Carnegie Mellon University in Qatar

Distributed Systems

15-440 - Fall 2023

Mock Quiz 1 Solutions

Name: ____

Andrew ID: ____________________________________

Total time: 45 minutes

Instructions:

- Write your answers in the spaces provided below the problems. If you make a mess, clearly indicate your final answers.
- This quiz has 8 questions over 4 pages, for a total of 25 points.
- Keep up with time.

Question	Points	Score
1	6	
2	3	
3	2	
4	2	
5	2	
6	4	
7	3	
8	3	
Total:	25	

6pts 1. Answer the following questions by selecting **True** or **False**:

(a) (**True** / [✓ **False**]) RPC allows passing parameters by value only.

(b) ([✓ **True**] / **False**) Marshalling and un-marshalling are performed within the stubs and skeletons of RPC.

(c) (**True** / [✓ **False**]) A workable distributed file system can be built without incorporating a naming service.

(d) ([✓ **True**] / **False**) The distributed Bellman-Ford algorithm uses essentially a flat naming service protocol for locating machines over the Internet.

(e) ([✓ **True**] / **False**) Blockchain uses an unstructured peer-to-peer architecture.

(f) (**True** / [✓ **False**]) Broadcasting is an effective naming service especially in WAN settings.

(g) (**True** / [✓ **False**]) An approach to locating mobile entities is to use forwarding pointers, whereby an entity moving from location A to location B can leave behind a server stub (or a skeleton) to its new location at B.

(h) (**True** / [✓ **False**]) No fault-tolerance measures need to be taken in RPC if it is layered on top of TCP.

(i) ([✓ **True**] / **False**) The at-least-once semantic in RPC can only be used with idempotent operations.

(j) (**True** / [✓ **False**]) Latency is a measure of throughput.

(k) (**True** / [✓ **False**]) If a server crashes before all the actions of a non-idempotent operation are done, the system has to redo all the actions that were executed before the crash.

(l) (**True** / [✓ **False**]) It is better to use synchronous RPC for operations like training a deep learning model (or what is referred to as batch processing).

3pts 2. What is the difference between *layered* and *tiered* architectures? Give an example of when you would you use a layered architecture, but not a tiered one?

Layered architecture: It *vertically* organizes the functionalities of a system, such that an upper layer can capitalize on a lower layer by simply calling any of such available functionalities, without needing to re-implement them.

Tiered architecture: It *horizontally* organizes the functionalities of a system (e.g. presentation logic, application logic, data logic) via mapping them across two or more distributed servers.

Example: The OSI networking layers as they can tap into each other on a single machine in a modular and more maintainable manner.

2pts 3. You have been asked to design and implement a distributed system for video streaming, which requires high Quality-of-Service. Would you use TCP or UDP for your middleware implementation? Justify your answer.

To achieve fast streaming, UDP is preferred since it is *lightweight* (i.e., it consumes less bandwidth as it does not require an ACK for *each* packet, re-transmissions, and a timeout mechanism). In addition, video streaming might tolerate a small number of packet losses (e.g., we can afford to lose a few frames from the set of frames that pertain to the same region or component like a green field in a streamed video of a soccer game). Note, however, that if this number of packet losses increases dramatically (e.g., due to network faults), the resolution will be significantly impacted. In this case, it might still be preferable to use UDP, but handle *non-common*, intense network failures through the middle-ware layer (in systems it is always better to optimize for common cases).

2pts 4. What is the weakest RPC semantic (i.e., *exactly-once*, *at-most-once*, or *at-least-once*) that you would suggest for requesting a taxi through *Uber*. Discuss.

Placing a ride request through the Uber app might necessitate paying for that ride upfront and shouldn't allow duplicate ride bookings. This suggests a non-idempotent operation (i.e., a request cannot be executed more than once), hence, the at-least-once semantic is not an option. Rather the at-most-once is the preferred option, opting for the *weakest* semantic (exactly-once is ideal but *stricter*).

2pts 5. Why are marshaling and unmarshaling important in exchanging data between communicating entities in a distributed system?

To mask heterogeneity. To elaborate, marshaling converts data from platform-dependent representations to platform-independent ones. Unmarshalling does the opposite. This allows entities with diverse architectures, operating systems, programming languages, and computer hardware to work together.

4pts 6. What are the three major problems that blockchain tries to solve? Discuss briefly how it solves only *one* of them?

The 3 major problems that Blockchain tries to solve are: (1) The need for a third trusted party (e.g., a central bank) in moving digital assets (e.g., coins) from one entity to another (or what is referred to as a transaction), (2) Long delays in executing transactions (e.g., as of now it may take 2-3 days to transfer money from one bank account to another), and (3) high transaction cuts (e.g., charged by banks involved in fulfilling transactions).

Blockchain solves the first problem by using a distributed peer-to-peer open ledger.

3pts 7. If a mobile computer is to remain accessible to clients when it moves across LANs, it must retain a single IP number. However, IP routing is subnet-based. Subnets are at fixed locations, and the correct routing of packets to them depends upon their positions on the network. Discuss a way of how location transparency can be achieved in such an environment (i.e., IP communication continues normally when a mobile computer moves between subnets at different locations).

A home node can be used where a mapping between a fixed single IP number (or address) and dynamic ones (that are assigned to the mobile computer while it is moving) is maintained. Entities will be able always to locate the mobile computer by checking first with the home node.

3pts 8. How to ensure the durability of non-idempotent operations?

By redoing the actions of the committed ones.