Virtual Memory III
CSE 351 Spring 2020

Instructor:
Ruth Anderson

Teaching Assistants:
Alex Olshanskyy
Rehaan Bhimani
Callum Walker
Chin Yeoh
Diya Joy
Eric Fan
Edan Sneh
Jonathan Chen
Jeffery Tian
Millicent Li
Melissa Birchfield
Porter Jones
Joseph Schafer
Connie Wang
Eddy (Tianyi) Zhou

https://xkcd.com/2308/
Administrivia

- Lab 4 – Due Friday 5/22
  - Cache parameter puzzles and code optimizations

- You must log on with your @uw google account to access!!
  - Google doc for 11:30 Lecture: https://tinyurl.com/351-05-20A
  - Google doc for 2:30 Lecture: https://tinyurl.com/351-05-20B
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
  - Split VPN into TLB Tag and TLB Index based on # of sets in TLB
- Maps virtual page numbers to physical page numbers
- Stores page table entries for a small number of pages
  - Modern Intel processors have 128 or 256 entries in TLB
- Much faster than a page table lookup in cache/memory
A TLB hit eliminates a memory access!
**TLB Miss**

- A TLB miss incurs an additional memory access (the PTE)
  - Fortunately, TLB misses are rare
Fetching Data on a Memory Read

1) Check TLB (translate \( VA \rightarrow PA \))
   - **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 (fetch requested data)
   - **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

Page Fault
(OS loads page)

Find in Disk

Val \neq 0
Page not in Mem

Val = 1
Page in Mem

Update TLB

Find in Mem

Protection Check

Protection Fault

Access Denied

SIGSEGV

Access Permitted

Physical Address

Check cache

Miss

Hit
Address Manipulation

request from CPU: \( n \)-bit virtual address

split to access TLB: TLB Tag, TLB Index, Page Offset

(on TLB miss) access PT: Virtual Page Number, Page offset

\( m \)-bit physical address:

split to access cache: Physical Page Number, Page offset

Cache Tag, Cache Index, Offset

TRANSLATION
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:
    - *Invalidate* all entries in TLB – mapping is for old process’ VAs
  - Cache: *Physically Indexed*
    - 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
Simple Memory System Example (small)

- **Addressing**
  - 14-bit virtual addresses \( n = 14 \text{ bits} \) \( \iff \) \( N = 16 \text{ KiB} \) \( \text{VA space} \)
  - 12-bit physical address \( m = 12 \text{ bits} \) \( \iff \) \( M = 4 \text{ KiB} \) \( \text{PA space} \)
  - Page size = 64 bytes \( P = 64 \text{ B} \) \( \iff \) \( p = 6 \text{ bits} \)

```
<table>
<thead>
<tr>
<th>13</th>
<th>12</th>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
</table>
```

- \( 2^{n-p} \) pages in \( \text{VA space} \)
- \( 2^{m-p} \) pages in \( \text{PA space} \)

```
<table>
<thead>
<tr>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
</table>
```

- \( \text{VPN width} = n-p \) 
- \( \text{VPO} \) 
- Virtual Page Number 
- Virtual Page Offset 

```
<table>
<thead>
<tr>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
</table>
```

- \( \text{PPN width} = m-p \) 
- \( \text{PPN} \) 
- Physical Page Number 
- Physical Page Offset 

```
<table>
<thead>
<tr>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
</table>
```
Simple Memory System: Page Table

- Only showing first 16 entries (out of $2^n = 2^{15} = 2^{56}$) one for every virtual page
  - **Note**: showing 2 hex digits for PPN even though only 6 bits
  - **Note**: other management bits not shown, but part of PTE

<table>
<thead>
<tr>
<th>VPN</th>
<th>PPN</th>
<th>Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>02</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0x13</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>09</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>2D</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>0D</td>
<td>1</td>
</tr>
</tbody>
</table>
Simple Memory System: TLB

- 16 entries total
- 4-way set associative

The TLB tag and index are used to look up entries in the TLB. The virtual page number and offset are used to determine the physical page number (PPN).

Why does the TLB ignore the page offset? It is not part of its job (address translation).

Why does the TLB have 16 entries total? The equation $16/4 = 4$ sets explains this. Each set is 4-way set associative, meaning each set can have up to 4 valid entries.

The diagram shows the organization of the TLB entries. Each set has 4 ways, and each way contains a tag, PPN, and valid bit.
Simple Memory System: Cache

- Direct-mapped with $K = 4$ B, $C/K = 16$
- Physically addressed

Note: It is just coincidence that the PPN is the same width as the cache Tag
Current State of Memory System

Page table (partial):

TLB:

<table>
<thead>
<tr>
<th>Set</th>
<th>Tag</th>
<th>PPN</th>
<th>V</th>
<th>Tag</th>
<th>PPN</th>
<th>V</th>
<th>Tag</th>
<th>PPN</th>
<th>V</th>
<th>Tag</th>
<th>PPN</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>03</td>
<td>–</td>
<td>0</td>
<td>09</td>
<td>0D</td>
<td>1</td>
<td>00</td>
<td>–</td>
<td>0</td>
<td>07</td>
<td>02</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>02</td>
<td>–</td>
<td>0</td>
<td>08</td>
<td>–</td>
<td>0</td>
<td>06</td>
<td>–</td>
<td>0</td>
<td>03</td>
<td>–</td>
<td>0X</td>
</tr>
<tr>
<td>2</td>
<td>07</td>
<td>–</td>
<td>0</td>
<td>03</td>
<td>0D</td>
<td>1</td>
<td>0A</td>
<td>34</td>
<td>1</td>
<td>02</td>
<td>–</td>
<td>0</td>
</tr>
</tbody>
</table>

Cache:

<table>
<thead>
<tr>
<th>Index</th>
<th>Tag</th>
<th>V</th>
<th>B0</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>1</td>
<td>99</td>
<td>11</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>1B</td>
<td>1</td>
<td>00</td>
<td>02</td>
<td>04</td>
<td>08</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>1</td>
<td>43</td>
<td>6D</td>
<td>8F</td>
<td>09</td>
</tr>
<tr>
<td>5</td>
<td>0D</td>
<td>1X</td>
<td>36</td>
<td>72</td>
<td>F0</td>
<td>1D</td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>1</td>
<td>11</td>
<td>C2</td>
<td>DF</td>
<td>03</td>
</tr>
</tbody>
</table>

Index:

<table>
<thead>
<tr>
<th>Tag</th>
<th>V</th>
<th>B0</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>28</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>0A</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>02</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>09</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>2D</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>0X</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>0D</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Polling Question [VM III]

Memory Request Example #1

- Virtual Address: 0x03D4

**TLBT**

\[0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\]

**VPO**

- VPN 0xF (check this entry of the page table)
- TLBT 0x03 (look for this tag within TLB set)
- TLBI 3 (check this set of the TLB)

TLB Hit? Y
Page Fault? N
PPN 0x0D

- Physical Address:

**CT**

\[0\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\]

**Co**

CT 0x0D (look for this tag within cache set)
CI 5 (check this set of the cache)
CO 0 (which byte of cache block)

Cache Hit? Y
Data (byte) 0x36

Note: It is just coincidence that the PPN is the same width as the cache Tag

Give your answer for Data(byte) at: http://pollev.com/reaction
Memory Request Example #2

- Virtual Address: \(0x038F\)
  - VPN: 0x0E
  - TLBT: 0x03
  - TLBI: 2
  - TLB Hit? N
  - Page Fault? Y
  - PPN \(\text{NA}\)

- Physical Address:
  - CT: 
  - CI: 
  - CO: 
  - Cache Hit? 
  - Data (byte) 

Note: It is just coincidence that the PPN is the same width as the cache Tag
Memory Request Example #3

- **Virtual Address:** 0x0020

  ![TLB Diagram](image)

  - VPN 0x00
  - TLBT 0x00
  - TLBI 0
  - TLB Hit? N
  - Page Fault? N
  - PPN 0x28

