CS 411 Translation-LookAside Buffers

Problem: Paging increases memory access time by 2.

Solution: Cache a subset of a process' page table in the MMU - TLB

What is caching?

- A technique used to ~~increase~~ decrease the latency of some part of the memory hierarchy or increase the apparent size of some part of the memory hierarchy.

Latency ≡ Response time

Memory Hierarchy

CPU Registers

→ Data / Instruction Caches | Decrease RAM latency
TLB

RAM

Block Cache - Windows Ready Boost
Linux Bcache

SSD / HDD Storage

"Offline" Storage
USB Flash

Virtual Memory - Next - is used to increase the apparent size of RAM.

Effective RAM Access time

~~without~~ No TLB: AT_R – RAM access time for a single R/W

$EAT = 2AT_R$

With a TLB:

$$EAT = AT_R + [p\,AT_T + (1-p)AT_R]$$

AT_T – TLB access time

p – probability of TLB "hit"

$0 \leq p \leq 1$

Caches are "content addressable memories"

– (Key, Value) pairs

like a dictionary or a hash table

TLB -

Key is page #

Value is page table entry
for the page #

Design issues -

- TLB size
- Replacement Policy
- Context Switches
  - Flush
  - Add "PID" to key