Chalmers Un. of Technology and Gothenburg Un. Comp. Science and Engineering Department

Operating Systems
DIT 400, EDA092

Exam 2010-03-09

Date, Time, Place: Tuesday 9/3 2010, 8:30-12:30, M building

Course Responsible: Arne Dahlberg, Marina Papatriantafilou (Tel: 772 1705, 772 5413)

Auxiliary material: You may have with you
- An English-Swedish, Swedish-English dictionary.
- No other books, notes, calculators, PDA's etc.

Grade-scale ("Betygsgränser"):
CTH:3:a 30-38 p, 4:a 39-47 p, 5:a 48-60 p
GU: Godkänd 30-47p, Väl godkänd 48-60 p

Instructions

- Do not forget to write your personal number, if you are a GU or CTH student and at which program ("linje").
- Start answering each assignment on a new page; number the pages and use only one side of each sheet of paper.
- Write in a **clear manner** and **motivate** (explain, justify) your answers. If it is not clear what is written, your answer will be considered wrong. If it is not explained/justified, even a correct answer will get **significantly** lower (possibly zero) marking.
- If you make any assumptions in answering any item, do not forget to clearly state what you assume.
- The exam is organized in groups of questions. The credit for each group of questions is mentioned in the beginning of the respective group. Unless otherwise stated, all questions in a group have equal weight.
- Please answer in English, if possible. If you have large difficulty with that and you think that your grade can be affected, feel free to write in Swedish.

Good luck !!!!

1. (10 p)

   (a) What are the two most important reasons to use RAID systems? (2p)

   (b) What is the most important difference between RAID 5 and RAID 6? (1p)

   (c) Mention two types of errors that RAID do not protect against. (1p)

   (d) What is a disk controller? (1p)

   (e) A disk controller usually have a large RAM memory. Give two possible uses for this RAM memory. (2p)

   (f) Explain the meaning of memory mapped files. (1p)

   (g) Explain the main steps needed for copying a file using memory mapped files. (2p)

2. (10 p)

   (a) A method to keep track of used blocks in a filesystem is FAT (File Allocation Table). Describe how the FAT method works. (3p)

   (b) In what type of memories are FAT in common use today? (1p)

   (c) File system operations in UNIX use a VFS (Virtual File System) layer for calling the actual file system code. Describe how the VFS interface works. (2p)

   (d) How is a file system directory implemented in a UNIX file system? (1p)

   (e) Why do UNIX hard links not generate cycles in the file system graph? (1p)

   (f) In almost all operating systems the first block in a disk partition is reserved for a special purpose. Which purpose? And is it necessary to reserve the block for this purpose? (2p)

3. (10 p)

   (a) The Andrew filesystem uses a consistency semantics called *Session Semantics*. Explain how *Session Semantics* works. 1(p)

   (b) Why is *location independent* internal file identifiers used in the Andrew filesystem? (1p)

   (c) Give one advantage and one disadvantage with using a stateless file server. (2p)

   (d) Explain the difference between a type 1 hypervisor and a type 2 hypervisor. (1p)

   (e) Give two examples of why a programmer might want to run an operating system in a virtual machine. (2p)

   (f) Explain why virtualization complicates the handling of virtual memory and how the problem is usually solved. (3p)

4. (10 p)

   (a) A computer has a cache, main memory, and a disk used for virtual memory. If a referenced word is in the cache, 20 ns are required to access it. If it is in main memory but not in the cache, 60 ns are needed to load it into the cache (this includes the time to originally check the cache), and then the reference is started again. If the word is not in main memory, 12 ms are required to fetch the word from disk, followed by 60 ns to copy it to the cache, and then the reference is started again. The cache hit ratio is 0.9 and the main-memory hit ratio is 0.6. What is the expected time in ns required to access a word on this system? Explain how it is computed. (3p)

   (b) What is the cause of thrashing? How can thrashing be detected and what can be done to eliminate the problem? (3p)

   (c) Why is the Least Recently Used (LRU) algorithm for page replacement not conceivable to implement without explicit hardware support? Describe a method that approximates LRU replacement. (4p)

5. (10 p)

   (a) Can a multithreaded program using user-level threads achieve better performance on a multiprocessor system than on a single-processor one? Explain your answer (2p)

   (b) Consider two periodic real-time processes, P1 and P2, with periods and processing times p1 = 50, t1 = 25 and p2 = 75, t2 = 30. Can they be scheduled using Earliest Deadline First scheduling without missing their deadlines? Explain why or why not. If yes, show also the schedule; if not, explain also whether they can be scheduled using another method. (4p)

   (c) Consider a system running 10 IO-bound tasks and one CPU-bound task. Each of these IO-bound tasks issues an IO operation once every msec of CPU computation and each IO operation takes 10 msec to complete. The context-switching takes 0.1 msec. Assuming that all processes are long-running tasks, describe the CPU utilization (fraction of time that the CPU spends on computation over the sum of computation and context-switching times), when the time-quantum is 1 msec and when it is 10 msec. (4p)

6. (10 p)

   (a) Why do Solaris, Linux and Windows XP use spin-locks as a synchronization mechanism in multiprocessor systems and not on single-processor systems? (2p)

   (b) Consider a system consisting of m resources of the same type being shared by n processes. Resources can be requested and released by processes only one at a time. Show that the system is deadlock free if the following two conditions hold: a) The maximum need of each process is between 1 and m resources. b) The sum of all maximum needs is less than m+ n. (4p)

   (c) Consider a server designed with a limit N on the number of open connections it is allowed to maintain. Explain how semaphores can be used to limit the number of concurrent connections. Explain your solution carefully and argue for its correctness, liveness (deadlock-freedom) and fairness properties. (4p)