SOLUTION

# *ECE 6102 Exam 2*

**Instructions:** You have 1 hour, 20 minutes to complete this test. The test is closed book and closed notes. No calculators, phones, or other electronic devices are allowed. Make sure to show your work on all problems. No credit will be given for answers without sufficient work.

| | |
|---|---|
| Multiple Choice | _____ |
| Problem 6 | _____ |
| Problem 7 | _____ |
| Problem 8 | _____ |
| **Total** | _____ |

**Multiple Choice:** 10 points (choose only one answer to each question)

1) Which of the following is *not* true of threshold masking Byzantine quorum systems, where the maximum number of faulty servers at any time is $f_{max}$?
   a) at least $4 f_{max} + 1$ total servers are required for correct operation
   b) the intersection between any read quorum and any write quorum must be at least of size $2 f_{max} + 1$
   c) for a read operation, the first value that matches from $f_{max} + 1$ servers must be the correct and up to date value of the object
   d) the load of any threshold masking Byzantine quorum system is at least 0.5
   e) all of the above are true

2) Which of the following statements about Shamir's $(k, n)$ secret sharing scheme *is* true?
   a) having at most $k$ out of $n$ shares reveals no information about the secret
   b) the secret value is a random point on a degree $k - 1$ polynomial
   c) after generation of an initial $n$ shares, additional shares cannot be added easily without changing the polynomial
   d) when generating shares of a secret, $k - 1$ coefficients of the polynomial can be selected at random
   e) all of the above are false

3) Which of the following statements about the "GridSharing" paper is *not* true? Recall that $c$, $l$, and $b$ are the maximum number of crash, leakage-only, and Byzantine faults, respectively.
   a) the total number of servers needed in the direct approach is $(l+b+1)\bullet(3b+c+1)$
   b) the GridSharing approach allows the total number of servers to be reduced at the cost of increasing the number of shares, as compared to the direct approach
   c) the number of shares required in the GridSharing approach increases more rapidly as $b$ increases than it does as $l$ increases
   d) the share assignment scheme developed by Ito, *et al.,* is used to determine which servers store which shares in the GridSharing approach
   e) all of the above are true

4) Which of the following statements about BitTorrent *is* true?
   a) BitTorrent was originally designed as a music file sharing system
   b) optimistic unchoking allows a peer to discover new peers that can increase its download rate
   c) pieces are always selected by the rarest-first strategy
   d) tit for tat allows a group of peers to isolate a cooperative peer
   e) all of the above are false

5) Which of the following is *not* a feature of the Dynamo system?
   a) eventual consistency of non-conflicting written values
   b) quorum-based read and write operations
   c) consistent hashing for mapping keys to nodes
   d) vector clocks to determine when different object versions are causally related
   e) all of the above are features of Dynamo

6) 30 points

Consider a Chord system that uses 7-bit hashes to represent node IDs and keys. Assume that there are 15 nodes in the system and their node ID hashes are: 7, 17, 21, 33, 45, 49, 62, 78, 91, 96, 105, 106, 113, 118, and 126.

a) List the finger table entries for node 49.

b) List the route (sequence of nodes) that a query by node 78 for the key 17 would follow.

(a) Routing Table for 49:

| $2^i$ offset | node ID |
|---|---|
| 1 | 62 |
| 2 | 62 |
| 4 | 62 |
| 8 | 62 |
| 16 | 78 |
| 32 | 91 |
| 64 | 113 |

(b) Key 17 is stored at node ID 17

$$\begin{array}{r} 78 \\ +64 \\ \hline 142 \end{array}$$

142 mod 128 = 14 ⇒ +64 routing table entry for node 78 is 17

∴ 78 routes request directly to node 17 in 1 hop!

7) 30 points

In Megastore, assume the write-ahead logs of 3 replicas for a particular entity group are as follows:

| Round | Replica 1 | Replica 2 | Replica 3 |
|---|---|---|---|
| 105 | a | a | a |
| 106 | b | b | g |
| 107 | | c | h |
| 108 | d | d | |

Recalling the concepts of choosing, accepting, and learning from the Paxos protocol, describe the status of each replica and the overall Paxos consensus for ***each*** of the above rounds.

R105: 1) each replica has accepted a
2) a has been chosen since it was accepted by a majority
3) don't know which replicas have learned that a was chosen

R106: 1) Replicas 1, 2 accepted b ; Replica 3 accepted g
2) b has been chosen since it was accepted by a majority
3) don't know if 1, 2 have learned b was chosen; 3 has definitely not learned that b was chosen

R107: 1) Replica 1 has not accepted a value; Replica 2 accepted c; Replica 3 accepted h
2) no value has been chosen
3) no replicas have learned a chosen value

R108: 1) Replicas 1, 2 accepted d; Replica 3 has not accepted a value
2) d has been chosen
3) don't know if 1, 2 have learned that d was chosen; 3 has not learned that d was chosen

8) 30 points

Using the MapReduce programming paradigm with either Python or Java, write a map() function and a reduce() function to find the total dollar amount of songs sold for each genre from a file of song purchases in the format of Assignment 3. Recall that the fields in the file format are (in this order): date, time, user, song title, album title, genre, song length, and price. Also, the different genres are: Alternative, Country, Dance, Hiphop, Jazz, Pop, and Rock.

```
def genre_amount_map (data):
    (entry, text_fn) = data
    text = text_fn()
    for l in text.split("\n"):
        fields = l.split("\t")
        yield (fields[5], fields[7])

def genre_amount_reduce(Key, amounts):
    prices = map(float, amounts)
    yield (Key, sum(prices))
```