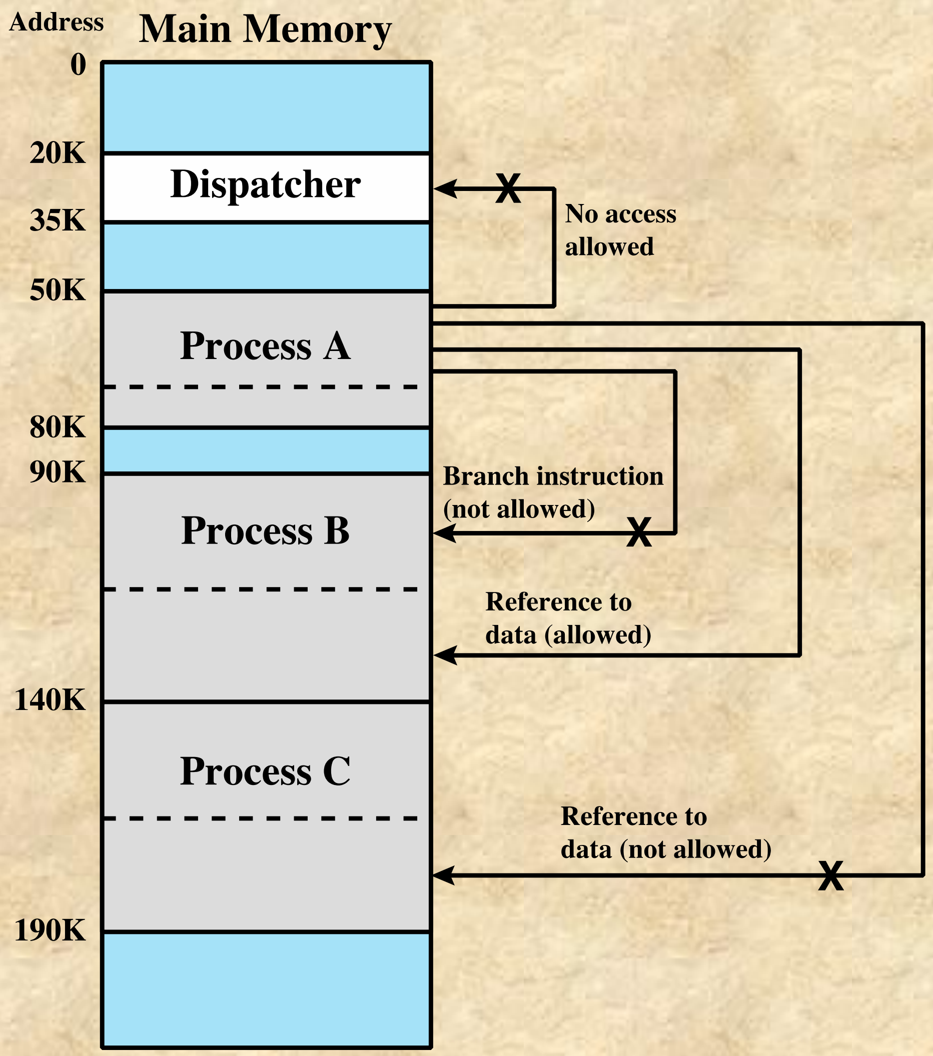


Figure 8.13 Protection Relationships Between Segments

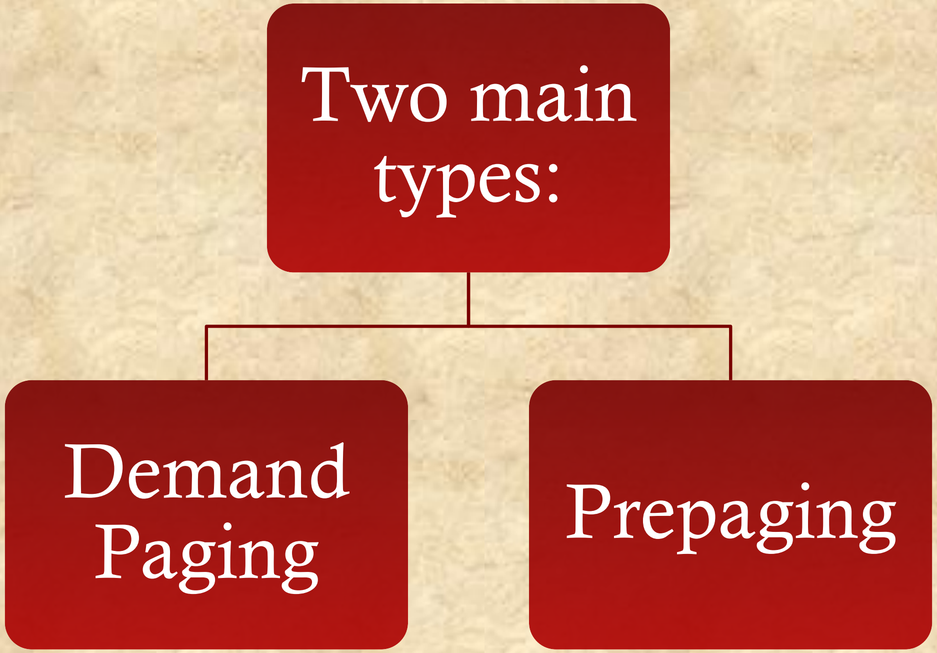
Operating System Software

The design of the memory management portion of an operating system depends on three fundamental areas of choice:

- whether or not to use virtual memory techniques
- the use of paging or segmentation or both
- the algorithms employed for various aspects of memory management

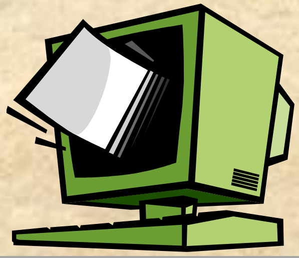
Fetch Policy

- Determines when a page should be brought into memory



Demand Paging

- **Demand Paging**
  - only brings pages into main memory when a reference is made to a location on the page
  - many page faults when process is first started
  - principle of locality suggests that as more and more pages are brought in, most future references will be to pages that have recently been brought in, and page faults should drop to a very low level



Prepaging

- **Prepaging**
  - pages other than the one demanded by a page fault are brought in
  - exploits the characteristics of most secondary memory devices
  - if pages of a process are stored contiguously in secondary memory it is more efficient to bring in a number of pages at one time
  - ineffective if extra pages are not referenced
  - should not be confused with "swapping"

Placement Policy

- Determines where in real memory a process piece is to reside
- Important design issue in a segmentation system
- Paging or combined paging with segmentation placing is irrelevant because hardware performs functions with equal efficiency
- For NUMA systems an automatic placement strategy is desirable

Replacement Policy

- Deals with the selection of a page in main memory to be replaced when a new page must be brought in
    - objective is that the page that is removed be the page least likely to be referenced in the near future
- The more elaborate the replacement policy the greater the hardware and software overhead to implement it

Frame Locking

- When a frame is locked the page currently stored in that frame may not be replaced
    - kernel of the OS as well as key control structures are held in locked frames
    - I/O buffers and time-critical areas may be locked into main memory frames
    - locking is achieved by associating a lock bit with each frame

Basic Algorithms

Algorithms used for the selection of a page to replace:

- Optimal
- Least recently used (LRU)
- First-in-first-out (FIFO)
- Clock

Least Recently Used (LRU)

