EECS 482
Introduction to Operating Systems

Fall 2019

Baris Kasikci

Slides by: Harsha V. Madhyastha

Page table contents

Physical page #	Resident	Read/Write enabled	Dirty	Referenced

- 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

Physical page #	Read/Write enabled	Dirty	Referenced

- Can we make do without dirty bit?
  - Use protection bits
    - » Make clean pages write-protected
    - » Change the write-enabled bit after page becomes dirty
      - If the page is indeed not write-protected (e.g., code pages are write-protected)
  - Won’t this increase # of page faults a lot?

Page table contents

Physical page #	Read/Write enabled	Referenced

- Can we make do without referenced bit?
- Application too may want to control protection
  - Not in project 3

Page table contents

Physical page #	read_enabled	write_enabled

Address Space Management

- How to manage a process's accesses to its address space?
  - Kernel sets up page table per process and manages which pages are resident
  - MMU looks up page table to translate any virtual address to a physical memory address

- What about kernel's address space?
- How does MMU handle kernel's loads and stores?

Storing Page Tables

- Two options:
  1. In physical memory
  2. In kernel’s virtual address space
- Difference: Is PTBR a physical or virtual addr?
- Pros and cons of option 2?

- Project 3 uses option 2
  - Kernel’s address space managed by infrastructure

Kernel vs. user address spaces

- Can you evict the kernel's virtual pages?
- How can kernel access specific physical memory addresses (e.g., to refer to translation data)?

How does kernel access user's address space?

- Kernel can manually translate a user virtual address to a physical address, then access the physical address

- Can map kernel address space into every process's address space

fffff . . . 80000	operating system
7ffff . . . 00000	user process

  - Trap to kernel doesn't change address spaces; it just enables access to both OS and user parts of that address space

Kernel vs. user mode

- How are we protecting a process' address space from other processes?
- Must ensure that only kernel can modify translation data
- How does CPU know kernel is running?

- Recap of protection:

Switching from user process into kernel

- Faults and interrupts
  - Timer interrupts
  - Page faults
  - Why are these safe to transfer control to kernel?

- System calls
  - Process management: fork/exec
  - I/O: open, close, read, write
  - System management: reboot
  - …

System calls

- When you call `cin` in your C++ program:
  - `cin` calls `read()`, which executes assembly-language instruction `syscall`
  - `syscall` traps to kernel at pre-specified location
  - kernel's syscall handler calls kernel's `read()`
- To handle trap to kernel, hardware atomically
  - Sets mode bit to kernel
  - Saves registers, PC, SP
  - Changes SP to kernel stack
  - Changes to kernel's address space
  - Jumps to exception handler

Arguments to system calls

- Two options:
  - Store in registers
  - Store in memory (in whose address space?)

- Kernel must check validity of arguments
  - e.g., `read(int fd, void *buf, size_t size)`

Protection summary

- Safe to switch from user to kernel mode because control only transferred to certain locations
  - Where are these locations stored?
- Who can modify interrupt vector table?
- Why is it easier to control access to interrupt vector table than mode bit?

Address Space Protection

- How are address spaces protected?
- How is translation data protected?
- How is mode bit protected?
- Protection boils down to init process which sets up interrupt vector table when system boots up

Project 3

- Memory management using paging
  - Due Nov 12th
- By the end of this lecture, we will cover all the material you need to know to do the project
- Begin drawing a state machine for a virtual page first
  - Focus on swap-backed pages first (before file-backed pages)
- Avoid doing unnecessary work

Project 3

- Incremental development critical
  - Swap-backed pages with a single process
  - File-backed pages
  - Fork
- Minimum amount of functionality to test
  - vm_init
  - vm_create (with parent process unknown)
  - vm_map (with filename == NULL)
  - Getting this combination right = ~1/3 of the grade

Process creation

- :(){ :|:&};:
  - : ()   -> define a function called :
  - { :|:&} -> the function sends its output to the : function again and runs that in the background.
  - ; is the command separator
  - : runs the function the first time

Unix process creation

- System calls to start a process:
  1. Fork() creates a copy of current process
  2. Exec(program, args) replaces current address space with specified program

- Why first copy and then overwrite?

- Any problems with child being an **exact** clone of parent?

Unix process creation

- Fork uses return code to differentiate
  - Child gets return code 0
  - Parent gets child's unique process id (pid)

```
If (fork() == 0) {
    exec ();          /* child */
} else {
    /* parent */
}
```

Subtleties in handling fork

- Buggy code from autograder:

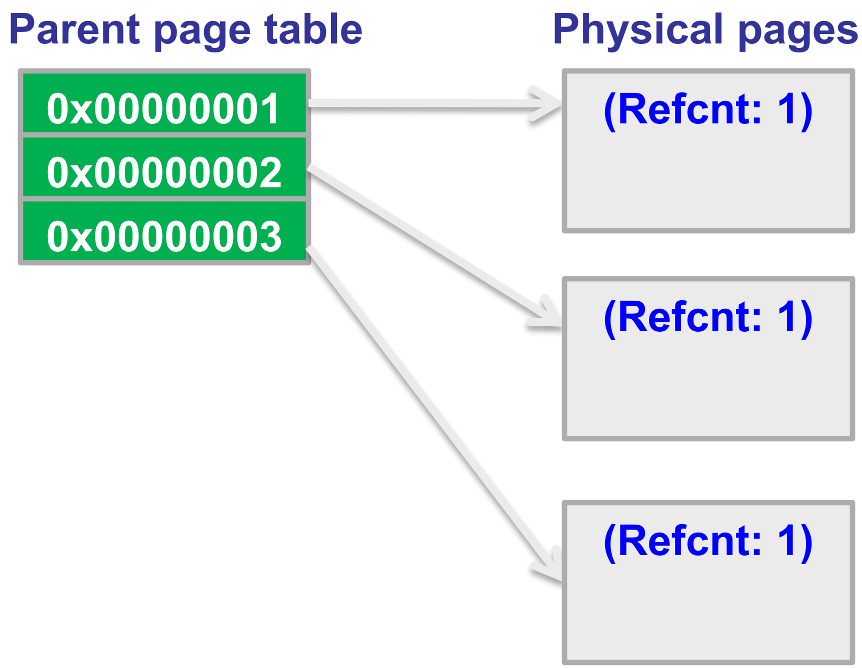
```
if (!fork()) {
        exec(command);
}
while(child is alive) {
        if (size of child address space > max) {
                print "process took too much
memory”;
                kill child;
                break;
        }
}
```

- What is the bug here?

Avoiding work on fork

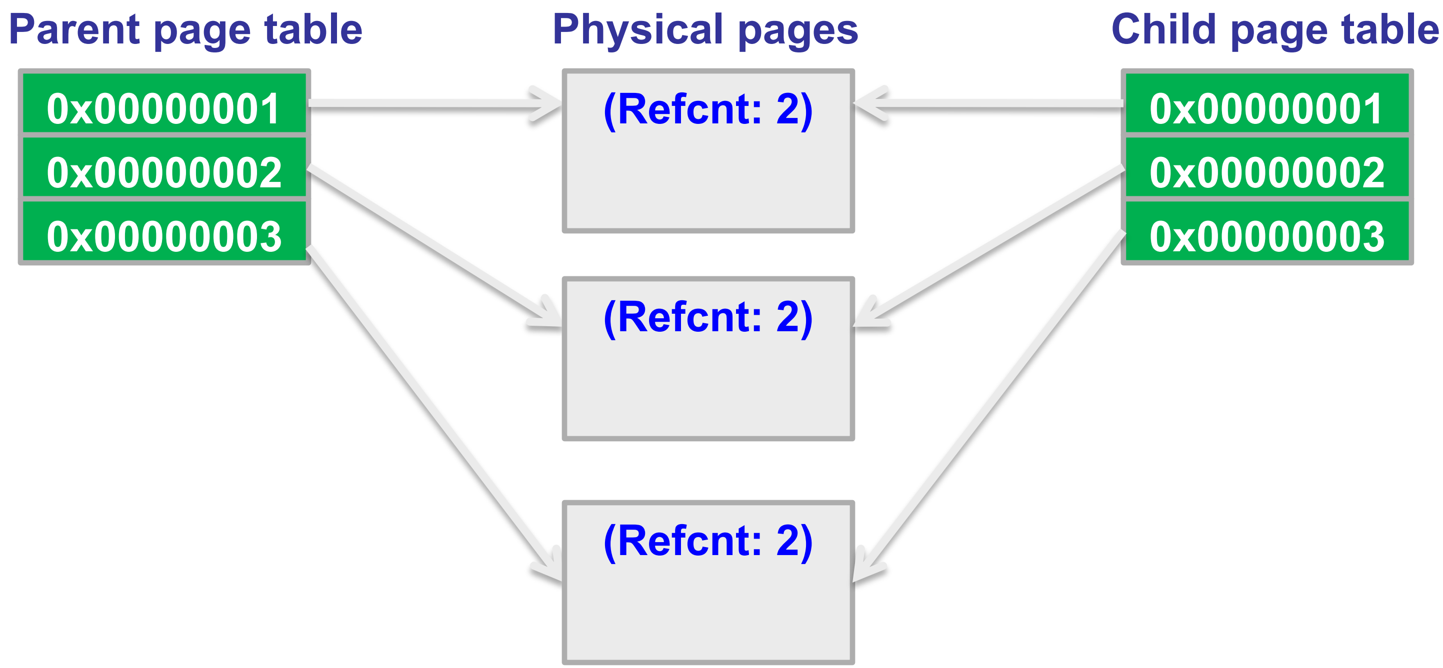
- Copying entire address space is expensive
- Instead, Unix uses **copy-on-write**
  - Assign reference count to each physical page
  - On fork(), copy only the page table of parent
    - » Increment reference count by one
  - On store by parent or child to page with refcnt > 1:
    - » Make a copy of the page; set refcnt to one for that page
    - » Modify PTE of modifier to point to new page
    - » Decrement reference count of old page

Copy-on-write: Example



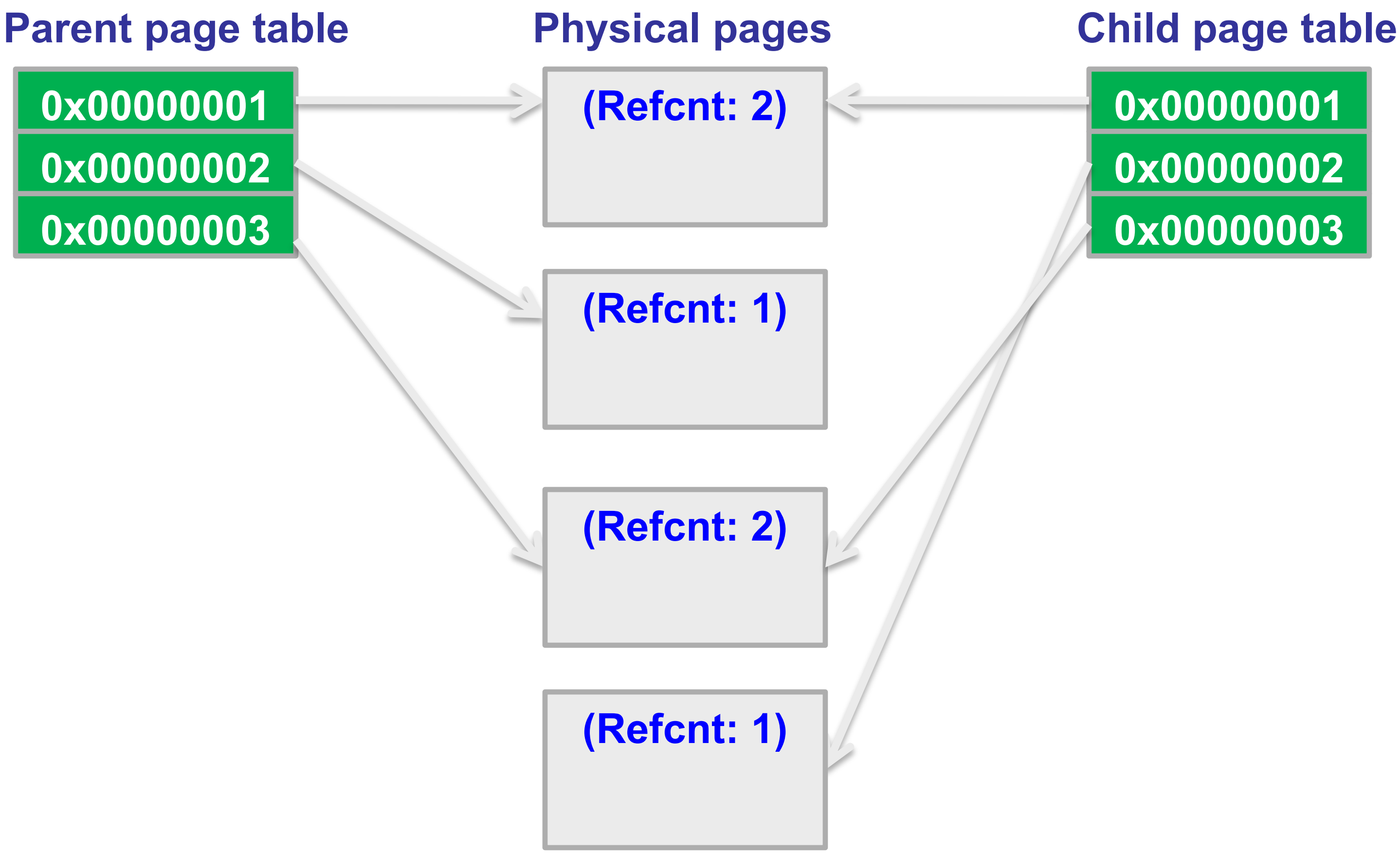
Parent about to fork()

Copy-on-write: Example



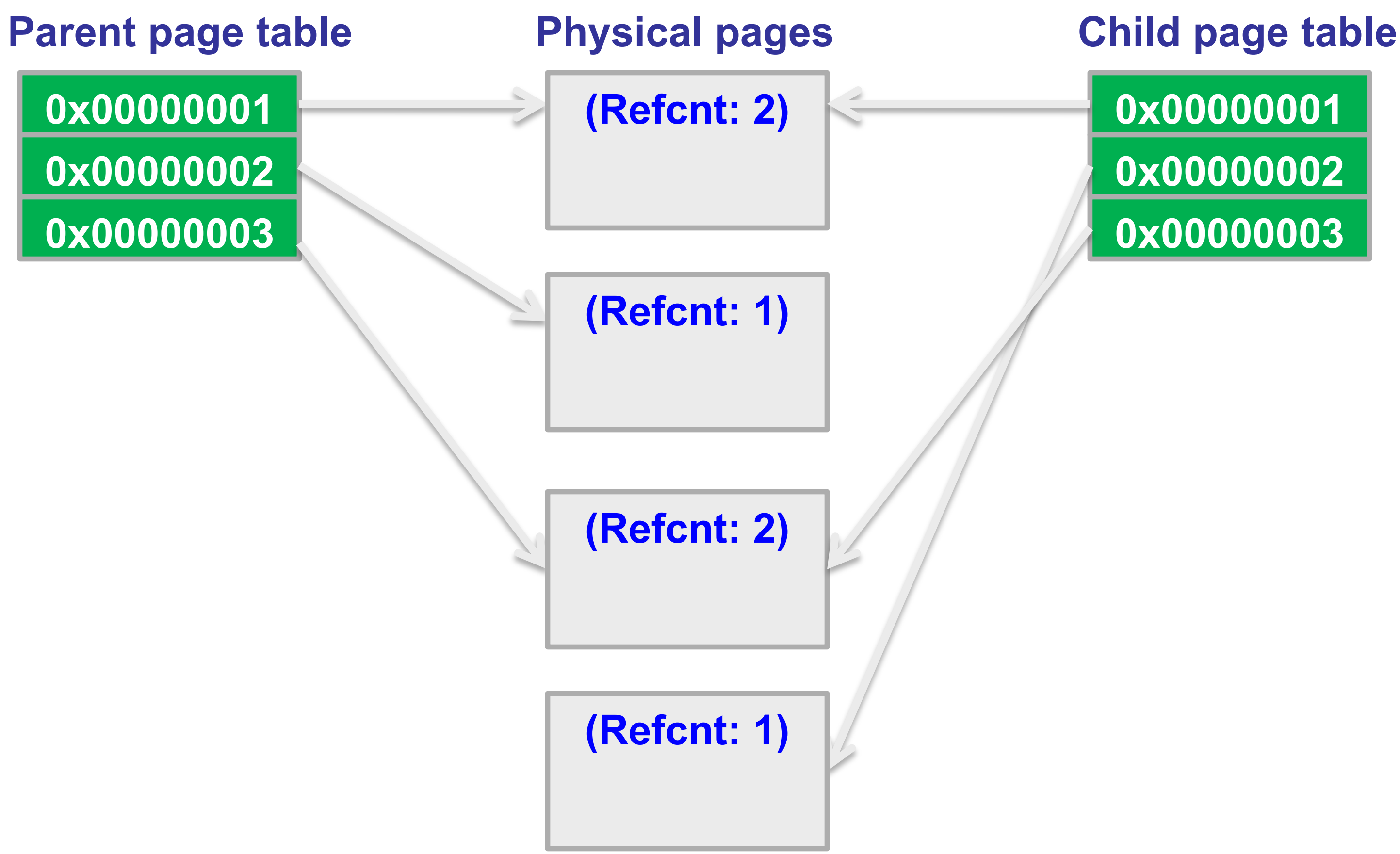
Copy-on-write of parent address space

Copy-on-write: Example



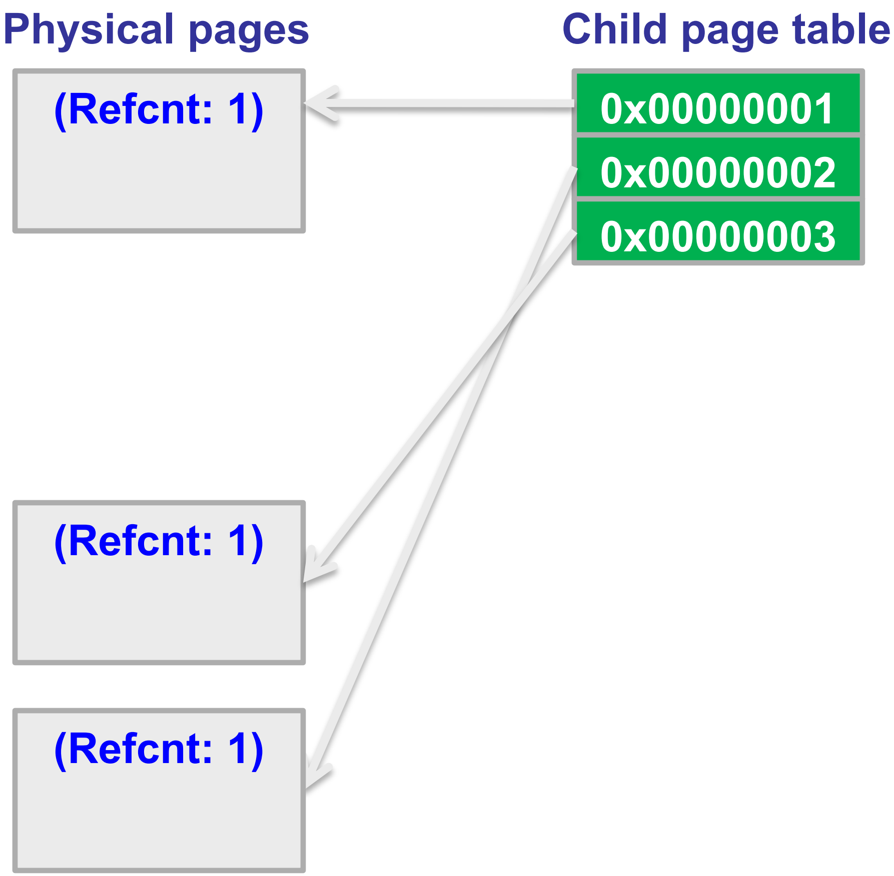
Child modifies 2nd virtual page

Copy-on-write: Example



Parent modifies 2nd virtual page

Copy-on-write: Example



Parent exits

Implementing a shell

```
while (1) {
    print prompt
    ask user for input (cin)
    parse input //split into command and args
    fork a copy of current process (the shell prog.)
    if (child) {
        redirect output to a file/pipe, if requested
        exec new program with arguments
    } else { //parent
        wait for child to finish, or
        run child in the background and ask for
another command
    }
}
```