CSCI-1380: Distributed Computer Systems
Homework #2
Assigned: 02/09/2021
Due: 02/16/2021

1   DNS and Networking

You recently joined the file system team at Amazin.io, a distributed storage system (i.e., a rival to dropbox). Your team is in charge of the ReadData(fileHandle, numBytes) function which reads the *next* numBytes from the file. Currently, your team exposes ReadData(...) with an RPC framework that provides *at-least-once* semantics. Prior to calling ReadData(...), the client must have called open(filepath), which opens the file, sets the read pointer to the beginning of the file, and returns the *fileHandle*.

1. Is ReadData(...) Idempotent? Explain.

2. If not idempotent, how would you make this idempotent? Describe the changes you would make within ReadData and present a new function signature (if required).

3. Currently, the system is using UDP for these RPCs. Your team has observed that for clients on cellular connections, the loss is significant. They are proposing to switch to TCP to benefit from the automatic loss recovery mechanism. Given your existing RPC semantics: is this change necessary to address loss? explain.

2   DNS and Load Balancing

- Large online service providers, e.g., Facebook and Google, often employ two levels of load balancing. First at a global level (which decides the data centers that a user should go to) and then a cluster level (which decides which servers within a cluster a user should go to). These two levels complement and support each other.

  1. While many employ "Round Robin" or "Random (e.g., modulo or consistent hashing)" based approaches within a cluster, they do not employ such strategies at the global level. Explain why.
  2. What are two strategies that a provider may employ at the global level? Explain the motivation for both strategies. (Note: While there are more than two strategies available, you only need to list two).

- DNS is a widely adopted mechanism for translating names to IP addresses. Large online services often use this to direct users to specific data centers (i.e., to control or load balancer users across data centers).

  1. One of the optimizations within DNS is to cache results. Explain how caching can impact a company's global load balancing policies.
  2. What is the primary mechanism that online services have at their disposal to control DNS caching? Explain how it works and why it sometimes ineffective.

3 Handing In

Once finished, you should hand in a PDF with your answers on Gradescope. Gradescope will allow you to select which pages contain your answers for each part of each question.

Please let us know if you find any mistakes, inconsistencies, or confusing language in this or any other CS138 document by filling out the anonymous feedback form: http://cs.brown.edu/courses/cs138/s21/feedback.html