CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

Note:

1. **Write your name and GT number on each page.**
2. The test is **CLOSED BOOK** and **NOTES**.
3. Please provide the answers in the space provided. You can use scratch paper (provided by us) to figure things out (if needed) but you get credit **only** for what you put down in the space provided for each answer.
4. For conceptual questions, **concise bullets** (**not wordy sentences**) are preferred.
5. Where appropriate use figures to convey your points (a figure is worth a thousand words!)
6. **Illegible answers are wrong answers.**
7. **DON'T GET STUCK ON ANY SINGLE QUESTION…FIRST PASS: ANSWER QUESTIONS YOU CAN WITHOUT MUCH THINK TIME; SECOND PASS: DO THE REST.**

Good luck!

Question number		Points earned	Running total
1 (0 min)	(Max: 1 pts)		
2 (12 min)	(Max: 30 pts)		
3 (12 min)	(Max: 20 pts)		
4 (12 min)	(Max: 29 pts)		
5 (12 min)	(Max: 20 pts)		
Total (48 min)	(Max: 100 pts)		

1. (0 min, 1 point)(circle the right choice)
Ellen DeGeneres used the following for the "Selfie" during OSCAR 2014 (don't worry…you get a point regardless of your answer)
   a. Fujitsu Lifebook
   b. iPhone
   c. Samsung Galaxy
   d. Blackberry
   e. BBN Butterfly
   f. None of the above
   g. I will pretend I don't know the right answer…!

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

OS Structures

2. (**12 min**, 30 points)

(a) (6 points)

Based on a reading of the SPIN and Exokernel papers, select the features best supported by each type of OS below

(NOTE: +1 for each right entry; -2 for each wrong entry; MAX 6; MIN 0)

Feature	Monolithic OS	DOS-like OS	Microkernel OS
Extensibility		X	X
Protection	X		X
Performance	X	X	

(b) (5 points)

What is the difference between "protection domain" and "hardware address space"? (Concise bullets please)

- **Protection domain is a logical concept that signifies the reach of a particular subsystem. For example, each specific service of an OS such as the memory manager, the file system, the network protocol stack, and the scheduler may each live in their own protection domains and interact with one another through well-defined interfaces implemented using IPC mechanism (*a la* L3 microkernel).**

  (+2.5 if first sentence is mentioned; 0 otherwise)

- **Hardware address space is a physical concept and pertains to the ability of the hardware to restrict the memory reach of the currently executing process. The OS can use this hardware mechanism to implement logical protection domains. A specific implementation of an OS may choose to pack multiple protection domains in the same hardware address space. Another may choose to associate a distinct hardware address space to each protection domain.**

  (+2.5 if first sentence mentioned; 0 otherwise)

(c) (5 points)

Consider an architecture that gives segmentation support in hardware. How will you implement small protection domains in such an architecture?

- **The entire hardware virtual address space carved into distinct non-overlapping regions. (+2)**
- **Each protection domain assigned to one of these distinct regions. (+1)**
- **The hardware segment register is loaded with the start address and length of each logical protection domain at the point of entry into that protection domain by the operating system. (+2)**

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

(d) (5 points)
What is the purpose of the "software TLB" in the Exokernel approach to providing extensibility?

- **All library OSes live above the kernel. Thus each library OS is in its own hardware address space. (+1)**
- **When a library OS is scheduled to run on the processor by Exokernel, the (hardware) TLB has to be flushed. (+1)**
- **Thus the newly scheduled library OS will have an inordinate amount of TLB misses. (+1)**
- **To mitigate the associated loss of performance, Exokernel maintains a software TLB data structure for each library OS. The translations contained in this data structure are loaded into the hardware TLB. (+2)**

(e) (5 points)(Answer **True/False** with **justification**)
"SPIN's approach of extending logical protection domains and Exokernel's approach of downloading code into the kernel are one and the same."

False. (+1)

- **Functionally, they accomplish the same thing, namely extend the kernel with additional functionality. (+2)**

- **However, in Exokernel once the code is downloaded into the kernel there is no protection against malicious or erroneous behavior of the downloaded code. (+1)**

- **In SPIN, since the extension via logical protection domains is achieved under the protection of the strong semantics of Modula-3, the other subsystems that live within the same hardware address space are protected from malicious or erroneous behavior of any given logical protection domain. (+1)**

(f) (4 points)
Enumerate the myths about a microkernel-based approach that are debunked by Liedtke in the L3 microkernel. (No need to say **how** he debunks it).

