How to streamline your life (lessons from computer architecture).

COS 116, Spring 2010
Guest: Szymon Rusinkiewicz
Lesson 1: Caching
(and the 80-20 rule)
The Tired Librarian

- 1000 checkouts/returns per day
- Distance covered = 50 x 2 x 1000 = 100,000 feet
  ~ 20 miles
- Please help!!!
80-20 “Rule”

- Pareto [1906]: 20% of the people own 80% of the wealth

- Juran [1930’s]: 20% of the organization does 80% of the work
Better Arrangement

- Distance covered per day?

Reserves

“Most popular” shelf: 20% most popular books

50 ft

5 ft
Even better arrangement?

- Distance covered per day?
- Reserves
  - "Most popular" shelf: 20% most popular books
  - Top 4%
Discussion
Time

- Is the librarian’s problem solved?
How to predict the 20% most popular books for next day?

- In general, no easy solution

- In practice, use rules of thumb
  - Example: “Least Recently Used”. When you need to create space on the desk (or shelf), move out the book that was used least recently
  - Many others (LRU is computationally expensive)
New and improved

<table>
<thead>
<tr>
<th>XPS 600</th>
<th>XPS 200</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Power Unleashed</strong></td>
<td><strong>Small, But Mighty</strong></td>
</tr>
<tr>
<td><strong>Processor</strong></td>
<td><strong>Processor</strong></td>
</tr>
<tr>
<td>Intel® Pentium® 4 Processor 640 with HT (3.20GHz, 800 FSB, 2MB L2 cache)</td>
<td>Intel® Pentium® 4 Processor with Hyper-Threading Technology - 600 Sequence - up to Pentium Extreme Edition Dual Core</td>
</tr>
<tr>
<td><strong>Operating System</strong></td>
<td><strong>Operating System</strong></td>
</tr>
</tbody>
</table>
Connection to Computer Organization

- Speed vs cost of various memories

<table>
<thead>
<tr>
<th>Memory Type</th>
<th>Cost: $ / GB</th>
<th>Speed: GB/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard drive</td>
<td>0.10</td>
<td>0.1</td>
</tr>
<tr>
<td>Flash (e.g., SSD, USB stick)</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>RAM</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>On-chip memory for CPU (L2 Cache)</td>
<td>40000</td>
<td>20</td>
</tr>
</tbody>
</table>
Often, today’s computers have even more levels of caching.
Moral

- Performance:
  - Speed is close to that of fastest memory (cache)
  - Overall capacity is that of largest memory (disk)
Question

- How does the same program (.exe file) run on different PCs with different memory configurations?

Answer: “Virtual Memory”

- All programs live a fiction: allowed to pretend they each have $2^{32}$ or $2^{64}$ bytes of memory
- Illusion is preserved by hardware
Virtual Memory

- Program’s view:

- Underlying truth:
Lesson 2: Multitasking

- “The Multitasking Generation”
An Evening’s Tasks for a Gen-M’er

- Homework
- Listen to music
- Instant Messaging
- Call Mom (goes to bed by 11 PM!)
- Answer phone
- Read a bit more of Joyce’s *Ulysses*
- Watch the Daily Show

How do you do it all?
Tasks done by my PC last night

- Word processing
- Play CD
- Download news updates
- Download email
- Run clock
- Hidden tasks: handle network traffic, manage disk and RAM traffic, scheduler, etc.

Managed by “Operating System” (WinXP, Linux, MacOS, etc.)
Multitasking versus Parallel Processing

Multitasking: A single CPU handles many tasks by switching rapidly among them. (e.g., all Wintel machines since early 1990s; all Unix machines since the 1970s)

Parallel Processing: Multiple CPUs that do the work of a single CPU. (But, 4 CPUs do not necessarily mean 4x speed.)
Scheduler’s objectives

- Fairness
- Timeliness
- Critical tasks processed promptly
- Low overhead

How can one achieve these (often conflicting) goals?