Department of Electrical Engineering and Computer Science

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.033 Computer Systems Engineering: Spring 2015

Quiz I

There are 16 questions and 12 pages in this quiz booklet. Answer each question according to the instructions given. You have **50 minutes** to answer the questions.

Some questions are harder than others and some questions earn more points than others—you may want to skim all questions before starting.

If you find a question ambiguous, be sure to write down any assumptions you make. **Be neat and legible.** If we can't understand your answer, we can't give you credit!

Write your name in the space below. Write your initials at the bottom of each page.

THIS IS AN OPEN BOOK, OPEN NOTES, OPEN LAPTOP QUIZ, BUT DON'T USE YOUR LAPTOP FOR COMMUNICATION WITH OTHERS. TURN YOUR NETWORK DEVICES OFF.

CIRCLE your recitation section:

10:00	1. Dina/Andrew	2. Arvind/Manali	3. Karen/Ellen	
11:00	4. Dina/Andrew	5. Arvind/Manali	6. Karen/Ellen	
12:00	7. Sam/David			
1:00	8. Peter/Cong	9. Sam/David	10. Martin/Webb	13. Mark/Amy
2:00	11. Peter/Cong	12. Mark/Amy	14. Martin/Webb	

Do not write in the boxes below

1-3 (xx/18)	4-7 (xx/32)	8-9 (xx/10)	10-11 (xx/14)	12-14 (xx/18)	15-16 (xx/8)	Total (xx/xx)

Name:

Initials:

I Modularity and Naming

1. [8 points]: While browsing the Web, you click on the link `www.course6.com`. Your computer asks your Domain Name System (DNS) name server, M, to find an IP address for this domain name. Which of the following is **always** true of the name resolution process, assuming that all name servers are configured correctly and no packets are lost?

(Circle True or False for each choice.)

A. True / False M must contact one of the root name servers to resolve the domain name.

False. For example, M could contact the `course6.com` nameserver, or the mapping could be cached.

B. True / False M must contact one of the name servers for `course6.com` to resolve the domain name.

False. The mapping could be cached (and valid), in which case M would not need to contact any name server.

C. True / False If M had answered a query for the IP address corresponding to `www.course6.com` at some time in the past, then it will always correctly respond to the current query without contacting any other name server.

False. The mapping in M's cache may have expired by now, in which case M will need to contact another name server.

D. True / False If M has a valid IP address of a functioning name server for `course6.com` in its cache, then M will always get a response from that name server without any other name servers being contacted.

True.

Initials:

II Virtual Memory

Program X is using a page table to map virtual addresses to physical pages. Below is a partial view of this page table, with the physical addresses replaced by question marks. Here, W is the writeable bit and P is the present bit.

virtual address	physical address	W	P
1	?	0	1
2	?	1	1
3	?	1	0

All memory is initialized with 0 values. Program X executes the following three write calls (the syntax here is `Write(a, b)` where `a` is a virtual address and `b` is a value):

```
Write(1, 10)
Write(2, 20)
Write(3, 30)
```

2. [6 points]: Which of the calls will cause an exception (e.g., a page fault)?
(Circle ALL that apply)

A. `Write(1, 10)`
B. `Write(2, 20)`
C. `Write(3, 30)`
D. None

A and C will cause exceptions. A because the writeable bit is zero and C because the present bit is zero.

3. [4 points]: Assume that X has attempted each of the three writes above. Fill in the following table with the contents of each *physical* address. You can assume that whenever a virtual address v is present in RAM, its corresponding physical address is $(v + 2) \mod 4$.

Physical address	Value
0	20
1	30
2	0
3	0

The writes to virtual addresses 2 and 3 (corresponding to physical addresses 0 and 1) both succeeded. Even though the write to 3 caused an exception, when P = 0 the write will still succeed (and the page will be loaded into the page table).