- **User/Kernel border crossing is expensive**
- **Address space switches are expensive**
- **Thread switches and IPC are expensive**
- **Locality loss (memory/cache effects) due to context switches when protection domains are implemented as server processes**

(+1 for each of the above points)

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

Virtualization

3. (**12 min**, 20 points)

(a) (12 points) (Full virtualization)

In the Intel architecture, the processor does the following on every memory access:

- Translates virtual page number to a physical page frame number using a page table in physical memory. Let's call this hardware page table (**HPT**).
- The processor also has a hardware TLB to cache recent address translations.

Full virtualization is implemented on top of this processor. You have to answer the following questions in the context of full virtualization.

(i) (2 points)

What is the shadow page table (**SPT**), and who maintains it?

- **Shadow page table gives the mapping between physical pages (which are deemed contiguous in physical memory so far as the guest OS is concerned), and the machine pages (i.e., the real physical memory in DRAM which is under the control of the hypervisor). (+1)**
- **It is maintained by the hypervisor (or virtual machine monitor) (+1)**

(ii) (2 points)

How many SPTs are there? Why?

- **As many as the number of guest OSes currently executing on top of the hypervisor. (+1)**
- **Each guest OS has its own illusion of a contiguous physical memory, which has to supported by the hypervisor by having a distinct SPT for each virtual machine. (+1)**

(iii) (4 points)

Each process created by a guest OS has its own page table. Let's call this physical page table (**PPT**). What is the relationship between PPT, SPT, and HPT?

- **Each process in a guest OS has its own PPT. The PPT contains the VPN to PPN mapping assigned by the guest OS. (+1)**
- **Hypervisor allocates the real machine pages to back the physical pages of each guest OS. SPT contains the mapping between PPN and the MPN. (+1)**
- **HPT is what the processor uses to translate the VA of the currently running process to the real physical address. (+2)**

CS 6210 Spring 2014 Midterm Solution

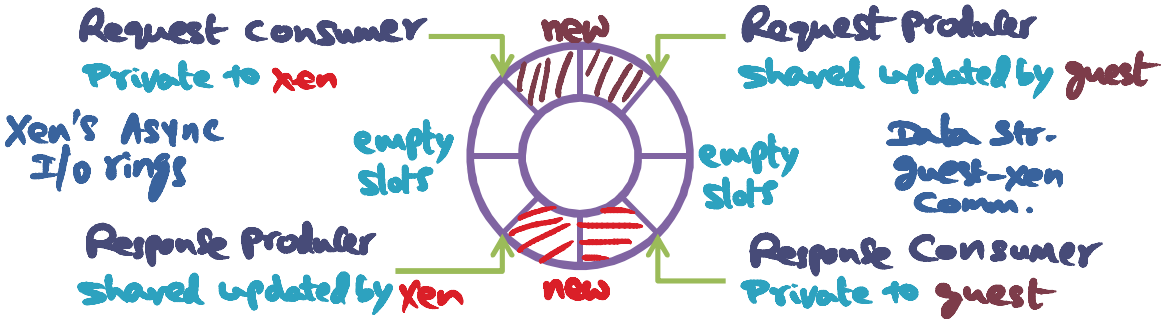
Name:_____Kishore_____GT Number:

(iv) (4 points)
How is page fault by a process handled with this set up? (Concise bullets please)

- **Hypervisor catches the page fault.** **(+1)**
- **Passes the faulting VA to the currently executing guest OS.** **(+1)**
- **Guest OS services the page fault.** **(+1)**
- **Hypervisor traps the PT/TLB updates of the guest OS to directly enter the VPN to MPN mapping into the HPT.** **(+1)**

(b) (8 points) (Para virtualization)
A process in a guest OS makes a request to read a file from the disk. Using figures, explain the steps taken by the guest OS and the Xen hypervisor using I/O rings to fulfill the guest OS's request.

