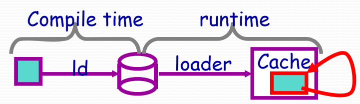
Memory Management

Ken Birman

How to create a process?

- On Unix systems, executable read by loader

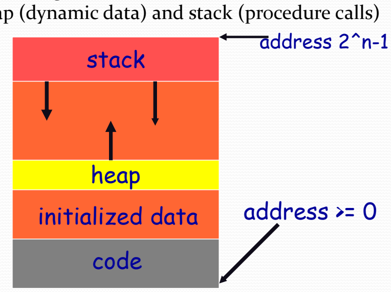


- Compiler: generates one object file per source file
- Linker: combines all object files into one executable
- Loader: loads executable in memory

2

What does process look like?

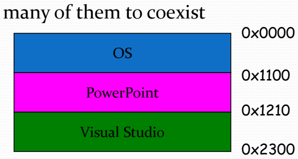
Process divided into segments, each has:
- Code, data, heap (dynamic data) and stack (procedure calls)



3

Processes in Memory

- We want many of them to coexist

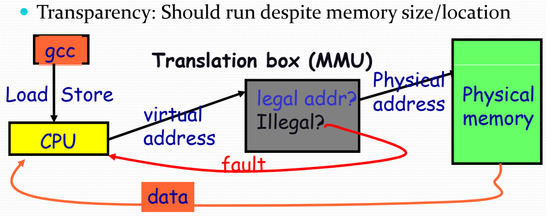


- Issues:
  - PowerPoint needs more memory than there is on machine?
  - What is visual studio tries to access 0x0005?
  - If PowerPoint is not using enough memory?
  - If OS needs to expand?

4

Issues

- Protection: Errors in process should not affect others
- Transparency: Should run despite memory size/location

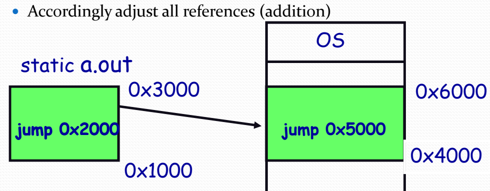


How to do this mapping?

5

Scheme 1: Load-time Linking

- Link as usual, but keep list of references
- At load time: determine the new base address
  - Accordingly adjust all references (addition)

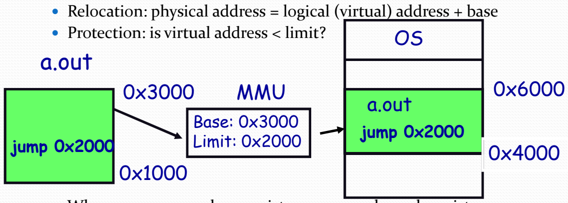


- Issues: handling multiple segments, moving in memory

6

Scheme 2: Execution-time Linking

- Use hardware (base + limit reg) to solve the problem
  - Done for every memory access
  - Relocation: physical address = logical (virtual) address + base
  - Protection: is virtual address < limit?

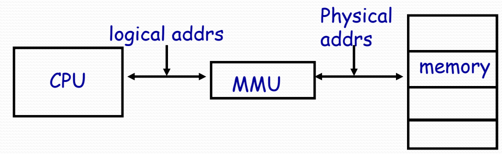


  - When process runs, base register = 0x3000, bounds register = 0x2000. Jump addr = 0x2000 + 0x3000 = 0x5000

7

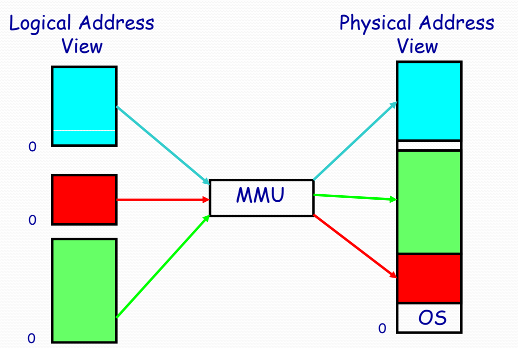
Dynamic Translation

- Memory Management Unit in hardware
  - Every process has its own address space



8

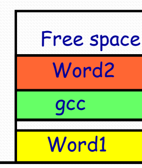
Logical and Physical Addresses



9

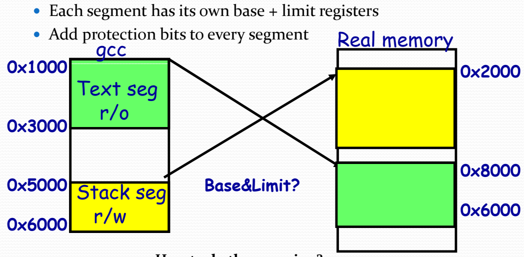
Scheme 2: Discussion

- Pro:
  - cheap in terms of hardware: only two registers
  - cheap in terms of cycles: do add and compare in parallel
- Con: only one segment
  - prob 1: growing processes. How to expand gcc?
  - prob 2: how to share code and data?? how can Word copies share code?
  - prob 3: how to separate code and data?
- A solution: multiple segments
  - **"segmentation"**



Segmentation

- Processes have multiple base + limit registers
- Processes address space has multiple segments
  - Each segment has its own base + limit registers
  - Add protection bits to every segment



How to do the mapping?

11

Types of Segments

- Each "file" of a program (or library) typically results in
  - An associated code segment (instructions)
  - The "initialized" data segment for that code
  - The zeroed part of the data segment ("bss")
- Programs also have
  - One heap segment – grows "upward"
  - A stack segment for each active thread (grows down; all but first have a size limit)

12

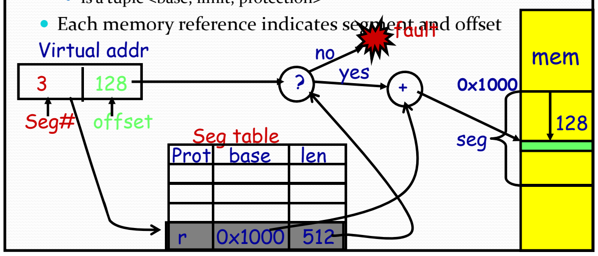
Lots of Segments

- Bottom line: a single process might have a *whole lot* of active segments!
  - And it can add them while it runs, for example by linking to a DLL or forking a thread
  - Address space is a big range with chunks of data separated by gaps
- It isn't uncommon for a process on a modern platform to have hundreds of segments in memory...

13

Mapping Segments

- Segment Table
  - An entry for each segment
  - Is a tuple <base, limit, protection>
- Each memory reference indicates segment and offset

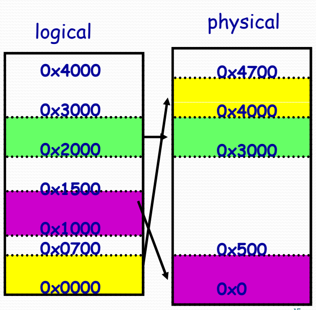


Segmentation Example

- If first two bits are for segments, next 12 for offset

Seg	base	bounds	rw
0	0x4000	0x6ff	10
1	0x0000	0x4ff	11
2	0x3000	0xfff	11
3			00

- where is 0x0240?
- 0x1108?
- 0x265c?
- 0x3002?
- 0x1700?



15

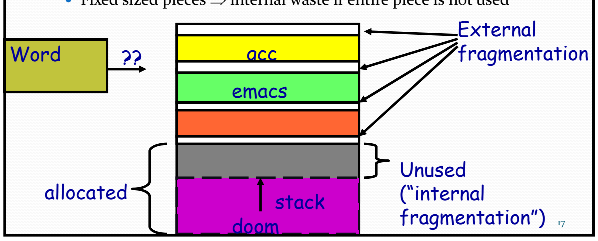
Segmentation: Discussion

- Advantages:
  - Allows multiple segments per process
  - Easy to allow sharing of code
  - Do not need to load entire process in memory
- Disadvantages:
  - Extra translation overhead:
    - Memory & speed
  - An entire segment needs to reside contiguously in memory!
    ⇒ Fragmentation

16

Fragmentation

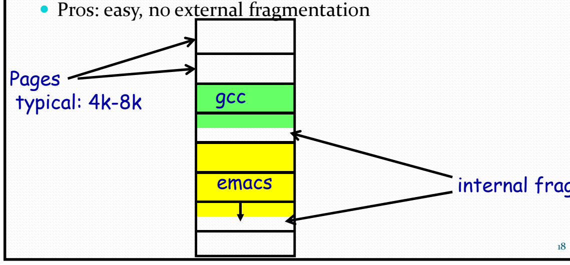
- "Free memory is scattered into tiny, useless pieces"
- External Fragmentation:
  - Variable sized pieces ⇒ many small holes over time
- Internal Fragmentation:
  - Fixed sized pieces ⇒ internal waste if entire piece is not used



17

Paging

- Divide memory into fixed size pieces
  - Called "frames" or "pages"
- Pros: easy, no external fragmentation

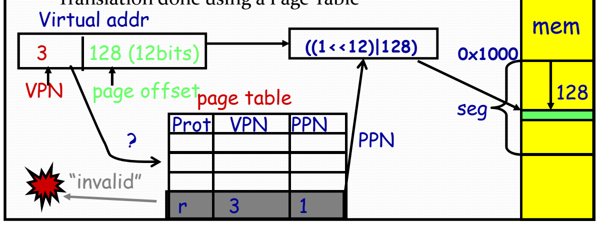


18

Mapping Pages

- If 2^m virtual address space, 2^n page size
  - ⇒ (m - n) bits to denote page number, n for offset within page