Initials:

III Concurrency

4. [10 points]: The following is the pseudo-code of a library that performs banking transactions, including deposit, balance, withdraw, and transfer. This library is designed for a multi-threaded environment where multiple clients are using it at the same time.

```
void deposit (account* a) {
    acquire(a->lock)
    a->balance = a->balance + 1
    release(a->lock)
}

void balance (account* a) {
    int b;
    acquire(a->lock)
    b = a->balance
    release(a->lock)
    return b
}

void withdraw (account* a) {
    acquire(a->lock)
    a->balance = a->balance - 1
    release(a->lock)
}

void transfer (account* a1, account* a2) {
    acquire(a1->lock);
    acquire(a2->lock);
    a1->balance = a1->balance - 1;
    a2->balance = a2->balance + 1;
    release(a2->lock);
    release(a1->lock);
}
```

Which of the following are true about the above code?

(Circle the BEST answer)

A. Deadlock could occur

B. A race condition exists

C. Both deadlock and races could occur

D. Neither deadlock nor races could occur

A is true. Deadlock could occur by calling `transfer(a, b)` concurrently with `transfer(b, a)` (this is essentially the same example we gave in lecture, when we studied filesystem move). Threads hold `a->lock` whenever they access `a->balance`, so there are no race conditions.

Initials:

5. [8 points]: Indicate which of the following statements about the paper "Eraser: A Dynamic Data Race Detector for Multithreaded Programs", by Savage et al. are true.

(Circle True or False for each choice.)

A. True / False As long as the scheduler ensures that the locking discipline is enforced, no races can occur.

B. True / False Savage et al. claim that Eraser is better than the *happens-before* technique at detecting races.

C. True / False Eraser is practical in part because the set of lock combinations observed in practice is small.

D. True / False Eraser is guaranteed to find all possible races, though it may also identify some potential races that are false positives.

False, True, True, False.

IV Operating System Structure and UNIX

6. [8 points]: Which of the following are true about virtual machines?
(Circle True or False for each choice.)

A. True / False Suppose Program A and Program B are two independent programs running inside the same guest OS. The virtual machine monitor prevents bugs in Program A from crashing Program B.

B. True / False Suppose Guest OS A and Guest OS B are two guest operating systems running on a single physical machine. The virtual machine monitor prevents bugs in Guest OS A from crashing Guest OS B

C. True / False There is no reason to use virtual machines if the guest operating system is a microkernel rather than a monolithic kernel.

D. True / False Using virtual machines ensures that no bug can crash an entire physical machine.

A - False. This is the job of the OS.
B - True. This is what VMMs are for.
C - False.
D - False. There could be a bug in the VMM itself.

7. [6 points]: Consider UNIX as described in the paper "The UNIX Time-Sharing System" by Ritchie and Thompson. Which of the following statements about processes in UNIX are true?
(Circle True or False for each choice.)

A. True / False Checking the return value of `fork()` enables a child to execute different instructions from its parent.

B. True / False Running `command &` causes the shell to create a new process for the command, start it, and then move on to read the next command. Running `command`, on the other hand, causes the shell to run `command` to completion rather than creating a new process.

C. True / False If the shell uses `fork()` to create a new process to run a command, then the child process, and not the parent process, executes the command.

A - True. `fork()` returns 0 to the child process, and the child's pid to the parent.
B - False. A new process is always created.
C - True. The parent process—the shell—calls `fork()`, and then has to use `wait()` to determine when the child is done so that it can move on to the next command.

Initials:

V Performance

Boole, LLC. is using the word count program in the MapReduce paper ("MapReduce: Simplified Data Processing on Large Clusters" by Dean and Ghemawat) to count the number of occurrences of "and" and "not" in a set of documents. They have **four machines** that execute `map` tasks and **two machines** that execute `reduce` tasks (the `reduce` machines are different than the `map` machines, i.e., there are six machines total). They run the computation on 4000 documents, all of which contain the string "and not" and nothing else.

The MapReduce infrastructure produces eight splits, with each split containing 500 of the 4000 documents. Unfortunately, because Boole decided to economize on their `map` machines, these machines are slow and unreliable. It takes each `map` machine **two seconds** to process a split (the start-up times, etc., of the machines are negligible). **Three seconds** into the `map` computations, two of the machines overheat, burst into flames, and crash.

8. [6 points]: How many `map` tasks will the system execute to completion?
(Circle the BEST answer)

A. 6

B. 8

C. 10

D. None of the above

10. In the first two seconds, four maps complete (one per machine). In the second two seconds, four maps are started, but then two machines crashed; call them A and B. This means that a) A and B's map jobs didn't complete and b) we lose all data from A and B, including the results of their earlier map jobs. So 10 total jobs complete, since we need to re-run the 2 previously-completed jobs on A and B.

Boole's system administrators are working overtime and unfortunately configured one of the `reduce` machines with the caches turned off so that it runs 100 times slower than the other machine. The partitioning function assigns "and" and "not" to different partitions. The MapReduce infrastructure does not start the `reduce` tasks until all of the `map` tasks finish.

9. [4 points]: How many `reduce` tasks will the MapReduce infrastructure start?
(Circle the BEST answer)

A. 1

B. 2

C. 3

D. None of the above

3. The system will start two tasks on the machines initially; let's call A the fast one and B the slow one. The task on B will be a straggler, and so will be restarted on A.

Initials:

VI Interdomain Routing and RON

10. [10 points]: Consider four Internet Service Providers (ISPs), W, X, Y, and Z, each a separate Autonomous System (AS). You find that an IP prefix, p, owned by Y, is advertised using BGP between these autonomous systems **only** as follows (there are no other advertisements concerning p between these four autonomous systems):

- W directly advertises p to Z.
- Y directly advertises p to W, X, Z.
- Z directly advertises p to W.

Which of the following transit and peering relationships between the four autonomous systems is **fully consistent** with the above advertisements?

(Circle True or False for each choice.)

A. True / False
Y is in a peering relationship with W, X, Z;
W and Z are in a peering relationship.

B. True / False
W and Z provide transit to Y;
X and Y are in a peering relationship;
W and Z are in a peering relationship.

False, True, True

C. True / False
X and Y are in a peering relationship;
W provides transit to Z and to Y;
Z provides transit to Y.

11. [4 points]: The paper on RON by Andersen *et al.* showed that an overlay network can improve the availability and performance of Internet routing. Their observed results are because:

(Circle True or False for each choice.)

A. True / False It is possible on the Internet for the IP path from host A to host B to have longer latency and a higher loss rate than the composition of IP paths from host A to host C and host C to host B.

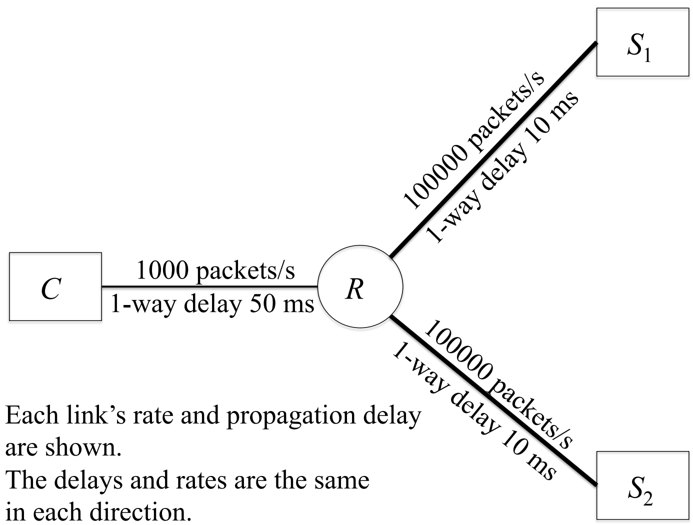
B. True / False RON's active probing reacts to congestion along paths and reroutes packets around paths with excessive congestion, whereas BGP usually does not.

True, True

Initials:

VII Annette's RPC Network

Annette Werker is building a network with two servers, S_1 and S_2, connected to a client, C, via a router, R, as shown here:



The packet transmission and processing delays are negligible and may be ignored: assume that the round-trip time (RTT) depends only on the propagation delays of the links and possibly on queueing delays.

The workload in this network consists of remote procedure calls (RPCs) issued by the client, C, to the servers, S_1 and S_2. The RPC request fits in one small packet (100 bytes). An RPC response is one of two types:

- an **elephant**, which is greater than or equal to 1000 packets long, or
- a **mouse**, which is between 8 and 16 packets long

The request and response messages are both sent over TCP, which provides an in-order reliable transport service with congestion control. The congestion control uses a congestion window (cwnd), which varies according to the "slow start", "additive increase", and "multiplicative decrease" rules described in lecture.