- **The file system of the guest OS translates the file name to a disk block address and passes it to the disk subsystem of the guest OS.**
- **The disk subsystem shares an I/O ring data structure with Xen.**
- **Guest OS (i.e., the disk subsystem) uses an available slot in the I/O ring to enqueue the disk read request. It embeds a physical page frame pointer in this descriptor.** **(+1)**
- **Hypervisor performs the disk I/O using DMA directly transferring the disk block into this physical page frame.** **(+1)**
- **Hypervisor enqueues a response using an available slot in the I/O ring data structure.** **(+1)**



(+5 for a nice descriptive figure)

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

Synchronization, Communication, and Scheduling in Parallel Systems

4. (**12 mins**, 29 points)

(a) (2 points)

Explain sequential consistency (Figures and/or concise bullets please).

- **Respect program order of memory access from each processor (+1)**
- **Arbitrary interleaving of the memory accesses from different processors that respects the program order (+1)**

(b) (3 points) (MCS)

Given the attributes:

(I) unfair
(II) fair but with contention on lock release
(III) fair and no contention on lock release

Associate the attribute (Roman numeral) with the correct spinlock algorithms below.

MCS lock _____III_____

T&S with delay _____I_____

Ticket lock _____II_____

(c) (5 points)(Explain why the following statement is true)(MCS)

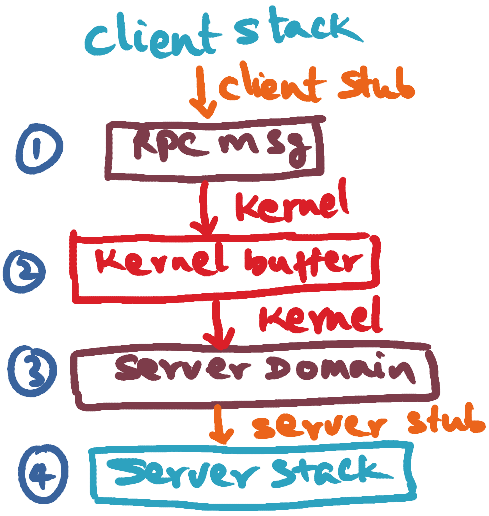
"Dissemination barrier is bound to do poorly on a bus-based shared memory multiprocessor compared to the MCS Tree barrier algorithm."

- **The algorithm has O(N) communication in each round. The communications in each round are independent and so far as the algorithm is concerned they can all go in parallel. (+1)**
- **In a bus-based shared memory multiprocessor, these O(N) communication events will get sequentialized. (+1)**
- **The total communication in dissemination barrier is O(N $\log_2$N) whereas it is O(N) for the MCS barrier. Thus dissemination barrier is more adversely affected by the serialization of bus requests. (+3)**

CS 6210 Spring 2014 Midterm Solution

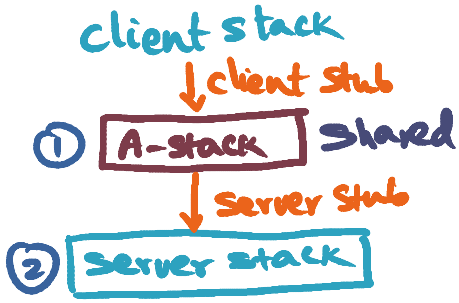
Name:_____Kishore_____GT Number:

(d) (4 points)(LRPC)
Without the LRPC optimization, there are 4 copies before the server procedure can start executing. Enumerate them with concise bullets or a diagram.



+1 for each copy

(e) (2 points)(LRPC)
How does LRPC reduce the number of copies from 4 to 2? (Concise bullets please)



+1 for each copy

(f) (4 points)(multiprocessor scheduling issues)
A process becomes runnable. Give two factors that are important to take into consideration in assigning it to one of the processors in a shared memory multiprocessor? (Concise bullets please)

- **Cache affinity; in particular choose the processor on which there has been a minimum number of "other processes" scheduled since the last time this process ran on that processor. (+2)**
- **Queue length; consider if the processor already has "other" processes already in its local queue before this process will get to run. (+2)**
- **Ensure working sets of all collocated processes including the newly runnable one fit in the last level cache of a multicore processor. (+2)**

(any two of the above three for full credit)

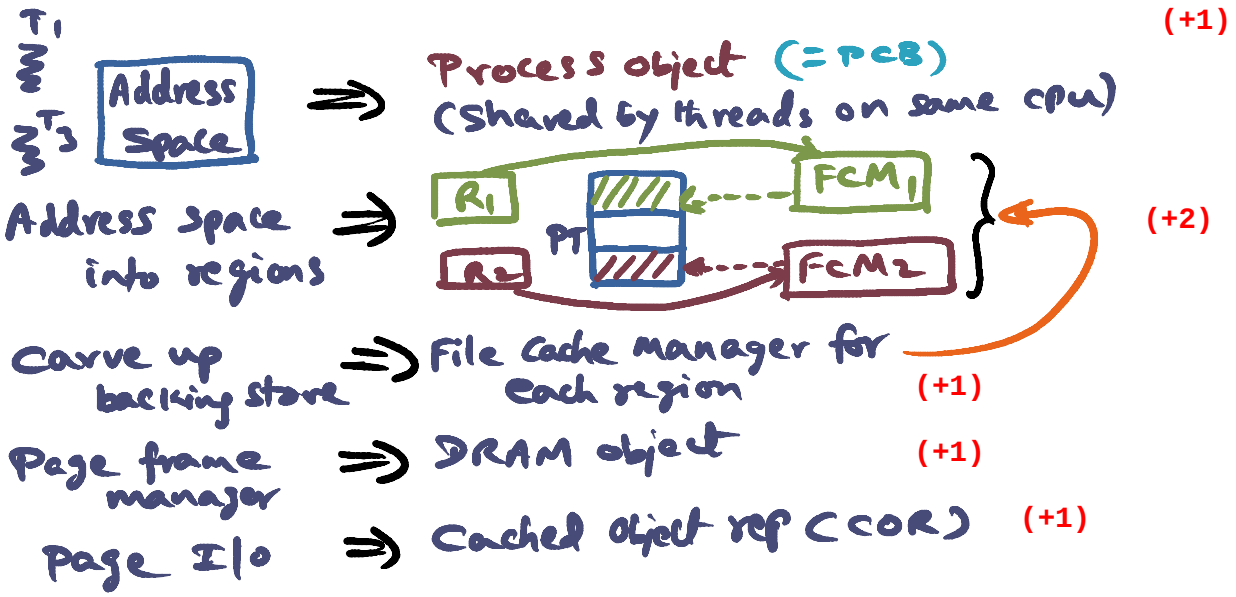
CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

(g) (3 points)(Explain why the following statement is true)(Tornado)
"Dealing with concurrent page faults experienced by the threads of a multi-threaded process is a challenge for a parallel OS."
(Concise bullets please).

- **A multithreaded process has a page table which is logically shared by all the threads.** **(+1)**
- **Concurrent page faults happening on different processors where the threads are executing could potentially end up getting serialized if careful attention is not paid to managing the shared page table.** **(+2)**

(h) (6 points)(Tornado)
How does Tornado address this challenge using the clustered object abstraction? (Use figures and concise bullets please)



(+1)
(+2)
(+1)
(+1)
(+1)

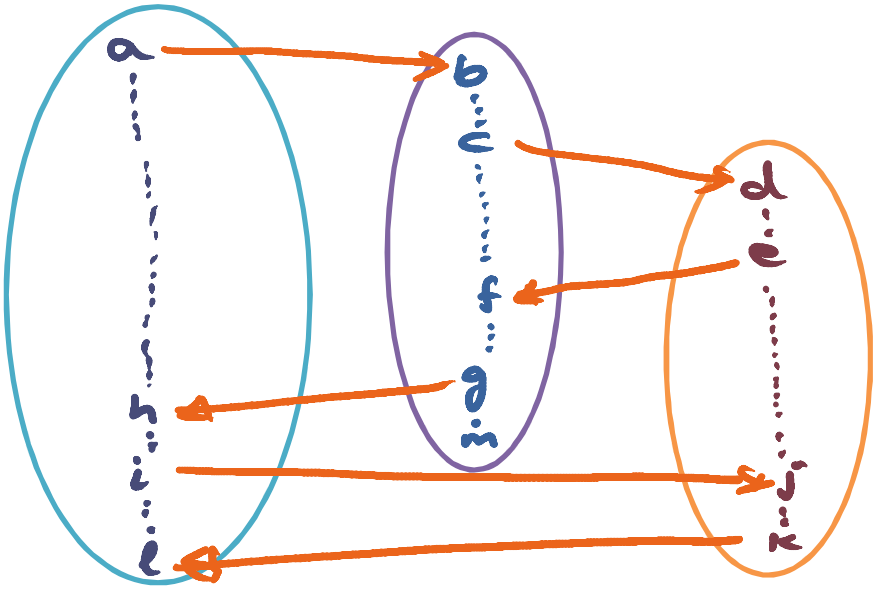
CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

Communication in Distributed Systems

5. (10 **mins**, 20 points)

(a) (5 points)(Lamport)



Use $\longrightarrow$ to indicate “happened before”

Use || to indicate “concurrent”

Connect each of the following events below with || or $\longrightarrow$

(i) a $\longrightarrow$ l

(ii) h $\longrightarrow$ j

(iii) m || k

(iv) l || m

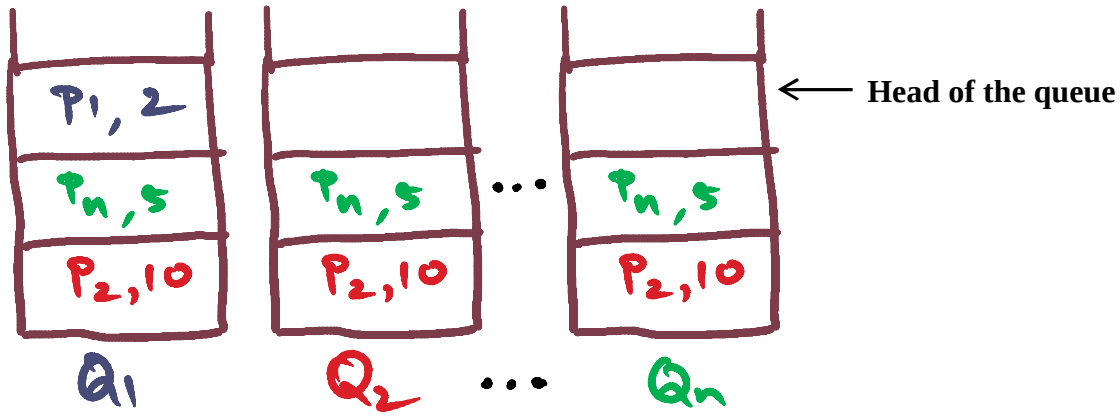
(v) b $\longrightarrow$ h

(+1 for each correct)

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

(b) (5 points)(Lamport)
The figure below shows the state of the local queues In Lamport's distributed mutual exclusion algorithm. <Pi, TS> denotes processor Pi made a lock request at time TS. Qi is associated with processor Pi. Top entry is the head of the respective queues.



When can P1 assume it has the lock?

- When P1 has received either ACKs (or requests timestamped later than its own request) from all the other processors.
  AND
- When P1's own request is at the top of its queue Q1

(-2 if either of the above two criteria are not mentioned)

CS 6210 Spring 2014 Midterm Solution

Name:_____Kishore_____GT Number:

(c) (5 points)(Thekkath and Levy paper on Latency limits)
How can we reduce the marshaling overhead in RPC at the point of the client call?

- Marshal into kernel buffer directly avoiding an extra copy in the client stub (requires having the marshaling code of the client side stub in the kernel),
  OR
- Use a shared descriptor between the client stub and the kernel so that the kernel can create the RPC packet without knowing anything about the semantics of the RPC call.

(+2.5 for each item)

(d) (5 points)(Thekkath and Levy paper on Latency limits)
How can we reduce the context switch overhead in RPC?

- Out of the potential 4 context switches (2 on the client side, and 2 on the server side), only 2 are in the critical path
  - Switching to the server when the remote call comes in
  - Switching to the client when the result comes back
- The other two context switches (to a different process on the client side upon an RPC call by a client; to a different process on the server side once the RPC call has been serviced) can be overlapped with communication.
  (+3 for recognizing that only 2 context switches are in the critical path)
  (+2 for mentioning the second bullet in some form)