- **Physical Address:**

  ![Physical Address Diagram](image)

  - CT 0x28
  - CI 8
  - CO 0
  - Cache Hit? N
  - Data (byte) n/a
Memory Request Example #4

- **Virtual Address:** 0x036B
  - TLBT: 00000110110110101011
  - TLBI: 00000110110110101011
  - VPN: 0x0D
  - TLBT: 0x03
  - TLBI: 1
  - TLB Hit? Y
  - Page Fault? N
  - PPN: 0x2D

- **Physical Address:**
  - CT: 0x2D
  - CI: 3
  - CO: 3
  - Cache Hit? Y
  - Data (byte): 0x3B

**Note:** It is just coincidence that the PPN is the same width as the cache Tag.
Memory Overview

- `movl 0x8043ab, %rdi`
Page Table Reality

- Just one issue... the numbers don’t work out for the story so far!

- The problem is the page table for each process:
  - Suppose 64-bit VAs, 8 KiB pages, 8 GiB physical memory
  - How many page table entries is that?
    - 1 PTE for every virtual page
    \[ 2^{n-p} \approx 2^{51} \text{ PTEs} \]
  - About how long is each PTE?
    \[ \text{PPN width } + \text{ management bits } = 20 + 5 = 25 \text{ bits } \approx 3 \text{ bytes} \]

- Moral: Cannot use this naïve implementation of the virtual→physical page mapping – it’s way too big
A Solution: Multi-level Page Tables

This is called a **page walk**

This is extra (non-testable) material
Multi-level Page Tables

- A tree of depth $k$ where each node at depth $i$ has up to $2^j$ children if part $i$ of the VPN has $j$ bits
- Hardware for multi-level page tables inherently more complicated
  - But it’s a necessary complexity – 1-level does not fit
- Why it works: Most subtrees are not used at all, so they are never created and definitely aren’t in physical memory
  - Parts created can be evicted from cache/memory when not being used
  - Each node can have a size of ~1-100KB
- But now for a $k$-level page table, a TLB miss requires $k + 1$ cache/memory accesses
  - Fine so long as TLB misses are rare – motivates larger TLBs
Practice VM Question

- Our system has the following properties
  - 1 MiB of physical address space
  - 4 GiB of virtual address space
  - 32 KiB page size
  - 4-entry fully associative TLB with LRU replacement

```
2^n-p ≤ # of virtual pages

2^{5} ≤ # of pages in physical memory
```

a) Fill in the following blanks:

- 2^{17} Entries in a page table
- 20 Minimum bit-width of PTBR
- 17 TLBT bits
- 2^{5} Max # of valid entries in a page table
**Practice VM Question**

- One process uses a page-aligned *square* matrix `mat[]` of 32-bit integers in the code shown below:
  
  ```c
  #define MAT_SIZE = 2048
  for(int i = 0; i < MAT_SIZE; i++)
    mat[i*(MAT_SIZE+1)] = i;
  ```

- b) What is the largest stride (in bytes) between successive memory accesses (in the VA space)?

  - The stride is always `2049` bytes, as each access is incremented by `2049` in the array index, hence the total space occupied by `2049` integers is `2049 * 4` bytes.
Practice VM Question

Page size = \(32 \text{ KiB} = 2^{15} \text{ B}\)

One process uses a page-aligned square matrix `mat[]` of 32-bit integers in the code shown below:

```c
#define MAT_SIZE = 2048
for(int i = 0; i < MAT_SIZE; i++)
  mat[i*(MAT_SIZE+1)] = i;
```

c) Assuming all of `mat[]` starts on disk, what are the following hit rates for the execution of the for-loop?

<table>
<thead>
<tr>
<th>Access Pattern</th>
<th>TLB Hit Rate</th>
<th>Page Table Hit Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>single write to index, never revisit indices (always increasing), we access every row of matrix exactly once</td>
<td>(\frac{3}{4} = 75%)</td>
<td>(0%)</td>
</tr>
</tbody>
</table>

Each page holds \(2^{15} / 2^{13} = 4\) rows of matrix within each page: MTHH
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)