Virtual Memory I
CSE 351 Winter 2019

Instructors:
Max Willsey
Luis Ceze

Teaching Assistants:
Britt Henderson
Lukas Joswiak
Josie Lee
Wei Lin
Daniel Snitkovsky
Luis Vega
Kory Watson
Ivy Yu

https://xkcd.com/1495/
Roadmap

C:

```c
car *c = malloc(sizeof(car));
c->miles = 100;
c->gals = 17;
float mpg = get_mpg(c);
free(c);
```

Java:

```java
Car c = new Car();
c.setMiles(100);
c.setGals(17);
float mpg = c.getMPG();
```

Assembly language:
```
get_mpg:
pushq %rbp
movq %rsp, %rbp
...
popq %rbp
ret
```

Machine code:
```
0111010000011000
100011010000010000
1000100111000010
1100000111111010000
```

Computer system:

Memory & data
Integers & floats
x86 assembly
Procedures & stacks
Executables
Arrays & structs
Memory & caches
Processes
Virtual memory
Memory allocation
Java vs. C
Virtual Memory (VM*)

- Overview and motivation
- VM as a tool for caching
- Address translation
- VM as a tool for memory management
- VM as a tool for memory protection

**Warning:** Virtual memory is pretty complex, but crucial for understanding how processes work and for debugging performance.

*Not to be confused with “Virtual Machine” which is a whole other thing.*
Memory as we know it so far... is *virtual*!

- Programs refer to virtual memory addresses
  - `movq (%rdi), %rax`
  - Conceptually memory is just a very large array of bytes
  - System provides private address space to each process

- Allocation: Compiler and run-time system
  - Where different program objects should be stored
  - All allocation within single virtual address space

- But...
  - *We probably* don’t have $2^w$ bytes of physical memory
  - *We certainly* don’t have $2^w$ bytes of physical memory *for every process*
  - Processes should not interfere with one another
    - Except in certain cases where they want to share code or data

![Diagram of memory allocation](image)
Problem 1: How Does Everything Fit?

64-bit virtual addresses can address several exabytes
(18,446,744,073,709,551,616 bytes)

Physical main memory offers a few gigabytes
(e.g. 8,589,934,592 bytes)

(Not to scale; physical memory would be smaller than the period at the end of this sentence compared to the virtual address space.)

1 virtual address space per process, with many processes...
Problem 2: Memory Management

We have multiple processes:
- Process 1
- Process 2
- Process 3
- ... Process n

Each process has...
- stack
- heap
- .text
- .data
- ...

What goes where?
Problem 3: How To Protect

Physical main memory

Process \( i \)

Process \( j \)

Problem 4: How To Share?

Physical main memory

Process \( i \)

Process \( j \)
How can we solve these problems?

- “Any problem in computer science can be solved by adding another level of indirection.” – David Wheeler, inventor of the subroutine

- Without Indirection

- With Indirection

What if I want to move Thing?
Indirection

- **Indirection**: The ability to reference something using a name, reference, or container instead of the value itself. A flexible mapping between a name and a thing allows changing the thing without notifying holders of the name.
  - Adds some work (now have to look up 2 things instead of 1)
  - But don’t have to track all uses of name/address (single source!)

- **Examples**:
  - **Phone system**: cell phone number portability
  - **Domain Name Service (DNS)**: translation from name to IP address
  - **Call centers**: route calls to available operators, etc.
  - **Dynamic Host Configuration Protocol (DHCP)**: local network address assignment
Indirection in Virtual Memory

- Each process gets its own private virtual address space
- Solves the previous problems!
Address Spaces

- **Virtual address space:** Set of $N = 2^n$ virtual addr
  - $\{0, 1, 2, 3, \ldots, N-1\}$
- **Physical address space:** Set of $M = 2^m$ physical addr
  - $\{0, 1, 2, 3, \ldots, M-1\}$

- Every byte in main memory has:
  - one physical address (PA)
  - zero, one, *or more* virtual addresses (VAs)
Mapping

- A virtual address (VA) can be mapped to either physical memory or disk
  - Unused VAs may not have a mapping
  - VAs from different processes may map to same location in memory/disk
A System Using Physical Addressing

- Used in “simple” systems with (usually) just one process:
  - Embedded microcontrollers in devices like cars, elevators, and digital picture frames
A System Using Virtual Addressing

- Physical addresses are completely invisible to programs
  - Used in all modern desktops, laptops, servers, smartphones...
  - One of the great ideas in computer science
Why Virtual Memory (VM)?

- Efficient use of limited main memory (RAM)
  - Use RAM as a cache for the parts of a virtual address space
    - Some non-cached parts stored on disk
    - Some (unallocated) non-cached parts stored nowhere
  - Keep only active areas of virtual address space in memory
    - Transfer data back and forth as needed

- Simplifies memory management for programmers
  - Each process “gets” the same full, private linear address space

- Isolates address spaces (protection)
  - One process can’t interfere with another’s memory
    - They operate in different address spaces
  - User process cannot access privileged information
    - Different sections of address spaces have different permissions
VM and the Memory Hierarchy

- Think of virtual memory as an array of \( N = 2^n \) contiguous bytes.
- *Pages* of virtual memory are usually stored in physical memory, but sometimes spill to disk:
  - Pages are another unit of aligned memory (size is \( P = 2^p \) bytes).
  - Each virtual page can be stored in *any* physical page (no fragmentation!).

```
Virtual memory

Virtual pages (VP's)

- VP 0
- VP 1
- VP 2^{n-p-1}

Physical memory

Physical pages (PP's)

- PP 0
- PP 1
- PP 2^{m-p-1}

Unallocated

Unallocated

Unallocated

Disk

"Swap Space"
```
or: Virtual Memory as DRAM Cache for Disk

- Think of virtual memory as an array of \( N = 2^n \) contiguous bytes stored *on a disk*
- Then physical main memory is used as a cache for the virtual memory array
  - These “cache blocks” are called *pages* (size is \( P = 2^p \) bytes)
Memory Hierarchy: Core 2 Duo

Not drawn to scale

CPU Reg
32 KB
L1 I-cache
L1 D-cache

SRAM
Static Random Access Memory

DRAM
Dynamic Random Access Memory

Main Memory

Disk

Miss Penalty (latency)
33x

Miss Penalty (latency)
10,000x

Throughput:
16 B/cycle
8 B/cycle
2 B/cycle
1 B/30 cycles

Latency:
3 cycles
14 cycles
100 cycles
millions

Latency:
14 cycles
100 cycles
millions
Virtual Memory Design Consequences

- Large page size: typically 4-8 KiB or 2-4 MiB
  - Can be up to 1 GiB (for “Big Data” apps on big computers)
  - Compared with 64-byte cache blocks

- Fully associative
  - Any virtual page can be placed in any physical page
  - Requires a “large” mapping function – different from CPU caches

- Highly sophisticated, expensive replacement algorithms in OS
  - Too complicated and open-ended to be implemented in hardware

- Write-back rather than write-through
  - Really don’t want to write to disk every time we modify something in memory
  - Some things may never end up on disk (e.g. stack for short-lived process)
Why does VM work on RAM/disk?

- Avoids disk accesses because of \textit{locality}
  - Same reason that L1 / L2 / L3 caches work

- The set of virtual pages that a program is “actively” accessing at any point in time is called its \textit{working set}
  - If (working set of one process $\leq$ physical memory):
    - Good performance for one process (after compulsory misses)
  - If (working sets of all processes $>$ physical memory):
    - \textit{Thrashing}: Performance meltdown where pages are swapped between memory and disk continuously (CPU always waiting or paging)
    - This is why your computer can feel faster when you add RAM
Virtual Memory (VM)

- Overview and motivation
- VM as a tool for caching
- **Address translation**
- VM as a tool for memory management
- VM as a tool for memory protection
Address Translation

How do we perform the virtual → physical address translation?
Address Translation: Page Tables

- CPU-generated address can be split into:
  
  * $n$-bit address: \[
  \begin{array}{c|c}
  \text{Virtual Page Number} & \text{Page Offset} \\
  \end{array}
  \]

  - Request is Virtual Address (VA), want Physical Address (PA)
  - Note that Physical Offset = Virtual Offset (page-aligned)

- Use lookup table that we call the page table (PT)
  
  - Replace Virtual Page Number (VPN) for Physical Page Number (PPN) to generate Physical Address
  - Index PT using VPN: page table entry (PTE) stores the PPN plus management bits (e.g. Valid, Dirty, access rights)
  - Has an entry for every virtual page – why?
Page Table Diagram

- Page tables stored in physical memory
  - Too big to fit elsewhere – managed by MMU & OS
- How many page tables in the system?
  - One per process

Page table has $2^{n-p}$ entries!
Page Table Address Translation

In most cases, the MMU can perform this translation without software assistance.
Page Hit

- **Page hit:** VM reference is in physical memory

**Example:** Page size = 4 KiB

```
Virtual Addr: 0x00740b
VPN:         
Physical Addr: 
PPN:         
```

**Page Table (DRAM)**

<table>
<thead>
<tr>
<th>Valid</th>
<th>PPN/Disk Addr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>null</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Physical memory (DRAM)**

- VP 1
- VP 2
- VP 7
- VP 4

**Virtual memory (DRAM/disk)**

- VP 1
- VP 2
- VP 3
- VP 4
- VP 6
- VP 7
Page Fault

- **Page fault:** VM reference is NOT in physical memory

**Example:** Page size = 4 KiB
Provide a virtual address request (in hex) that results in this particular page fault:

**Virtual Addr:** 00
**Page Fault Exception**

- User writes to memory location
- That portion (page) of user’s memory is currently on disk

Page fault handler must load page into physical memory

Returns to faulting instruction: `mov` is executed again!

- Successful on second try
Handling a Page Fault

- Page miss causes page fault (an exception)
Handling a Page Fault

- Page miss causes page fault (an exception)
- Page fault handler selects a victim to be evicted (here VP 4)
Handling a Page Fault

- Page miss causes page fault (an exception)
- Page fault handler selects a \textit{victim} to be evicted (here VP 4)
Handling a Page Fault

- Page miss causes page fault (an exception)
- Page fault handler selects a *victim* to be evicted (here VP 4)
- Offending instruction is restarted: page hit!
Peer Instruction Question

- How many bits wide are the following fields?
  - 16 KiB pages
  - 48-bit virtual addresses
  - 16 GiB physical memory

<table>
<thead>
<tr>
<th></th>
<th>VPN</th>
<th>PPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>(B)</td>
<td>32</td>
<td>18</td>
</tr>
<tr>
<td>(C)</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>(D)</td>
<td>34</td>
<td>20</td>
</tr>
</tbody>
</table>
Virtual Memory (VM)

- Overview and motivation
- VM as a tool for caching
- Address translation
- VM as a tool for memory management
- VM as a tool for memory protection
VM for Managing Multiple Processes

- Key abstraction: each process has its own virtual address space
  - It can view memory as a simple linear array
- With virtual memory, this simple linear virtual address space need not be contiguous in physical memory
  - Process needs to store data in another VP? Just map it to any PP!

![Virtual Address Space Diagram]

- Virtual Address Space for Process 1:
  - VP 1
  - VP 2
  - ... (N-1)

- Virtual Address Space for Process 2:
  - VP 1
  - VP 2
  - ... (N-1)

- Physical Address Space (DRAM):
  - PP 2
  - PP 6
  - PP 8
  - ... (M-1)

Address translation
Simplifying Linking and Loading

**Linking**
- Each program has similar virtual address space
- Code, Data, and Heap always start at the same addresses

**Loading**
- `execve` allocates virtual pages for `.text` and `.data` sections & creates PTEs marked as invalid
- The `.text` and `.data` sections are copied, page by page, on demand by the virtual memory system
VM for Protection and Sharing

- The mapping of VPs to PPs provides a simple mechanism to *protect* memory and to *share* memory between processes
  - **Sharing**: map virtual pages in separate address spaces to the same physical page (here: PP 6)
  - **Protection**: process can’t access physical pages to which none of its virtual pages are mapped (here: Process 2 can’t access PP 2)
Memory Protection Within Process

* VM implements read/write/execute permissions
  * Extend page table entries with permission bits
  * MMU checks these permission bits on every memory access
    * If violated, raises exception and OS sends SIGSEGV signal to process (segmentation fault)

<table>
<thead>
<tr>
<th>Process i:</th>
<th>Valid</th>
<th>READ</th>
<th>WRITE</th>
<th>EXEC</th>
<th>PPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP 0:</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>PP 6</td>
</tr>
<tr>
<td>VP 1:</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>PP 4</td>
</tr>
<tr>
<td>VP 2:</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>PP 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Address Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP 2</td>
</tr>
<tr>
<td>PP 4</td>
</tr>
<tr>
<td>PP 6</td>
</tr>
<tr>
<td>PP 8</td>
</tr>
<tr>
<td>PP 9</td>
</tr>
<tr>
<td>PP 11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process j:</th>
<th>Valid</th>
<th>READ</th>
<th>WRITE</th>
<th>EXEC</th>
<th>PPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP 0:</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>PP 9</td>
</tr>
<tr>
<td>VP 1:</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>PP 6</td>
</tr>
<tr>
<td>VP 2:</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>PP 11</td>
</tr>
</tbody>
</table>
Review Question

What should the permission bits be for pages from the following sections of virtual memory?