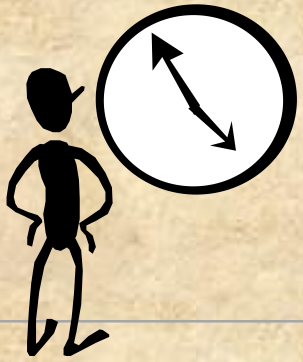
- Replaces the page that has not been referenced for the longest time
- By the principle of locality, this should be the page least likely to be referenced in the near future
- Difficult to implement
  - one approach is to tag each page with the time of last reference
    - this requires a great deal of overhead

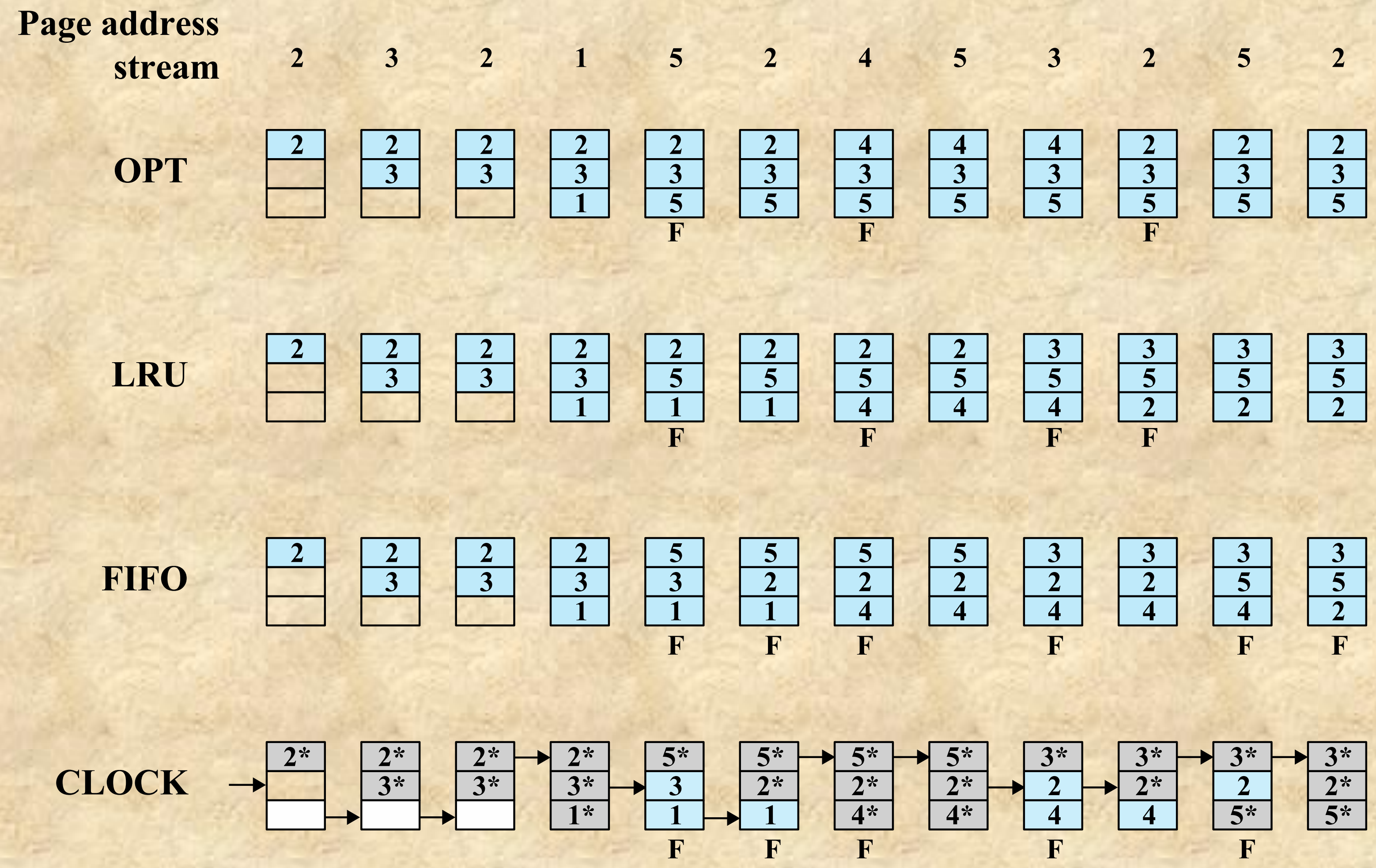
First-in-First-out (FIFO)

- Treats page frames allocated to a process as a circular buffer
- Pages are removed in round-robin style
  - simple replacement policy to implement
- Page that has been in memory the longest is replaced

Clock Policy

- Requires the association of an additional bit with each frame
  - referred to as the *use* bit
- When a page is first loaded in memory or referenced, the use bit is set to 1
- The set of frames is considered to be a circular buffer
- Any frame with a use bit of 1 is passed over by the algorithm
- Page frames visualized as laid out in a circle



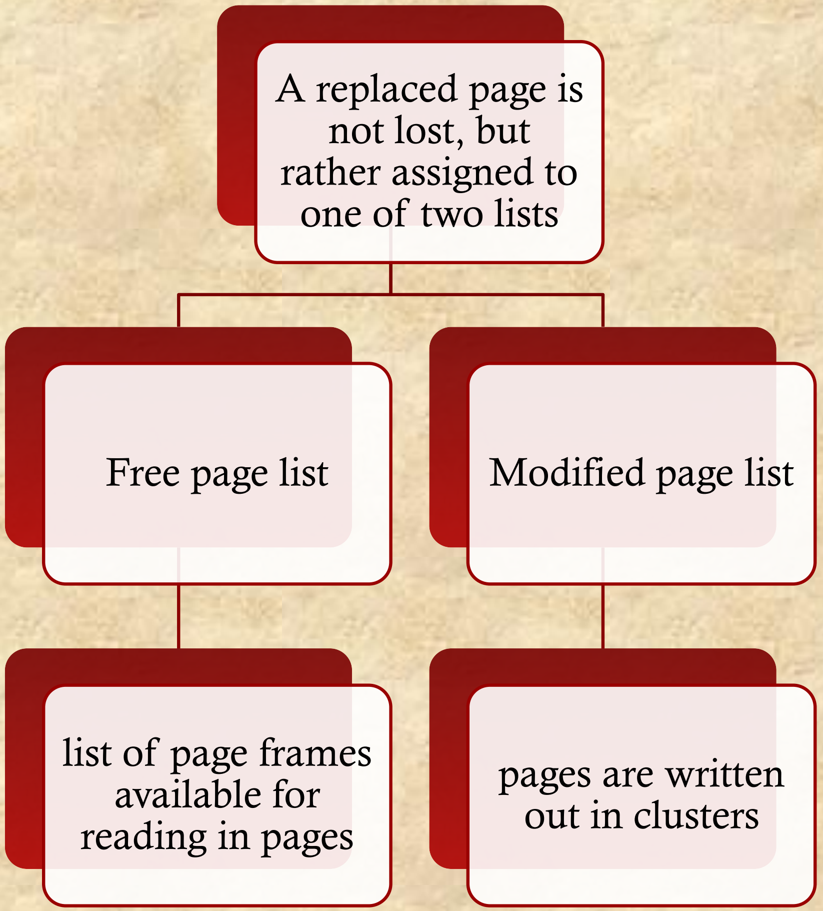


F = page fault occurring after the frame allocation is initially filled

Figure 8.14 Behavior of Four Page-Replacement Algorithms

Page Buffering

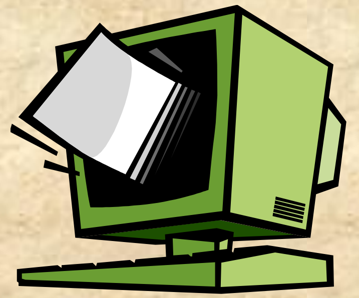
- Improves paging performance and allows the use of a simpler page replacement policy



A replaced page is not lost, but rather assigned to one of two lists
Free page list
Modified page list
list of page frames available for reading in pages
pages are written out in clusters

Replacement Policy and Cache Size

- With large caches, replacement of pages can have a performance impact
  - if the page frame selected for replacement is in the cache, that cache block is lost as well as the page that it holds
  - in systems using page buffering, cache performance can be improved with a policy for page placement in the page buffer
  - most operating systems place pages by selecting an arbitrary page frame from the page buffer



Resident Set Management

- The OS must decide how many pages to bring into main memory
  - the smaller the amount of memory allocated to each process, the more processes can reside in memory
  - small number of pages loaded increases page faults
  - beyond a certain size, further allocations of pages will not effect the page fault rate

Resident Set Size

Fixed-allocation

- gives a process a fixed number of frames in main memory within which to execute
  - when a page fault occurs, one of the pages of that process must be replaced

Variable-allocation

- allows the number of page frames allocated to a process to be varied over the lifetime of the process

Replacement Scope

- The scope of a replacement strategy can be categorized as ***global*** or ***local***
  - both types are activated by a page fault when there are no free page frames

Local

- chooses only among the resident pages of the process that generated the page fault

Global

- considers all unlocked pages in main memory

Fixed Allocation, Local Scope

- Necessary to decide ahead of time the amount of allocation to give a process
- If allocation is too small, there will be a high page fault rate

If allocation is too large, there will be too few programs in main memory

- increased processor idle time
- increased time spent in swapping

Variable Allocation Global Scope

- Easiest to implement
  - adopted in a number of operating systems
- OS maintains a list of free frames
- Free frame is added to resident set of process when a page fault occurs
- If no frames are available the OS must choose a page currently in memory
- One way to counter potential problems is to use page buffering

Variable Allocation Local Scope

- When a new process is loaded into main memory, allocate to it a certain number of page frames as its resident set
- When a page fault occurs, select the page to replace from among the resident set of the process that suffers the fault
- Reevaluate the allocation provided to the process and increase or decrease it to improve overall performance

Variable Allocation Local Scope

- Decision to increase or decrease a resident set size is based on the assessment of the likely future demands of active processes

Key elements:

- criteria used to determine resident set size
- the timing of changes

Page Fault Frequency (PFF)

- Requires a use bit to be associated with each page in memory
- Bit is set to 1 when that page is accessed
- When a page fault occurs, the OS notes the virtual time since the last page fault for that process
- Does not perform well during the transient periods when there is a shift to a new locality

Variable-Interval Sampled Working Set (VSWS)

- Evaluates the working set of a process at sampling instances based on elapsed virtual time
- Driven by three parameters:

the minimum duration of the sampling interval

the maximum duration of the sampling interval

the number of page faults that are allowed to occur between sampling instances

Cleaning Policy

- Concerned with determining when a modified page should be written out to secondary memory

Demand Cleaning

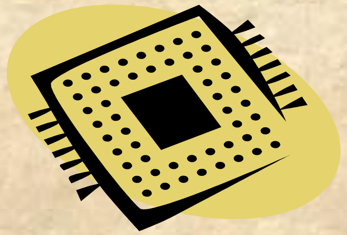
a page is written out to secondary memory only when it has been selected for replacement

Precleaning

allows the writing of pages in batches

Load Control

- Determines the number of processes that will be resident in main memory
  - *multiprogramming* level
- Critical in effective memory management
- Too few processes, many occasions when all processes will be blocked and much time will be spent in swapping
- Too many processes will lead to thrashing



Process Suspension

- If the degree of multiprogramming is to be reduced, one or more of the currently resident processes must be swapped out

Six possibilities exist:

- lowest-priority process
- faulting process
- last process activated
- process with the smallest resident set
- largest process
- process with the largest remaining execution window