Department of Electrical Engineering and Computer Science

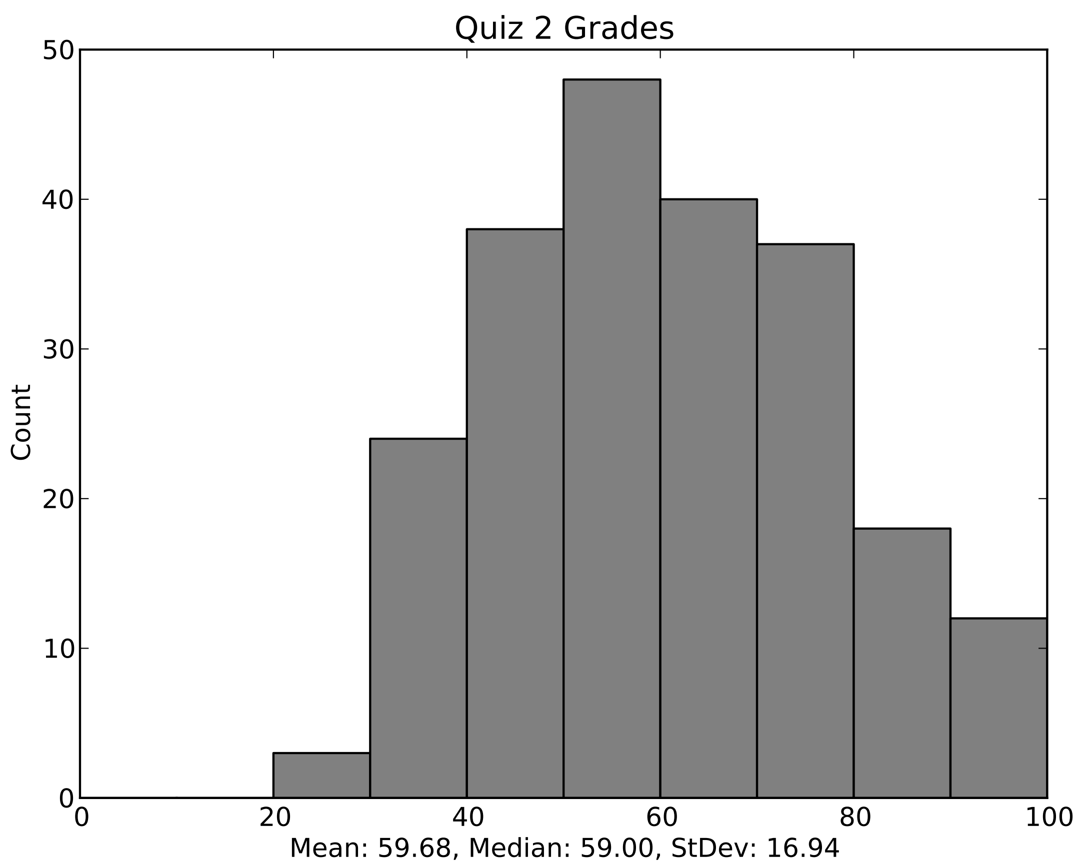
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.033 Computer Systems Engineering: Spring 2012

Quiz 2 Solutions

There are 12 questions and 10 pages in this quiz booklet. Answer each question according to the instructions given. You have **50 minutes** to answer the questions.

Grade distribution histogram:



I Reading Questions

1. [10 points]: Answer the following question based on the paper "Do incentives build robustness in BitTorrent?". The main feature that distinguishes BitTyrant from BitTorrent is:

(Circle the BEST answer)

A. Data transfers with BitTyrant clients are all transactions.

B. BitTyrant data transfers can be rolled back.

C. BitTyrant clients find their peers with the use of distributed hash tables.

D. BitTyrant clients choose their peers in a way that leads them towards downloading more information than they upload.

E. BitTyrant's algorithms model altruism by using cumulative distribution functions (CDF).

Answer: D.

2. [8 points]: Choose the correct statement about System R's recovery process, based on the paper you read in recitation.

(Circle the BEST answer)

A. During the recovery process in system R there is a complete redo of all transactions that were committed after the checkpoint, and for uncommitted transactions there is an undo of the parts that happened before the last checkpoint.

B. During the recovery process in system R there is a complete redo of all transactions that were committed after the checkpoint, and for uncommitted transactions there is a redo of the parts that happened before the last checkpoint.

C. During the recovery process in system R there is a complete redo of all transactions that were committed after the checkpoint, and for uncommitted transactions there is an undo of the parts that happened after the last checkpoint.

D. During the recovery process in system R there is a complete redo of all transactions that were committed after the checkpoint, and a complete undo of all uncommitted transactions.

Answer: A.

3. [8 points]: In the LFS system, as described in the paper that you read in recitation, what is the commit point, meaning the point at which a single data block overwritten in a file will be available after a crash, assuming that there are no software bugs and the contents of the disk survive the crash?

(Circle the BEST answer)

A. The application's `write()` system call returns.

B. The application calls `close()` on the file.

C. The file's data block is flushed to the log.

D. The file's updated inode is flushed to the log.

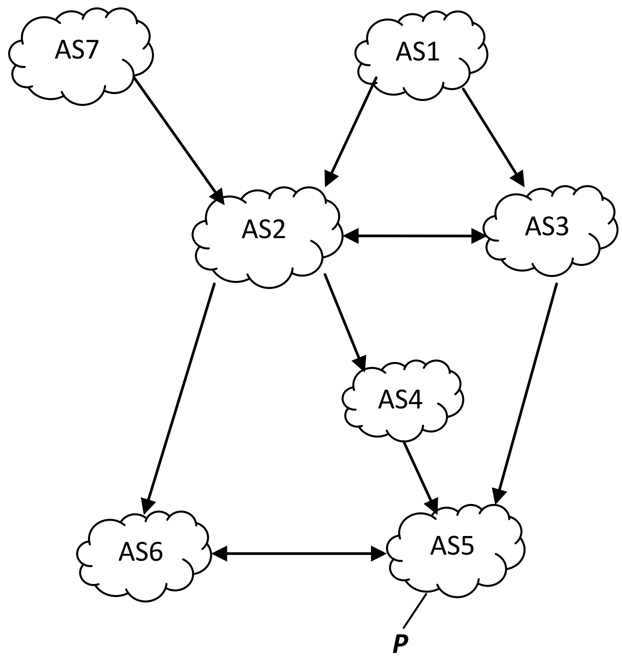
E. The segment summary for the segment containing the file's data and updated inode is flushed to the log.

F. The checkpoint region is updated to include the location of the updated inode.

Answer: E.

II BGP

Consider the topology below, where arrows point from provider to customer and bi-directional arrows refer to a peering links. Assume that the ASes follow the import and export rules described in the required reading titled "Wide-Area Internet Routing". The figure shows that the address prefix P belongs to AS 5.



Answer the questions on the next page.

Answer note: During the quiz many students overlooked the definition of arrows as pointing from provider to customer in the text of the question, and instead assumed it was the opposite. We graded this question under each of the two interpretations, and gave each student the higher of the two scores.

4. [5 points]: Which AS path does AS1 follow to reach prefix P in AS5? You only need to list the ASes on the path in order, starting with AS1 and ending with AS5.

Answer (for both arrow interpretations): AS1 $\rightarrow$ AS3 $\rightarrow$ AS5.

5. [5 points]: How many paths does AS2 learn for prefix P? Which of these AS paths does it use to deliver packets to prefix P? Again, write the path as an ordered list of ASes.

Answer (arrows from provider to customer): Number of paths: 3
Answer (arrows from customer to provider): Number of paths: 2

Answer (arrows from provider to customer): AS2 $\rightarrow$ AS4 $\rightarrow$ AS5
Answer (arrows from customer to provider): AS2 $\rightarrow$ AS4 $\rightarrow$ AS5 is a valid answer, AS2 $\rightarrow$ AS6 $\rightarrow$ AS5 as well (not both)

6. [8 points]: Say that the link between AS2 and AS4 and the link between AS6 and AS5 both fail. Would any of the ASes in the figure lose the ability to reach prefix P (i.e., will have no routes to P even after the routing stabilizes)? If "yes", tell us which AS or ASes?

Answer (arrows from provider to customer): Yes. AS7 loses the ability to reach prefix P. All other ASes will have routes. The interesting path to consider is AS7 $\rightarrow$ AS2 $\rightarrow$ AS3 $\rightarrow$ AS5. Because AS2 is a customer of AS7, and a peer of AS3, it will not announce routes it learns from AS3 to AS7. As a result, AS7 will not know about this path, and will not have any routes to P.

Answer (arrows from customer to provider): Yes. AS2, AS6, and AS7 all lose connectivity to P.

7. [8 points]: Say that the topology is as in the Figure except that the link between AS2 and AS3 fails (this is the only link that fails). Let the routes stabilize after the failure. Does any of the ASes change the route it uses to reach prefix P before and after the failure? If "yes", which AS, and what is the new route?

Answer (both interpretations): No AS changes routes.

III Fault-tolerance

8. [12 points]: Alyssa P. Hacker stores her data on a fault-tolerant storage system, which has 5 1 TB disks that together provide 4 TB of aggregate space, and can tolerate any one (but no more) disk failure. Each disk has a MTTF of 10^6 hours, and you can assume that all disk failures are independent. When a disk fails, Alyssa orders a replacement from Amazon which arrives in approximately 4 days (100 hours). Once Alyssa plugs in the replacement disk, assume that its content is reconstructed instantaneously. What is the MTTF for Alyssa's storage system?

(Circle the BEST answer)

A. 10^6 hours

B. 4×10^7 hours

C. 5×10^8 hours

D. 2×10^9 hours

E. 10^{10} hours

Answer:

The probability that a single disk fails in a given hour is $\frac{1}{10^6}$. Since Alyssa's system has five disks, the probability that one of these five disks fails in a given hour is $\sim \frac{5}{10^6}$. Once one disk fails, the overall system will fail if, in the next 100 hours, another disk fails. The probability that one out of four disks will fail in a given hour is $\frac{4}{10^6}$, so the probability that one of four disks will fail in a 100 hour interval is $\frac{4 \cdot 100}{10^6}$. The MTTF of Alyssa's system is the inverse of the probability of the system failing in any given hour, and that probability is $\frac{5}{10^6} \times \frac{4 \cdot 100}{10^6}$.

Thus, the final answer is $\frac{1}{\frac{5}{10^6} \times \frac{4 \cdot 100}{10^6}} = \frac{1}{\frac{20}{10^{10}}} = 5 \times 10^8$. C.

IV Isolation

9. [18 points]: Alyssa P. Hacker switches her transactional system to a weaker isolation mode, where one transaction that's still running can observe the effects of any other committed transaction, even if that committed transaction started later. (In Postgres, this is called `READ COMMITTED`.) Alyssa runs two transactions, `xfer(A, B, 10)` and `audit(50)`, whose code is below, both of which eventually commit but can run concurrently:

```
xfer(a, b, amt):
  accounts[a].balance = accounts[a].balance - amt
  accounts[b].balance = accounts[b].balance + amt

audit(thresh):
  for a in accounts:
    if a.balance > thresh:
      sum += a.balance
  return sum
```

Every reference to the `balance` field of an account is a separate read or write. Do not make any assumptions about the order in which `audit` iterates over the accounts. The database starts in the following state:

Account	*Balance*
A	55
B	45

Which of the following tuples of values (A, B, audit), representing the final state of the database and the return value of the `audit` function, can result after running these two transactions in Alice's weaker isolation mode?

(Circle True or False for each choice.)

A. True / False A=45, B=55, audit=0

Answer: True. Audit reads B=45, fails if check. Then xfer commits. Then audit reads A=45, fails if check.

B. True / False A=45, B=55, audit=45

Answer: True. Audit reads B=45, fails if check. Then audit reads A=55, passes if check. Then xfer commits. Then audit reads A=45, adds to sum.

C. True / False A=45, B=55, audit=55

Answer: True. xfer commits. Then audit reads A=45, fails if check. Then audit reads B=55, passes if check. Then audit reads B=55, adds it to sum.

D. True / False A=45, B=55, audit=90

Answer: False. The only way to get 90 would be to get audit() to add 45 twice. audit() only adds 45 if it first sees 55, then xfer commits, and then it reads 45 when adding to sum. Since there is only one xfer that can commit, there's only one 45 that audit can add.

E. True / False A=45, B=55, audit=100

Answer: True. Audit reads A=55, passes if check. Then xfer commits. Then audit reads A=45, adds it to sum. Then audit reads B=55, passes if check. Then audit reads B=55, adds it to sum.

F. True / False A=45, B=55, audit=110

Answer: True. Audit reads A=55, passes if check. Then audit reads A=55, adds it to sum. Then xfer commits. Then audit reads B=55, passes if check. Then audit reads B=55, adds it to sum.

V Performance

Ben Bitdiddle runs the database for a credit card processing company. Ben's clients are online stores that each have a single program running the following pseudocode to charge customer credit cards:

```
while True:
  ccnum, amount = get_next_order()
  send(ben, {ccnum, amount})
  status = recv(ben)
```

Ben's server looks like the following pseudocode, where `write_log()` prepares the log records, but only `flush_log()` writes to disk. Assume that every log flush requires seeking to a new location. Assume that `reply()` does not block.

```
while True:
  ccnum, amount = recv()
  status = check(ccnum, amount)
  if status == OK:
    write_log(ccnum, amount)
    flush_log()
  reply(status)
```

Ben's database must commit each credit card charge operation before responding to the client. Ben uses a typical rotational disk, with a 10 msec average seek time, and 100 MB/sec sequential throughput. Each log record is 512 bytes. The round-trip latency between each client and the server is 100 msec. Assume that the `check()` function is instantaneous.

10. [4 points]: How many credit cards can Ben's system successfully charge per second, if Ben has a large number of clients?

Answer: approximately 100 per second: limited by the server only being able to issue 1 `flush_log` per 10 msec seek.

11. [6 points]: How many credit cards can Ben's system successfully charge per second, if Ben has 5 clients, the clients spend almost no time in `get_next_order()`, and `get_next_order()` never blocks?

Answer: approximately 50 per second: limited by the requests that a client can issue: one request per 100 msec roundtrip to the server.

Initials:

Ben decides that he wants to process more transactions per second, and changes the server code to flush the log after a batch of records have been written to the log:

```
while True:
  for i = 1..N:
    ccnum[i], amount[i] = recv()
  for i = 1..N:
    status[i] = check(ccnum[i], amount[i])
    if status[i] == OK:
      write_log(ccnum[i], amount[i])
  flush_log()
  for i = 1..N:
    reply(status[i])
```

12. [8 points]: How many credit cards can Ben's modified system successfully charge per second, if Ben has a large number of clients? Assume the best choice of N.

Answer: about 204,800 per second: limited by the bandwidth of the disk: 100MB/sec / 512 bytes per record = 204,800 records per second.

End of Quiz II

Double-check that you wrote your name on the front of the quiz, and circled your recitation.

Initials: