Disable all interrupts [KERNEL ONLY or NOT KERNEL ONLY] => KERNEL. Prevents hardware events from being serviced in a timely manner causing devices to malfunction. With preemption, the scheduler is not guaranteed to be notified at and-of-quantum. Read the time-of-day clock [KERNEL ONLY or NOT KERNEL ONLY] => NOT KERNEL. Reading a value does not change/expose kernel state, other processes’ state, or change external hardware state. Set the time-of-day clock [KERNEL ONLY or NOT KERNEL ONLY] => KERNEL. The clock is shared and no longer has a reliable value. No computation (like timers) is reliable. Change the memory map [KERNEL ONLY or NOT KERNEL ONLY] => KERNEL. Changing the may give access to the kernel memory or other processes’ memory exposing data or changing data inappropriately. Synchronize Multi-processor memory accesses [KERNEL ONLY or NOT KERNEL ONLY] => NOT KERNEL. This takes some possible race conditions and makes them predictable. Used for spinlocks or other synchronization. Note, however, it may have a performance penalty, but cannot cause any incorrect behavior by kernel or other processes. Interrupt: Event generated by H/W external to the processor. Causes “interrupt behavior” where CPU is put into kernel mode and a jump is made to a specific location. Exception: Software generated event. Causes “interrupt behavior” Trap: “expected” exception in correct program flow. Usually used to request kernel services Fault: “unexpected” exception, may or may not result from correct flow. Sometimes (page fault) the state of the CPU/memory can be adjusted in order to let the instruction be restarted. Other times this causes the process to be terminated.

Must all the General Purpose Registers be saved when handling an interrupt? Why or why not. => Interrupts are hardware generated events, like device completion or timer expiration. When an interrupt occurs, the kernel will perform some operations and then restart the thread code at the point where the interrupt occurred. If the kernel did not save the registers, it would be impossible to restart the thread correctly. In essence, interrupts are “invisible”: they happen but they do not influence the behavior of the program.

Define monolithic kernel, compare to microkernel: => A monolithic kernel essentially is designed to have all the system services encapsulated in one module (image) with a tight communication and sharing of structures between the individual OS components. Microkernel which takes the different tact of providing a set of minimal functions in the kernel and implement higher level OS features in other modules (running either in user mode or kernel mode). Monolithic kernel if it can be made bug free can on average perform better than a microkernel system that has some additional communication overhead between components.

Draw the complete state diagram for kernel threads. Label each state, transition and describe what actions transpire for each transition. (Hint: start with a NEW state and a TERMINATED state with three other states in-between)

NEW 🡪 READY 🡪 RUNNING 🡪 TERMINATED  
 🡩 🡨 🡫  
 🡩 🡫   
 🡩 🡫   
 WAITING  
NEW 🡪 READY: processes creation  
READY 🡪 RUNNING: Processing scheduled to execute  
RUNNING 🡪 WAITING: Waiting for something to happen  
RUNNING 🡪 READY: scheduling algorithm preempts process time slice  
RUNNING 🡪 TERMINATED: Exit  
WAITING 🡪 READY: Whatever the process was waiting for has completed and now the processes is ready to run again.

S ≤

Compare the advantages and disadvantage of preemptive over a non-preemptive scheduling strategy? => The advantage of a preemptive scheduler is that it allows for more equitable sharing and response time when running multiple processes/threads. A preemptive scheduler also will not allow a CPU bound or runaway thread to essentially shut down the system. << >> A disadvantage of a preemptive scheduler is that there is more bookkeeping needed to handle interrupting the thread because the thread is no longer reentering the system (via a system) at well-defined points.

In most systems for programming considerations the first page in a processes virtual address is usually marked as invalid. Why should this be true? (Hint: think pointers) And why isn’t this a complete solution to the problem it’s trying to solve? => The first page is often marked as invalid to catch cases where a programmer has incorrectly coded a dereference of a NULL pointer. This is an incomplete solution for all invalid pointer accesses since many other pages may be valid but for which there should be no valid pointer in the program. It is also incomplete when the dereference through NULL occurs with a large offset (say, a big field offset in a struct).

**●● \_F\_** The size in bytes of a variable of type (char \*) is 1 byte. **\_T\_** In a C program the expressions a[i] and \*(a + i) are equivalent. **\_T\_** The strdup function calls malloc. **\_F\_** On a 1-CPU computer, a program that runs in time T will run in time T/n if decomposed in n threads. **\_F\_** The "S" in SMP stands for Simultaneous. **\_F\_** A call to "strcmp" will show in the truss output. **\_T\_** The time command in mentor could show that the user time is larger than the real time. **\_T\_** A long time quantum may cause a program to finish sooner. **\_T\_** The arguments of a system call are checked in kernel mode. **\_F\_** The file descriptors of a process are closed when a process calls execvp(). **\_T\_** A process that uses pipes may hang due to unclosed file descriptors. **\_T\_** A program that runs with non-preemptive scheduling runs faster than one with preemptive scheduling. **\_F\_** Most of the processes' CPU bursts do not finish before a context switch. **\_F\_** Programs that run round-robin scheduling have faster average response time than programs that run SJF. **\_F\_** When a process calls fork, the number of open file objects in the kernel is duplicated. **\_F\_** POSIX threads are better than Solaris threads because the former are faster. **\_T\_** The input/output redirection to files can be done by the child. **\_T\_** Kernel threads in a process share the same file descriptors. **\_F\_** A section of code that is guarded by sema\_wait/sema\_post calls can be executed by only one thread at a time. **\_F\_** A process table entry contains one set of registers for each user and kernel level thread in a process. **\_T\_** The two primary purposes of an operating system are to manage the resources of the computer and to provide a convenient interface to the hardware for programmers. **\_F\_** When the CPU is in kernel mode, the CPU is prevented from accessing areas of memory that are not owned by the current process. **\_T\_** An interrupt table contains the addresses of the handlers for the various interrupts. **\_F\_** Each thread of a process has its own virtual address space. **\_F\_** Every time a clock interrupt occurs, a context switch from one process to another is performed. **\_F\_** In a virtual memory system, a virtual address and a physical address must be the same size. **\_T\_** In a virtual memory system, a virtual page and a physical page frame must be the same size. **\_T\_** In a multiprogrammed system using partitioning (i.e. each process occupies a contiguous block of memory), addresses can be relocated at run time using base registers. **\_F\_** In a multiprogrammed system using partitioning, the Best Fit strategy (where a process is placed in the smallest hole in memory large enough for the process) provides the most effective use of memory. **\_T\_** The operating system kernel consists of the portion of the operating system that is always running. **\_T\_** The difference between a program and a process is that a process is an active entity, whereas a program is a passive entity. **\_T\_** System calls can be run only in kernel mode. **\_F\_** Interrupts can be triggered only by hardware. **\_T\_** In UNIX systems, the exec() system call causes the calling process to run a different program. **\_F\_** Named pipes in UNIX require a parent-child relationship between the communicating processes. **\_F\_** Concurrency means that multiple tasks can execute simultaneously if multiple cores or processors are available, whereas parallelism means that multiple tasks can achieve progress via serial execution on a single core or processor. **\_T\_** It is possible to create a thread library without any user-level support. **\_F\_** Each thread has its own register set and virtual memory space. **\_T\_** Practical solutions to the critical section problem require hardware support. **\_T\_** A microkernel is a kernel that is stripped of all nonessential core components. \_**T**\_ A system call is triggered by software. \_**F**\_ Privileged instructions can be executed directly in user mode.

●● Enumerate the fields of a process table entry: Process ID, Process state, Saved Registers, File Descriptors, Page Table Mention the checks done by the kernel during the open() system call: ● File Permissions: If the file is opened in write-mode, the user should have write permissions to the file either as user, group or others. The same for read-mode.