In the next few questions, until we specify otherwise, assume that:

- There is **exactly one** TCP connection in the network. The questions specify the type of the connection (mouse or elephant).
- The router, R, has a buffer size of 100 packets.

12. [4 points]: Suppose the client issues a 100-byte mouse request to server S_1 at time 0. At what time does the client receive the complete response from the server? Assume that no packets are lost, that there are no other concurrent connections, and that the initial value of cwnd is 1. **Pick the answer closest to the correct value.**

A. 120 milliseconds.

B. 240 milliseconds.

C. 360 milliseconds.

D. 480 milliseconds.

D, which is equal to four round trip times. The client issues the request, and then receives a single packet from the server, who is beginning slow start. In the next round trip, the client receives two packets from the server, and in the next one, four. At this point, the client has received 7 packets. It will take at least one more RTT to complete the mouse transfer (which is at least 8 packets), possibly two RTTs if the mouse is 16 packets. Thus, it takes 4–5 RTTs, or 480-600ms. Either way, the correct answer is D.

13. [6 points]: Suppose a single elephant connection has been running for some time and is now in **additive increase** mode, with 1000 packets left to send. Assuming that R is using DropTail for queue management, what is the value of cwnd when the connection experiences its next packet loss? **Pick the answer closest to the correct value.**

A. 61 packets.

B. 101 packets.

C. 121 packets.

D. 221 packets.

D, 221 packets. The bandwidth-delay product + the number of packets the queue can hold is 1 packet-per-millisecond*120milliseconds + 100 = 220. Thus, with a window size of 221 ($>$ 220), loss occurs.

Initials:

New assumptions: When Annette isn't looking, Louis Reasoner increases R's buffer size to 4000 packets. The workload consists of one elephant RPC and one mouse RPC between each S_i and C. The workload is "continuous"; i.e., whenever an RPC of a given type (mouse or elephant) between S_i and C finishes, another of the same type starts, between the same server and C, so there are always two elephants and two mice RPCs running at any point in time.

14. [8 points]: Each statement below describes a network resource management scheme. Compare each scheme to Annette's network running DropTail queue management and first-in-first-out (FIFO) scheduling.

(Circle True or False for each choice.)

A. True / False Suppose R runs the RED active queue management scheme, with parameters set to make the average queue length be 100 packets for this workload. Then, the average completion time of the mouse RPCs would decrease.

B. True / False Suppose R runs deficit round-robin with two queues, where queue i holds packets sent by S_i. Then, the average completion time of the mouse RPCs would decrease.

C. True / False Suppose R runs priority queueing, separating mouse and elephant data into separate queues and prioritizing the mouse queue over the elephant queue. Then, the average completion time of the mouse RPCs would decrease *and* the average completion time of the elephant RPCs would not increase.

D. True / False Suppose R runs deficit round-robin, separating mouse and elephant data into two separate queues, and assigning equal rates to the two queues. Then, the average completion time of the mouse RPCs would decrease.

A - True
B - False
C - False
D - True

Initials:

15. [4 points]: Annette sees what Louis has done and decides to run Alizadeh et al.'s Data Center TCP (DCTCP) scheme, keeping R's buffer at 4000 packets. The workload remains the same (one elephant and two mouse RPCs from each server).

(Circle True or False for each choice.)

A. True / False A correctly implemented DCTCP will reduce the average completion time of mouse RPCs, while maintaining high network utilization.

B. True / False R sets explicit congestion notification (ECN) marks when the queue length exceeds a threshold, and S_i reduces cwnd by half if it receives an acknowledgment with ECN set.

A - True
B - False, S_i reduces cwnd in proportion to the congestion, not always by half.

16. [4 points]: Which of these is an example of an **end-to-end argument** in Annette's network and RPC system?

(Circle True or False for each choice.)

A. True / False Annette's network uses topological addressing.

B. True / False Annette's RPC implementation does not assume that what it gets from TCP is accurate, even though TCP is a reliable protocol, but uses its own method to verify that the RPC query and response are correct.

A - False
B - True

End of Quiz I

Please double check that you wrote your name on the front of the quiz, and circled your recitation section.

Initials: