EECS 482
Introduction to Operating Systems

Winter 2019

Manos Kapritsos

Thanks to Harsha Madhyastha and Peter Chen for the slides and notes
Recap: Paging

- Both address spaces and physical memory broken up into fixed size pages
Paging

if (virtual page is invalid or non-resident or protected) {
    trap to OS fault handler; retry
} else {
    physical page # = pageTable[virtual page #].physPageNum
}
Paging

● Pros?
  ◆ Flexible memory allocation/growing address space
  ◆ Virtual memory
  ◆ No external fragmentation
  ◆ Flexible sharing

● Cons?
  ◆ Large page tables

● How to modify paging to reduce space needed for translation data?
Multi-level Paging

- Standard page table is a simple array
- Multi-level paging generalizes this into a tree

Example: Two-level page table with 4KB pages
  - Index into level 1 page table: virtual address bits 31-22
  - Index into level 2 page table: virtual address bits 21-12
  - Page offset: bits 11-0
Multi-level Paging

level 1 page table

<table>
<thead>
<tr>
<th>virtual address bits 21-12</th>
<th>physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

level 2 page tables

<table>
<thead>
<tr>
<th>virtual address bits 21-12</th>
<th>physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

How does this let translation data take less space?
Sparse Address Space

<table>
<thead>
<tr>
<th>Virtual page #</th>
<th>Physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>105</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>283</td>
</tr>
<tr>
<td>3</td>
<td>invalid</td>
</tr>
<tr>
<td>...</td>
<td>invalid</td>
</tr>
<tr>
<td>1048572</td>
<td>invalid</td>
</tr>
<tr>
<td>1048573</td>
<td>1078</td>
</tr>
<tr>
<td>1048574</td>
<td>48136</td>
</tr>
<tr>
<td>1048575</td>
<td>60</td>
</tr>
</tbody>
</table>
### Sparse Address Space

<table>
<thead>
<tr>
<th>Bits 31-22</th>
<th>Physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>389</td>
</tr>
<tr>
<td>1</td>
<td>invalid</td>
</tr>
<tr>
<td>2</td>
<td>invalid</td>
</tr>
<tr>
<td>...</td>
<td>invalid</td>
</tr>
<tr>
<td>1021</td>
<td>invalid</td>
</tr>
<tr>
<td>1022</td>
<td>invalid</td>
</tr>
<tr>
<td>1023</td>
<td>7046</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bits 21-12</th>
<th>Physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>105</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>283</td>
</tr>
<tr>
<td>3</td>
<td>invalid</td>
</tr>
<tr>
<td>...</td>
<td>invalid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bits 21-12</th>
<th>Physical page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>invalid</td>
</tr>
<tr>
<td>1020</td>
<td>invalid</td>
</tr>
<tr>
<td>1021</td>
<td>1078</td>
</tr>
<tr>
<td>1022</td>
<td>48136</td>
</tr>
<tr>
<td>1023</td>
<td>60</td>
</tr>
</tbody>
</table>
Multi-level paging

- How to share memory between address spaces?
  - page, level-2 page table, or entire address space
- What must be changed on a context switch?
  - PTBR

- Pros:
  - Easy memory allocation
  - Flexible sharing
  - Space efficient for sparse address spaces

- Cons?
  - Two extra lookups per memory reference
Translation Lookaside Buffer

- TLB caches virtual page # to PTE mapping
  - Cache hit $\rightarrow$ Skip all the translation steps
  - Cache miss $\rightarrow$ Get PTE, store in TLB, restart instruction

- Does this change what happens on a context switch?
  - Need to invalidate TLB cache
Page Replacement

- Not all valid pages may fit in physical memory
  - Some pages are swapped out to disk

- To read in a page from disk, some resident page must be swapped out to disk

- Which page to evict when you need a free page?
  - Goal: minimize page faults
Replacement policies

- Random

- FIFO
  - Replace page brought into memory longest time ago
  - May replace pages that continue to be frequently used

- Optimal?
  - Replace page that won’t be used for the longest time in the future
  - Minimizes misses, but requires knowledge of the future
Replacement policies

- **LRU (least recently used)**
  - Approximates OPT by using past reference pattern
    » If page hasn’t been used for a while, it probably won’t be used for a long time in the future

- **LRU is hard to implement exactly**
  - Can we simplify LRU by approximating it?
The “referenced” bit

- Most MMUs maintain a “referenced” bit for each resident page
  - Set by MMU when page is read or written
  - Can be cleared by OS
- How to use reference bit to identify old pages?
  - Periodically reset it for all pages
- Why maintain reference bit in hardware?
  - It’s faster
Clock replacement algorithm

- Arrange resident pages around a clock

- Algorithm to select page for eviction:
  - Consider page pointed to by clock hand
  - If not referenced, page hasn’t been accessed since last sweep → Evict
  - If referenced, page has been referenced since last sweep
    » What to do?

- What if all pages have been referenced since last sweep?
  - First sweep will de-reference them all

- What about new pages?
  - Put them to the left of the hand
A nice feature of clock is that it only adds overhead when you need to evict a page.
Administrivia

- Project 2 due today!
- Midterm on Friday
- Traveling this week
  - Office hours canceled
  - Wednesday lecture by Baris Kasikci

- You can start working on Project 3
  - Covered most of the material you need to know to do the project
Page eviction

- Where to evict page to? Disk
- When do you NOT need to write page to disk?
  - Rely on hardware/MMU to maintain dirty bit in PTE
- Why not write to disk on every store?
  - Too expensive
- The page that is brought from disk must wait while some page is evicted
  - How could you avoid this (sometimes)?
    » Prefer to evict clean pages over dirty ones
- DON’T use these optimizations in Project 3!
Defer work (clarification)

How to defer work?
I am deferring work as per the spec and numerous instructors. Is it still too early to start?

Please help. I am confused.

- The OS should defer work, not you!
if (virtual page is non-resident or protected) {
    trap to OS fault handler
    retry access
} else {
    physical page # = pageTable[virtual page #].physPageNum
    pageTable[virtual page #].referenced = true
    if (access is write) {
        pageTable[virtual page #].dirty = true
    }
    access physical memory
}
## Page table contents

<table>
<thead>
<tr>
<th>Physical page #</th>
<th>Resident</th>
<th>Read/Write enabled</th>
<th>Dirty</th>
<th>Referenced</th>
</tr>
</thead>
</table>

- Why no valid bit in PTE?
  - All invalid virtual pages are non-resident

- For valid non-resident pages, does PTE contain disk block?
  - OS must maintain this, MMU simply traps to OS

- Can we make do without resident bit?
  - Use protection bits
Page table contents

<table>
<thead>
<tr>
<th>Physical page #</th>
<th>Read/Write enabled</th>
<th>Dirty</th>
<th>Referenced</th>
</tr>
</thead>
</table>

- Can we make do without dirty bit?
  - Use protection bits
  - Won’t this increase # of page faults?
    » Only trap when the page is clean
## Page table contents

<table>
<thead>
<tr>
<th>Physical page #</th>
<th>Read/Write enabled</th>
<th>Referenced</th>
</tr>
</thead>
</table>

- Can we make do without referenced bit?
  - Same trick: use protection bits
- Application too may want to control protection
  - Not in project 3
## Page table contents

<table>
<thead>
<tr>
<th>Physical page #</th>
<th>read_enabled</th>
<th>write_enabled</th>
</tr>
</thead>
</table>