● If the file is opened with the flag O\_CREAT and the file does not exist, the user should have write permissions to the directory the file will be created into. What are the steps involved in a context switch? => Save registers in process-table entry ● Jump to timer interrupt-handler ● Change the state of the process from running to ready ● Choose the next process to run from the ready processes. ● Set this process in running state. ● Restore the registers of the next process in the CPU ● Return from interrupt. What are the advantages and disadvantages of using kernel threads vs. user threads? ***Disadvantages of kernel-threads:*** ● Context Switch of kernel threads is slower because it needs to switch to kernel-mode ● Programs may take longer because of the context switch overhead of preemptive scheduling. ***Advantages of kernel-threads:*** ● User-level threads use non-preemptive scheduling so one non-cooperative thread may hang others. That does not happen with kernel-threads that use preemptive scheduling. ● Kernel-level threads may use multiple processors in SMP machines. What are the advantages and disadvantages of using threads vs. using processes? ***Advantages of threads:*** ● Context switches among threads is faster than among processes. ● Thread creation is faster than process creation ***Disadvantages of threads:*** ● If one thread crashes the entire process crashes. ● Thread synchronization is necessary to prevent the corruption of shared data structures. What factors have to be considered when choosing the length of a quantum time? ● Response Time ● Context Switch Overhead ● Average CPU burst length What does the CPU hardware do when a trap or interrupt occurs? Describe the specific steps that the hardware performs (but just the hardware - not the OS or any process). When a trap or interrupt occurs, the CPU pushes the program counter and other registers (e.g. the stack pointer, status register, etc.) onto the stack, switches to kernel mode, and jumps to the address contained in the interrupt table entry corresponding to the interrupt that occurred. When a process executes a fork() system call, a duplicate process (i.e. the child process) is created. How does the code in the processes – since it is identical in both the parent and child processes – know which process is the parent and which is the child? => In the child process the fork() call returns 0, whereas in the parent process the fork() call returns the PID of the child process.

Give a simple example of code that would operate differently in the parent and in the child. => if (fork() == 0) printf("I'm the child.\n"); else printf("I'm the parent.\n");

What is the CPU scheduler? Whenever the CPU becomes idle, the OS must select one of the processes in the ready queue to be executed. The selection process is carried out by the short-term scheduler (or CPU scheduler). The scheduler selects a process from the processes in memory that are ready to execute and allocates the CPU to that process.

Describe the four conditions that CPU scheduling decisions may take place. => CPU scheduling decisions may take place when a process: (1) Switches from running to waiting state: ● The result of an I/O request ● Wait for the termination of one of the child processes (2) Switches from running to ready state: An interrupt occurs (3) Switches from waiting to ready: Completion of I/O (4) Terminates. Describe the benefits of thread pools => (1) Servicing a request with an existing thread is usually faster than waiting to create a thread. (2) A thread pool limits the number of threads that exist at any one point. This is particular important on systems that cannot support a large number of concurrent threads. Describe four benefits of using multithreading => Responsiveness, resource sharing, economy, and utilization of MP Architectures. In the original UNIX operating system, a process executing in kernel mode may not be pre-empted. Explain why this makes (unmodified) UNIX unsuitable for real-time applications. => In a real-time system, the OS must always be able to pre-empt a lower-priority process in order to allow a higher-priority process to run. This is not possible in unmodified UNIX if the lower-priority process is executing a system call when the higher-priority process needs to run. Assume process P1 has threads T1 and T2. Will T1 and T2 continue to run after P1 exits? => No - because T1 and T2 are part of the address space of P1, the threads have no independent existence and will “disappear” when P1 exits. (This contrasts with child processes, which do have an independent existence from the parent process.) Give a reason why a context switch between threads may be cheaper than a context switch between processes. => A process has a much larger “state” than a thread, because a thread shares much of its state with the process that created it (e.g., virtual address space, file descriptor table, etc.). Thus, a switch between two threads within a process typically will require less time to save/restore state than a switch between two processes.

  

