CS425 Fall 2015 – Homework 4
(a.k.a. “The Interview”)

Out: Nov 17, 2015. Due: Dec 8, 2015

Topics: Lecture 21 onwards

Instructions:

1. **Attempt any 8 out of the 10 problems** in this homework (regardless of how many credits you’re taking the course for). If you attempt more, we will grade only the first 8 solutions that appear in your homework (and ignore the rest). Choose wisely!
2. Please hand in **hardcopy solutions that are typed** (you may use your favorite word processor. We will not accept handwritten solutions. Figures and equations (if any) may be drawn by hand.
3. Please **start each problem on a fresh sheet (not just page), and type your name at the top of each sheet.**
4. Homworks will be **due at the beginning of class on the day of the deadline. No extensions.**
5. Each problem has the same grade value as the others (10 points each).
6. Unless otherwise specified, the only resources you can avail of in your HWs are the provided course materials (slides, textbooks, etc.), and communication with instructor/TA via discussion forum and e-mail.
7. You can discuss lecture concepts and the questions on Piazza and with your friends, but you cannot discuss solutions or ideas. All work must be your own.

**Prologue:** After the success of your Mars mission and your return to Earth, you receive many offers of employment at top universities as faculty (CS or ECE departments only!), in industry and in government. But before you get hired, you have to go through interviews at these places. Your goal is to attempt 8 interviews and ace them so that you have the maximum choice of where you want to (After all, if you’re attached, the choice will likely be made by your significant other, so you want to give her/him the most choice! If you’re not attached, you want to have the most choice anyway, right? ;).

The storylines, statements, events, and games in this homework are purely fictitious. Any resemblance to persons, places, or events, living or dead, past, present, or future, is purely coincidental.
Problems:

1. In your interview season, your first interview is at Brandeis University, where they are proud of logical timestamps because Leslie Lamport was one of their PhD alumni (true fact!). The faculty decide to play a game of chess with you, with their department head as the player opposing you. During the game, the head says he is not happy however that the key-value store Riak prefers to use dotted version vectors (DVV) instead of Lamport’s timestamps or vector clocks. They ask you to give two advantages and two disadvantages of DVVs versus vector clocks. Use the web to (find out and) write briefly about what Riak version vectors are and how they are used. Then answer the interview question. In your answer, be sure to include URLs/links pointing to specific features, otherwise you may not get points.

2. When you are visiting the University of Texas at Austin, four faculty play the Dijkstra self-stabilizing game with you (true fact – Edsger Dijkstra spent many years on the faculty at Austin). A ring of 5 processes \( p_0 \) through \( p_4 \) is using the Dijkstra’s self-stabilizing algorithm discussed in class. The ring is connected anticlockwise with \( p_i \)’s neighbor being \( p_{(i+1) \mod 5} \). At a given point of time, the \( x[i] \) values for \( p_0 \) through \( p_4 \) are respectively 0, 1, 3, 3, 0. You can assume \( K = 5 \). The faculty would like you to answer the following questions: (i) how many tokens are present in the above state, (ii) show one path of steps (by processes) that leads the above state to converge to a safe state.

3. Deciding that you want to interview at some exciting companies too, you decide to interview at Twitter. At the company, while playing ping-pong with the CTO, he brags to you about their new stream processing system called Heron and how it’s much better than Apache Storm. Give two advantages of Heron over Storm, and two disadvantages. You can use the web as a resource for this. (You must write about fault-tolerance). In your answer, be sure to include URLs/links pointing to specific features, otherwise you may not get points.

4. You cannot use the Web for this question. Your next interview is at Harvard University. Here they are proud of their alumnus George Kingsley Zipf (true fact!). The department head here asks you what is (i) a Zipf distribution, (ii) a heavy-tailed distribution, (iii) the difference between the above two i and ii, and (iv) give two real-world examples for each of i and ii.

5. You cannot use the Web for this question. At Berkeley, where they invented the Mica Mote (true fact!), they say they are building a sensor network on an African preserve to monitor movements of lions (and ensure that none of them are killed for sport). The sensors measure sound. The deployment spreads 2000 MICA motes over several hundreds of square miles. They would like to periodically
(every minute) measure the average sound across all your sensors. You have two options: either having the sensor nodes route all their sound measurements to a base station (routed via other sensor nodes) which in turn then calculates the average temperature reading, or have the sensor nodes talk to each other and calculate the average amongst each other. Answer the following questions:

a. Which of these two options would you choose? Give at least one major reason why you chose that option.

b. To calculate the average via a spanning tree among the sensor nodes, what data would you pass along (up the tree)?

6. Of course, you must interview at Google! Here they are proud (among other things) of having built the first engine for distributed graph processing, namely – Pregel. They point to Apache Giraph as the open-source version that is closest to Pregel. OTOH, during your Berkeley interview you learnt about Spark and its graph processing engine GraphX. Give four differences between Giraph and GraphX – two of these differences must be advantages of Giraph (over GraphX) and the other two must be disadvantages. In your answer, be sure to include URLs/links pointing to specific features, otherwise you may not get points.

7. When you interview at IBM, you find that they still love supercomputers and distributed shared memory. They ask you the following question. In a distributed shared memory system using invalidate, a process P3 wants to read a page. In each of the following cases, say what is the series of operations that needs to happen for P3 to be able to read (warning: there be tricks below!) -- You cannot use the Web for this question. If the setup seems wrong to you, you should point out ALL errors in it.

   a. P1, P2 and P3 are currently holding the page in Read mode, P2 is the owner
   b. P2 and P4 are holding the page in Read mode, and P4 is the owner
   c. P3 and P4 are both holding the page in write mode
   d. P1 is holding the page in write mode

8. During your exciting interview at MIT, you find that they too seem to like distributed shared memory. They ask you the following question. In a distributed shared memory system using invalidate, a process P3 wants to write a page. In each of the following cases, say what is the series of operations that needs to happen for P3 to be able to write (warning: there be tricks below!) -- You cannot use the Web for this question. If the setup seems wrong to you, you should point out ALL errors in it.

   a. P3 is holding the page in Read mode and P4 is holding it in Write mode and P3 is the owner
   b. P4 is the owner and is holding the page in a Write mode
c. P1 and P2 are each holding the page in a Write mode, and P4 is the owner
   d. P4 and P5 are each holding the page in a Read mode, and P4 is the owner

9. You also interview at the NSA (why not, eh?). They ask you the following questions. What is the difference between ACLs and capabilities? When would you use ACLs, and when would you use capabilities (one scenario for each is sufficient). You cannot use the Web for this question.

10. Your last interview is at Facebook. Facebook says they have built a new distributed file system, which implements (at its low level) Unix file system read/write-like semantics, i.e., internally it maintains file descriptor data structures which contain an automated read-write pointer (at the server side). Your friend claims this makes it easier to program with, since Unix programs can use the same API. You cannot use the Web for this question.
   a. Do you agree with this design? If yes, say why. If not, say what the problem with this design is (be specific)?
   b. How would improve this design given your knowledge of distributed file systems?
   c. (Optional, no points for answering, answer only if you want to) If you had offers from all the 10 places listed in this homework (faculty at universities, starting positions at companies and government agencies), where would you join?
   d. PS (not a question): If you’re wondering why Illinois/UIUC is not listed in the above places, it’s because we encourage cross-pollination and would like our alumni to spread their knowledge everywhere around the world (true fact!).