Translation done using a Page Table



Paging: Hardware Support

- Entire page table (PT) in registers
  - PT can be huge ~ 1 million entries
- Store PT in main memory
  - Have PTBR point to start of PT
  - Con: 2 memory accesses to get to any physical address
  - Solution: a special cache (next slide)
- Each time we context switch, must switch the current PT too, so that the process we execute will be able to see its memory region

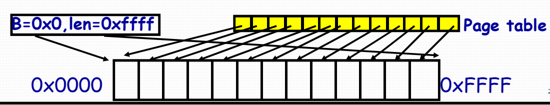
20

MMU includes a cache

- We call it the *translation lookaside buffer (TLB)*
  - Just a cache, with a special name
  - This cache contains page table entries
- The OS needs to "clear" the TLB on context switch
  - So first few operations by a freshly started process are very slow!
  - But then the cache will warm up
  - Goal is that MMU can find the PTE entries it needs without actually fetching them from the page table

Paging: Discussion

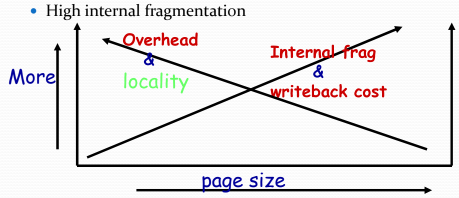
- Advantages:
  - No external fragmentation
  - Easy to allocate
  - Easy to swap, since page size usually same as disk block size
- Disadvantages:
  - Space and speed
    - One PT entry for every page, vs. one entry for contiguous memory for segmentation



22

Size of the page

- Small page size:
  - High overhead:
    - What is size of PT if page size is 512 bytes, and 32-bit addr space?
- Large page size:
  - High internal fragmentation



23

Paging + Segmentation

- Paged segmentation
  - Handles very long segments
  - The segments are paged
- Segmented Paging
  - When the page table is very big
  - Segment the page table
  - Let's consider System 370 (24-bit address space)

Seg # (4 bits)	page # (8 bits)	page offset (12 bits)

24

Segmented Paging Example

Table entry addresses

	Base	bound	prot
0	0x2000	0x14	R
1	0x0000	0x00	
2	0x1000	0x0d	RW

Segment table

Page table (2-byte entries)

Page table	Table entry addresses
0x001f	
0x0011	0x2020
…	
0x0003	
0x002a	
0x0013	0x2000
…	
0x000c	
0x0007	0x1020
…	
0x004	
0x00b	
0x006	0x1000

0x202016 read?
0x104c84 read?
0x011424 read?
0x210014 write?

Seg 2: page 2, PTE @ 0x1004, addr = 0x4016
Seg 1: page 4, PTE @ protection error
Seg 0, page 0x11, PTE@0x2022, addr = 0x1f424
Seg 2, page 0x10, PTE@ bounds violation

25

Handling PT size

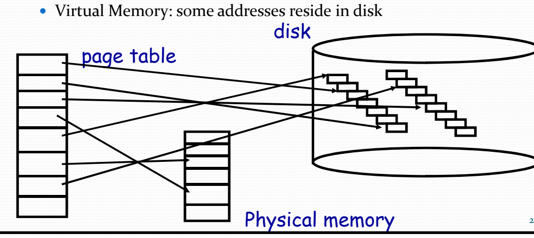
- Segmented Paging removes need for contiguous alloc
- OS maintains one copy of PT for every process
  - And a global copy of all frames
- Other approaches:
  - Hierarchical Paging: Page the page table
  - Hashed Page Table: Each entry maps to linked list of pages
  - Inverted Page Table:
    - Map from Frame to (VA, process) instead of VA to frame per process

26

Virtual Memory

What is virtual memory?

- Each process has illusion of large address space
  - 2^{32} for 32-bit addressing
- However, physical memory is much smaller
- How do we give this illusion to multiple processes?
  - Virtual Memory: some addresses reside in disk



28

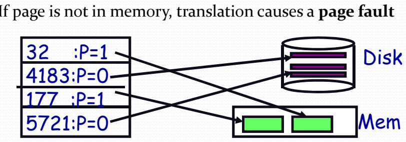
Virtual Memory

- Load entire process in memory (swapping), run it, exit
  - Is slow (for big processes)
  - Wasteful (might not require everything)
- Solutions: partial residency
  - Paging: only bring in pages, not all pages of process
  - Demand paging: bring only pages that are required
- Where to fetch page from?
  - Have a contiguous space in disk: swap file (pagefile.sys)

29

How does VM work?

