CS 241 Midterm study guide

10 May 2012

Contents

1 POSIX

1.1 Examples

2 System Calls

3 Libraries

3.1 strlen vs sizeof

4 Pointers

4.1 Pointer Operations

4.1.1 Pointer Arithmetic

4.2 NULL

4.3 Function Pointers

4.4 Strings

4.4.1 What's the difference between char $c[80]$ and char* c

- What about when they're used in sizeof()?

4.4.2 What's the difference between a string and a string literal?

4.4.3 How do strcpy, strcat, and strncat work?

5 Memory

5.1 Memory Fragmentation

5.1.1 Internal

Interprocesses fragmentation causes by poor "mallocing" (doesn't have malloc)

5.1.2 External

Physical to Virtual fragmentation managed by the kernel

5.2 Cache/Page Table replacment policies

5.2.1 OPT

Theortically optimal page replacement algorithm. Swaps out page used furtherst in the future. . . .this is impossible for general systems

5.2.2 LRU

Least recently used. . . .see cs232 notes

5.2.3 Working Set

The complete set of data needed to complete an operation. Often very localized in either time or space.

- Locality

  see cs232

5.2.4 Thrashing

Access pattern that assures high miss rate in the cache due to limitations of replacment algorithm

5.2.5 Belady's anomaly

Sometimes as the size of the cache increases, miss rate also increases. Proved for page faults using FIFO

5.2.6 When is a process swapped out to disk

When it is evicted by the replacment policy

5.3 Stack

5.3.1 When is the stack full?

When it meets the heap in the address space

5.4 Heap

Dynamic memory

5.4.1 Page faults

- SEGFAULT

5.4.2 How does malloc and free work?

- Memory Selection Algorithms
  - Implicit Free Lists
  - Explicit Free Lists
  - Segregated Free Lists
  - Buddy System

5.5 Virtual Memory and Paging/Segmentation

5.5.1 Virtual vs Physical memory

- Advantages of virtual memory

5.5.2 Paging vs Segmentation

5.5.3 Virtual to Physical address translation in multi-level page tables

- MMU
  - Example: How does the virtual memory subsystem know the exact location where a particular page is stored on disk, if it is swapped out of memory?
- TLB

- Algorithm for address translation goes here

  - Example: Assuming a 32-bit address space and 4 KB pages, what is the virtual page # and offset for virtual address 0xd34f6a5?

  - Example: Suppose we have a 64-bit address space and 16 KB pages. How big is the page table of a single process? What if it was multi-level?

5.5.4 Advantages of multi-level page tables

5.5.5 Determining optimal page size

5.5.6 Calculating the number of pages per page table

6 Threads and Processes

6.1 Process

6.1.1 Creating a process using fork()

- Starts new process with an incremented PC count

6.1.2 exec()

- Example: Explain how a shell process can execute a different program.

6.1.3 Orphans and Zombies

6.2 Threads

6.2.1 Shared Resources

6.2.2 Creating a thread using $\text{pthread}_{\text{create}}$()

6.2.3 $\text{pthread}_{\text{detach}}$() and $\text{pthread}_{\text{join}}$

- Example: Explain how one process can wait on the return value of another process.

6.2.4 Exiting a thread with out a thread library exit call

- How it happens: calling exit(), return, or termination

- Problems

6.2.5 What are the maximum number of threads that can be run concurrently? How is this number determined?

6.3 Context Switching

6.3.1 In Processes

6.3.2 In Threads

6.3.3 Kernel-Space vs User-Space thread managment

6.4 Memory Consistency

6.4.1 Shared memory

- Example: X is a global variable and initially X=0. What are the possible values for X after two threads both try to increment X?

6.4.2 Locking, Blocking, and Semaphores

- Mutual exclusion

- Semaphore and mutex

- Designing a a lock system for concurrent programming

6.4.3 POSIX wait()

7 Scheduling

7.1 Five state model: started ,running, ready, blocked, terminated

7.2 Scheduling schemes

7.2.1 Wait Time

7.2.2 Turnaround time

7.2.3 Response time

7.2.4 Preempting

7.2.5 Quanta

7.2.6 Fairness, progress guarentees, and interactive systems

7.2.7 Schemes

- Round Robin
  - Quanta length vs performance
- First Come First Serve (FCFS)
- Pre-emptive SJF
- Non-preemptive
  - Smallest Initial response time?
  - Smallest Initial wait?
  - Smallest Initial turnaround time?

- smalled average wait time?

- longest average wait time?

7.3 Execution Order

7.4 Starvation

7.5 Blocking

7.6 Signals and Interrupts

7.6.1 Explain how re-entrant functions are used in C.

7.7 Convoy Effect

The slow down of traffic due to queuing slowing down the whole system

Final

8 C Programming

1. What is POSIX?
2. What is a library function? What is a system call? What is the difference? Given an example of a pure library function and a pure system call.
3. How does pointer arithmetic work?
4. What is the * operator? What does it do? What is the & operator? What does it do? What is a function pointer? How do you define a function pointer?
5. What functions have you learned about in CS 241 that take a function pointer as a parameter?
6. What is a "C string"? How is a "C string" represented in memory?
7. What is NULL?
8. What is the difference between strlen() and sizeof()?
9. What's the difference between a stack and a heap variable? What about global and static variables?

10. How do malloc() and free() work?
11. What's the difference between char c[1] and char *c? ...what about when they're used in sizeof()?
12. What is the difference between a string and a string literal?
13. How do strcpy(), strcat(), strncpy(), and strncat() work?
14. How do printf() and scanf() work? What are the common formatting arguments?
15. How do you read a series of lines from a file or stdin using fgets()?

9 Memory

1. What is the difference between physical and virtual memory?
2. What are common memory allocation algorithms and what are the advantages of each?
3. How are virtual addresses translated to physical addresses in multi-level page tables?
4. How do page size and the number of levels of page tables affect the number of entries in a page table?
5. What is the difference between internal and external fragmentation?
6. What are the different page replacement policies and the advantages of each?
7. Describe how the buddy system works and the run time for its operations.
8. What is thrashing? When does it occur?
9. What causes a SEGFAULT and what happens when one occurs?
10. When is a process swapped out to disk?
11. What is the difference between the MMU and the TLB? Describe the function of each.

[1]FOOTNOTE DEFINITION NOT FOUND: 80

12. Name three benefits of virtual memory (as opposed to allowing programs to directly access physical memory).

13. Name one advantage of segmentation over paging, and one advantage of paging over segmentation.

14. How is a page table similar to an inode? What is the difference between these structures?

15. Assuming a 32-bit address space and 4 KB pages, what is the virtual page # and offset for virtual address 0xd34f6a5?

16. Give an example of a page fault that is an error, and an example of a page fault that is not an error.

17. Assume LRU page eviction and three pages of physical memory. Describe what happens

18. when the application accesses virtual memory pages in this sequence: 3,4,5,4,1,6,9,3,9,8,4,8,8,2.

19. How many page faults occur in the above example?

20. Why are pages set to read-only in the copy-on-write technique?

21. Suppose we have a 64-bit address space and 16 KB pages. How big is the page table of a single process, if the system uses single-level page tables? What is the problem here? How would multi-level page tables help solve this problem?

22. Which page replacement scheme is better, OPT or LRU? Why?

23. Why does LRU not suffer from Belady's anomaly?

24. How does the virtual memory subsystem know the exact location where a particular page is stored on disk, if it is swapped out of memory?

25. How is the working set computed? How is the notion of a working set useful for managing memory of processes?

26. Compare and contrast (give one benefit and one disadvantage) for: implicit, explicit, segregated, and buddy free lists.

10 Processes and Threads

1. What resources are shared between threads of the same process?
2. Invent some code using $\text{pthread}_{\text{create}}()$ statements. What could be its output?
3. What are the possible values for X after both threads complete execution? (X is a global variable and initially X = 0.)
4. What happens when a thread calls exit()?
5. What happens to a process's resources when it terminates normally?
6. Describe what happens when a process calls fork(). Be able to trace through the code.
7. Under what conditions would a process exit normally?
8. Explain the actions needed to perform a process context switch.
9. Explain the actions needed to perform a thread process switch.
10. What are the advantages and disadvantages of kernel-level threads over user-level threads?
11. Compare the use of fork() to the use of $\text{pthread}_{\text{create}}()$.
12. In a multiprocessor system, what system characteristics will cause other threads of the same process to block?
13. How can a process become orphaned and what does the OS do with it? What's a zombie?
14. Write a piece of code using fork() to create a process tree of depth n, where each process (a node in the tree) except for the "leaf" processes has exactly m child processes.
15. Describe how to use the POSIX call wait().
16. Explain what happens when a process calls exec().
17. Explain how reentrant functions are used in C.
18. What are the maximum number of threads that can be run concurrently? How is this number determined?

19. If a process spawns a number of threads, in what order will these threads run?

20. Explain how to use $\text{pthread}_{\text{detach}}()$ and $\text{pthread}_{\text{join}}()$.

21. Explain how a shell process can execute a different program without using system().

22. Explain how one process can wait on the return value of another process.

23. Describe the transitions between running, ready and blocked in the 5 state model.

24. Understand how $\text{pthread}_{\text{exit}}()$ differs from exit().

11 Scheduling

1. What is starvation? Which scheduling policies have the possibility of resulting in starvation?

2. Which scheduling algorithm results the smallest average wait time?

3. What scheduling algorithm has the longest average response time?

4. Define turnaround time, waiting time and response time in the context of scheduling algorithms.

5. What is the convoy effect?

6. Why do processes need to be scheduled?

7. How does bounded wait apply to scheduling?

8. Which scheduling algorithm minimizes average initial response time? Waiting time? Total response time?

9. Why is SJF/PSJF hard to implement in real systems?

10. What does it mean to preempt a process?

11. What does it mean for a scheduling algorithm to be preemptive?

12. Describe Round-Robin scheduling and its performance advantages and disadvantages.

13. Describe the First Come First Serve (FCFS) scheduling algorithm. Explain the performance advantages and disadvantages.

14. Describe the Pre-emptive and Non-preemptive SJF scheduling algorithms. Explain the performance advantages and disadvantages.

15. Describe the Preemptive Priority-based scheduling algorithm. Explain the performance advantages and disadvantages.

16. How does the length of the time quantum affect Round-Robin scheduling?

17. Define fairness in terms of scheduling algorithms. What are the fairness properties of each of the scheduling disciplines discussed in class?

18. Which scheduling algorithms guarantee progress?

19. A process was switched from running to ready state. Describe the characteristics of the scheduling algorithm being used.

20. Which properties of scheduling algorithms affect the performance of interactive systems?

12 Synchronization

1. What is the readers-writers problem?

2. What is the producers-consumers problem?

3. What is the dining philosopher problem?

4. Recognize a correct solution to the readers-writers problem, the producers-consumers problem, and the dining philosopher. Be able to identify and explain an error in a specific implementation of any of the classic synchronization problems.

5. What happens when readers are prioritized over writers in the classic "readers writer problem"? How about if writers are prioritized over readers?

6. What is required so that deadlock and starvation do not occur in the dining philosopher's problem? Give examples of solutions.

7. What is the difference between starvation, deadlock, race conditions and critical sections? Describe each.

8. What would happen if a system's hardware synchronization primitive were replaced with a software function?

9. Which type of variables must be protected against concurrent readers and writers in any combination?

10. Given two threads running example code that contains a critical section, be able to identify if progress and mutual exclusion are ensured.

13 Mutexes and Semaphores

1. Understand the common semaphore and mutex functions ($\text{sem}_{\text{wait}}()$, $\text{sem}_{\text{post}}()$, etc).

2. How does a semaphore perform, in both the parent and child, after a fork()?

3. What are proper and improper code replacements for a $\text{test}_{\text{and}}\text{set}()$ operation?

4. How does the internal counter of a POSIX semaphore work? What does it mean if the value of the semaphore is 1?

5. How can the reader-writer problem be solved using only POSIX mutexes?

6. Using only one mutex, is it possible to create a semaphore? If so, how? If not, why?

7. Understand how to solve the producer-consumer problem using mutexes and semaphores.

8. What is a buffer overflow? What is a buffer underflow? Understand how failures in synchronization could cause buffer over and underflows.

9. What is progress?

10. What is mutual exclusion?

11. What are condition variables? Understand how they can be used in code.

12. Understand how to fix deadlocks and starvation in code involving mutexes, semaphores, and conditional variables.

14 Processes and Deadlock

1. Define deadlock.
2. Define circular wait, mutual exclusion, hold and wait, and no preemption. How are these related to deadlock?
3. How would the implementation of a web server using threads differ from one using processes?
4. What can happen if synchronization in a multiple-threaded program is not programmed carefully?
5. Why might an operating system use a resource allocation graph?
6. What are the conditions of a deadlock? How could you guarantee that each one of these conditions can be prevented?
7. What does waitpid() do?
8. What are the approaches for solving deadlock?
9. What is the difference between Deadlock Prevention, Deadlock Detection & Recovery, and Deadlock Avoidance? What deadlock handling mechanism would you use?
10. What are the components of a resource allocation graph?
11. What problem does the Banker's Algorithm solve? Given a set of processes how would you use the Banker's Algorithm?
12. What is a safe state and how can you determine if a system is in a safe state?

15 IPC

1. What is the difference between a FIFO and a pipe?
2. How would you redirect standard out to a file?

3. What is the difference between a pipe, a FIFO, and an ordinary file on disk?
4. What happens when two processes read and write to a memory mapped file?
5. Explain how two processes can share memory using shmem.
6. Explain how a process can set which signals are caught or ignored using a signal set.
7. How can one process send a signal to another?
8. Describe the purpose of a POSIX signal.
9. Some signals cannot caught or ignored. Which signals are they and why shouldn't they be allowed to be caught?
10. What does "kill -<parameter> pid" do?
11. How is the function sigwait() used?
12. How does the function alarm() work?

16 Networking

1. When do you use the close() system call with sockets?
2. Discuss how a multithreaded web server running on a single processor system could be optimized using the process scheduling methods discussed in class. Which do you recommend?
3. How do select() and poll() work? What problem do they it solve?
4. Describe the Posix accept() function.
5. How does HTTP work?
6. Describe the services provided by TCP.
7. How does TCP connection establishment work?
8. Describe the services provided by UDP.
9. Explain the difference between a regular and a connected UDP socket.

10. How do sockets support the client-server model?
11. Which is better, UDP or TCP? Which one would you use?
12. How does the Domain Name System (DNS) work?
13. How does DNS use caching?
14. How is DNS related to IP?

17 File systems and I/O

1. Given a description of the block size and i-node structure, what is the maximum size of a file?
2. How many i-node operations are required to fetch a file at /path/to/file?
3. What information is stored in an i-node? What information isn't?
4. What data structure best describes an i-node?
5. What are the advantages and disadvantages of an i-node based file system?
6. Given the description of an i-node file system, how many i-node accesses are required to read the entire contents of a file of a given size? How many blocks does this file consume on disk?
7. What is an advantage of a soft link over a hard link?
8. What is I/O polling? What are advantages and disadvantages?
9. Describe disk I/O access using DMA.
10. How are file descriptors shared between threads in a single process? How are they shared between after a process executes a fork()?
11. When the size of a block changes in an i-node based file system, how does this change the maximum size of a file?
12. How do polling and interrupt driven I/O differ? What are the advantages and disadvantages of each?
13. How does the page-out process work?

14. Understand how hard-links result in different file names affecting the same i-node.

15. If an i-node based file system has a certain number of direct and single-indirect blocks, how large is the file?

16. Where does fstat() look to find the information that it returns?

17. How does a file system use caching?