Incentives in Computer Science (COMS 4995-6): Exercise Set #7

Due by Noon on Wednesday, March 11, 2020

Instructions:

(1) You can work individually or in a pair. If you work in a pair, the two of you should submit a single write-up.

(2) Submission instructions: We are using Gradescope for the homework submissions. Go to www.gradescope.com to either login or create a new account. Use the course code MKRKK6 to register for COMS 4995-6. Only one person needs to submit the assignment. When submitting, please remember to add your partner's name (if any) in Gradescope.

(3) Please type your solutions if possible. We encourage you to use the LaTeX template provided on the course home page.

(4) Write convincingly but not excessively. You should be able to fit all of your solutions into 2–3 pages, if not less.

(5) Except where otherwise noted, you may refer to the course lecture notes and the specific supplementary readings listed on the course Web page *only*.

(6) You can discuss the exercises verbally at a high level with other groups. And of course, you are encouraged to contact the course staff (via Piazza or office hours) for additional help.

(7) If you discuss solution approaches with anyone outside of your group, you must list their names on the front page of your write-up.

(8) Refer to the course Web site for the late day policy.

Exercise 37

Spend some quality time with the Bitcoin blockchain (e.g., at `https://btc.com`, or a similar site of your choosing). Report on your findings. Describe anything you found interesting, surprising, or mysterious. For example, you could choose a statistic of interest (number of transactions, transaction fees, frequency of forks, amount of time between consecutive blocks, etc.), plot or otherwise summarize it, and speculate on why it may have changed over time.

Exercise 38

Recall that, in Bitcoin today, each attempt to authorize a block has a roughly 2^{-80} probability of succeeding. Look up the computing power of state-of-the-art CPUs and GPUs. Roughly how long should you expect to wait (on average) before a successful authorization if your using a CPU? What about with a GPU? Be sure to justify your answer, at least informally. (Feel free to count the application of SHA-256 to a proposed block as one operation.)

Exercise 39

Look up the definition of and motivation behind an exponential distribution. (It might help to also look up Poisson distributions.) Argue that, with only honest miners and fixed amounts of computational power, the time between consecutive blocks in Bitcoin should be more or less exponentially distributed. Under this distributional assumption, and given that the expectation of the distribution is 10 minutes, what are the 1% and 99% quantiles for the waiting time between blocks?

Exercise 40

For this exercise, a *blockchain* is a directed tree in which one vertex has out-degree 0 (the root vertex, a.k.a. the "genesis block") and every other vertex has out-degree 1, and in which every vertex has a (unique) directed path back to the root. The vertices of the tree correspond to blocks, and the directed edges to block predecessors pointers (recall in Bitcoin a block other than the genesis block must contain a hash of another block on the blockchain, its predecessor). Bitcoin uses the *longest chain* rule, meaning that the valid blocks are precisely those that correspond to the vertices on the longest leaf-root path in the tree. (It's not important for this exercise how ties are broken.)

An alternative approach (known as the GHOST protocol) is to define the chain of valid blocks inductively, starting from the root and ending at a leaf. The genesis block is always valid. Suppose v is a valid block and the vertices $w_1, \ldots, w_k$ point to it. Then exactly one of the w_i's is deemed valid—the vertex with the largest number of descendants (i.e., the largest induced subtree). (Again, for this exercise it's not important how ties are broken.)

Give an example of a blockchain for which the longest chain rule and the GHOST protocol's rule authorize different sets of blocks. Speculate on why the authors of the GHOST protocol thought their rule might be preferable to the longest-chain rule.

Exercise 41

Here's one way to implement stake-based sampling in a proof-of-stake blockchain. Let s_i denote the current balance of user i, and consider the following randomized algorithm:

- For each user $i = 1, 2, \ldots, n$: (in arbitrary order)
  - Flip a biased coin with probability
  $$\frac{s_i}{\sum_{j=i}^{n} s_j}$$
  of "heads."
  - If the coin comes up "heads," deem user i the winner and halt.
  - Otherwise, continue.

Prove that the distribution of winners produced by this randomized algorithm matches what we want for a proof-of-stake blockchain protocol: for every user i,

$$\mathbf{Pr}[i \text{ wins}] = \frac{s_i}{\sum_{j=1}^{n} s_j}.$$

[Hint: induction on i.]

Exercise 42

Recall that a Sybil attack involves a single entity creating multiple identities to manipulate a system. Both proof-of-work and proof-of-state blockchain protocols are unaffected by Sybil attacks.

Is a Vickrey (or eBay) single-item auction vulnerable to a Sybil attack? Argue why not, or alternatively come up with a concrete attack that you could plausibly implement on eBay.