<table>
<thead>
<tr>
<th>Section</th>
<th>Read</th>
<th>Write</th>
<th>Execute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stack</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Address Translation: Page Hit

1) Processor sends *virtual* address to MMU (*memory management unit*)

2-3) MMU fetches PTE from page table in cache/memory
(Uses PTBR to find beginning of page table for current process)

4) MMU sends *physical* address to cache/memory requesting data

5) Cache/memory sends data to processor

**VA = Virtual Address**  **PTEA = Page Table Entry Address**  **PTE = Page Table Entry**

**PA = Physical Address**  **Data = Contents of memory stored at VA originally requested by CPU**
Address Translation: Page Fault

1) Processor sends virtual address to MMU
2-3) MMU fetches PTE from page table in cache/memory
4) Valid bit is zero, so MMU triggers page fault exception
5) Handler identifies victim (and, if dirty, pages it out to disk)
6) Handler pages in new page and updates PTE in memory
7) Handler returns to original process, restarting faulting instruction
Hmm... Translation Sounds Slow

- The MMU accesses memory *twice*: once to get the PTE for translation, and then again for the actual memory request
  - The PTEs *may* be cached in L1 like any other memory word
    - But they may be evicted by other data references
    - And a hit in the L1 cache still requires 1-3 cycles

What can we do to make this faster?
- **Solution**: add another cache! 🎉
Speeding up Translation with a TLB

Translation "Lookaside Buffer" (TLB):
- Small hardware cache in MMU
- Maps virtual page numbers to physical page numbers
- Contains complete page table entries for small number of pages
  - Modern Intel processors have 128 or 256 entries in TLB
- Much faster than a page table lookup in cache/memory

Diagram: TLB

- VPN → PTE
- VPN → PTE
- VPN → PTE
TLB Hit
(page has been)
(used recently)

A TLB hit eliminates a memory access!
A TLB miss incurs an additional memory access (the PTE)

- Fortunately, TLB misses are rare
Fetching Data on a Memory Read

1) Check TLB
   - Input: VPN, Output: PPN
   - **TLB Hit:** Fetch translation, return PPN
   - **TLB Miss:** Check page table (in memory)
     * **Page Table Hit:** Load page table entry into TLB
     * **Page Fault:** Fetch page from disk to memory, update corresponding page table entry, then load entry into TLB

2) Check cache
   - Input: physical address, Output: data
   - **Cache Hit:** Return data value to processor
   - **Cache Miss:** Fetch data value from memory, store it in cache, return it to processor
Address Translation

Virtual Address

TLB Lookup

TLB Miss

Check the Page Table

(Valid = 0)

Page not in Mem

Page Fault (OS loads page)

Find in Disk

TLB Hit

Protection Check

PTE

PTE

PTE

Protection Fault

SIGSEGV

Physical Address

Access Denied

Access Permitted

Check cache

Access Permitted

Find in Mem

Update TLB

Find in Mem

Access Denied

Find in Disk
Context Switching Revisited

- What needs to happen when the CPU switches processes?
  - Registers:
    - Save state of old process, load state of new process
    - Including the Page Table Base Register (PTBR)
  - Memory:
    - Nothing to do! Pages for processes already exist in memory/disk and protected from each other
  - TLB:
    - \textit{Invalidate} all entries in TLB – mapping is for old process’ VAs
  - Cache:
    - Can leave alone because storing based on PAs – good for shared data
Summary of Address Translation Symbols

❖ Basic Parameters
  ▪ $N = 2^n$ Number of addresses in virtual address space
  ▪ $M = 2^m$ Number of addresses in physical address space
  ▪ $P = 2^p$ Page size (bytes)

❖ Components of the virtual address (VA)
  ▪ VPO Virtual page offset
  ▪ VPN Virtual page number
  ▪ TLBI TLB index
  ▪ TLBT TLB tag

❖ Components of the physical address (PA)
  ▪ PPO Physical page offset (same as VPO)
  ▪ PPN Physical page number
Virtual Memory Summary

- **Programmer’s view of virtual memory**
  - Each process has its own private linear address space
  - Cannot be corrupted by other processes

- **System view of virtual memory**
  - Uses memory efficiently by caching virtual memory pages
    - Efficient only because of locality
  - Simplifies memory management and sharing
  - Simplifies protection by providing permissions checking
Memory System Summary

- Memory Caches (L1/L2/L3)
  - Purely a speed-up technique
  - Behavior invisible to application programmer and (mostly) OS
  - Implemented totally in hardware

- Virtual Memory
  - Supports many OS-related functions
    - Process creation, task switching, protection
  - Operating System (software)
    - Allocates/shares physical memory among processes
    - Maintains high-level tables tracking memory type, source, sharing
    - Handles exceptions, fills in hardware-defined mapping tables
  - Hardware
    - Translates virtual addresses via mapping tables, enforcing permissions
    - Accelerates mapping via translation cache (TLB)