- Modify Page Tables with another bit ("is present")
  - If page in memory, *is_present* = 1, else *is_present* = 0
  - If page is in memory, translation works as before
  - If page is not in memory, translation causes a **page fault**



30

Page Faults

- On a page fault:
  - OS finds a free frame, or evicts one from memory (which one?)
    - Want knowledge of the future?
  - Issues disk request to fetch data for page (what to fetch?)
    - Just the requested page, or more?
  - Block current process, context switch to new process (how?)
    - Process might be executing an instruction
  - When disk completes, set present bit to 1, and current process in ready queue

31

Resuming after a page fault

- Should be able to restart the instruction
- For RISC processors this is simple:
  - Instructions are idempotent until references are done
- More complicated for CISC:
  - E.g. move 256 bytes from one location to another
  - Possible Solutions:
    - Ensure pages are in memory before the instruction executes

32

When to fetch?

- Just before the page is used!
  - Need to know the future
- Demand paging:
  - Fetch a page when it faults
- Prepaging:
  - Get the page on fault + some of its neighbors, or
  - Get all pages in use last time process was swapped

33

What to replace?

- Page Replacement
  - When process has used up all frames it is allowed to use
  - OS must select a page to eject from memory to allow new page
  - The page to eject is selected using the Page Replacement Algo
- Goal: Select page that minimizes future page faults

34

Page Replacement Algorithms

- Random: Pick any page to eject at random
  - Used mainly for comparison
- FIFO: The page brought in earliest is evicted
  - Ignores usage
  - Suffers from "Belady's Anomaly"
    - Fault rate could increase on increasing number of pages
    - E.g. 0 1 2 3 0 1 4 0 1 2 3 4 with frame sizes 3 and 4
- OPT: Belady's algorithm
  - Select page not used for longest time
- LRU: Evict page that hasn't been used the longest
  - Past could be a good predictor of the future

35

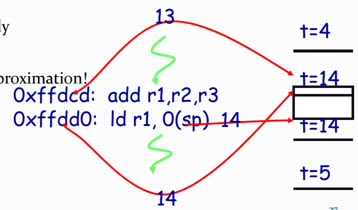
Example: FIFO, OPT

```
Reference stream is A B C A B D A D B C

OPTIMAL
        A  B  C  A  B  D  A  D  B  C  B
        5 Faults        toss C         toss A or D

A
B
C   FIFO
D       A  B  C  A  B  D  A  D  B  C  B
A
B       7 Faults        toss A       toss ?
C
```

36

Implementing Perfect LRU

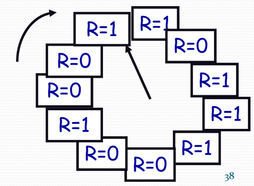
- On reference: Time stamp each page
- On eviction: Scan for oldest frame
- Problems:
  - Large page lists
  - Timestamps are costly
- Approximate LRU
  - LRU is already an approximation!



37

LRU: Clock Algorithm

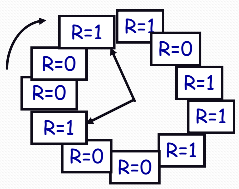
- Each page has a reference bit
  - Set on use, reset periodically by the OS
- Algorithm:
  - FIFO + reference bit (keep pages in circular list)
    - Scan: if ref bit is 1, set to 0, and proceed. If ref bit is 0, stop and evict.
- Problem:
  - Low accuracy for large memory



38

LRU with large memory

- Solution: Add another hand
  - Leading edge clears ref bits
  - Trailing edge evicts pages with ref bit 0
- What if angle small?
- What if angle big?



39

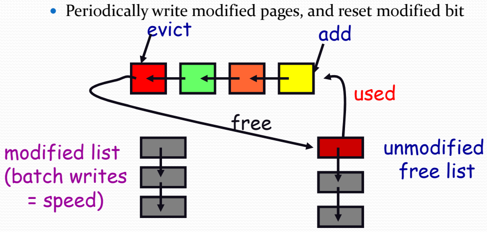
Clock Algorithm: Discussion

- Sensitive to sweeping interval
  - Fast: lose usage information
  - Slow: all pages look used
- Clock: add reference bits
  - Could use (ref bit, modified bit) as ordered pair
  - Might have to scan all pages
- LFU: Remove page with lowest count
  - No track of when the page was referenced
  - Use multiple bits. Shift right by 1 at regular intervals.
- MFU: remove the most frequently used page
- LFU and MFU do not approximate OPT well

40

Page Buffering

- Cute simple trick: (XP, 2K, Mach, VMS)
  - Keep a list of free pages
  - Track which page the free page corresponds to
  - Periodically write modified pages, and reset modified bit



41

Allocating Pages to Processes

- Global replacement
  - Single memory pool for entire system
  - On page fault, evict oldest page in the system
  - Problem: protection
- Local (per-process) replacement
  - Have a separate pool of pages for each process
  - Page fault in one process can only replace pages from its own process
  - Problem: might have idle resources

42