CSCI 4510/6510: Distributed Systems and Algorithms - Homework 2 Problem Set

Due Wednesday, October 16, 2024 at 11:59pm in Gradescope

Submission requirements

- Homework must be typed and submitted as a pdf file in Gradescope.
- If you include figures, they may be hand-drawn or digitally created.
- Solutions must be relevant to your specific code submission to receive full credit. You should reference and/or explain key details of your implementation to illustrate your point when applicable.
- Solutions must be self-contained. The grader should not need to check your program code to grade the solutions.
- You may discuss the problems and solutions with others, but your must write up your solutions independently.

Required for CSCI 4510 and 6510

Problem 1 Give a simple example where multiple sites generate conflicting operations in your system, and these conflicting operations are replicated to all sites. Explain what commands are entered at each site, what messages are sent, and how your implementation orders these conflicting operations so that the system achieves causal consistency.

Problem 2 One of the efficiencies of the Wuu Bernstein dictionary algorithm is that a site updates its dictionary incrementally, rather than recomputing it whenever a new event record is added to the site's log. In pseudocode or a paragraph, explain how you implemented this efficient update to the dictionary after a site receives a message, and how you truncated the site's log. Give the running time of your dictionary update and log truncation solutions in terms of the relevant properties of the system, e.g., the number of event records in PL_i and NP. *In class, these updates were presented in set notation. This question asks you to describe algorithms that achieve the outcomes given in the set notation.*

Problem 3 Is it necessary for every site to apply every operation in the same order to guarantee causal consistency of the key-value store Answer YES or NO and justify your answer.

Required for CSCI 6510 only

Problem 4 The paper by Wuu and Bernstein suggests several strategies to improve the storage utilization and communication costs of the replicated log and dictionary algorithms. Consider strategy 2 in the paper. Discuss the pros and cons of using this strategy in your implementation of the key-value store instead of the original solution. In your answer, consider the three types of resources: compute, storage, and network, as well as different types of workloads. *Note: you are not asked to implement strategy 2.*