Homework #1 Solution

Due: Fri, 22-January-2021, 11:59pm – Gradescope entry code: N8XV23

Please upload your answers timely to Gradescope. Start a new page for every problem. We strongly suggest LaTeX to type your answers. For the programming/simulation questions you can use any reasonable programming language (please no assembly, brainfuck, etc. ☹). Comment your source code and include the code and a brief overall explanation with your answers. A tentative point distribution (in % of the total) is provided in brackets. For most problems there is more than one valid way of solving them!

1. (30%) In this question we discuss the stochastic modelling of the mining times of Bitcoin.

   a) A reasonable model for the distribution of the time between two consecutive mining events is exponential. Write down this probability density function with specific choice of all parameters to model ideal Bitcoin.

   Answer:

   Mining consists of trying many nonces until a block \( b \triangleq (\text{prev}, \text{txs}, \text{nonce}) \) satisfies a hash inequality \( H(b) \leq \text{threshold} \). Here, \( \text{threshold} \in [0, 1] \) is the mining threshold and as of Jan 2021, \( \text{threshold} \approx 2^{-76} \). It is assumed that the output of \( H \) is uniform in \([0, 1] \).

   Trials in the PoW mining process can be modeled as independent Bernoulli random variables with success probability \( \text{threshold} \). The number of trials until the first success follows a geometric distribution, and as the success probability is very low and the number of trials very large, the block inter-arrival time is well approximated by an exponential distribution.

   The probability density function of an exponential random variable \( T \), with rate \( \lambda \) is \( f_T(t) = \lambda e^{-\lambda t} \). The mean of this exponential distribution is \( 1/\lambda \). (See https://en.wikipedia.org/wiki/Exponential_distribution)

   In Bitcoin, the average time between consecutive mining events is 10 minutes. This means \( 1/\lambda = 600 \) s. With \( t \) in seconds, the probability density function is

   \[
   f_T(t) = \frac{1}{600} e^{-t/600}.
   \]
b) What is the standard deviation of the inter-mining time under this model? What is the ratio of the standard deviation over the mean?

Answer:

For an exponential random variable with rate $\lambda$, both the mean and standard deviation are equal to $1/\lambda$. So, the standard deviation is 10 minutes, and the ratio of the standard deviation over the mean is 1.

c) What is the mean of the time it takes to mine 10 blocks? What is the standard deviation, and the ratio of the standard deviation over the mean? Be explicit about any assumptions you made to get this conclusion, and justify your modeling assumptions.

Answer:

Suppose the time taken to mine each of the 10 blocks, after their previous block was mined, is $T_1, T_2, \ldots, T_{10}$. We know that each of $T_1, T_2, \ldots, T_{10}$ has an exponential distribution with mean 10 mins. We can further make a reasonable assumption that $T_1, T_2, \ldots, T_{10}$ are independent random variables. This is because the hash values for different blocks are independent. Due to this independence,

$$\mathbb{E}[T_1 + \cdots + T_{10}] = \mathbb{E}[T_1] + \cdots + \mathbb{E}[T_{10}] = 100 \text{ mins},$$

$$\text{Var}(T_1 + \cdots + T_{10}) = \text{Var}(T_1) + \cdots + \text{Var}(T_{10}) = 10 \cdot (10 \text{ mins})^2.$$ 

The standard deviation of $T_1 + \cdots + T_{10}$ is $10\sqrt{10}$ mins $\approx 31.62$ mins. Thus the ratio of standard deviation over mean is $1/\sqrt{10} \approx 0.316$. Comparing with the ratio for mining a single block (above), this indicates that the time taken to mine a large number of blocks becomes relatively less spread around its mean.

d) Using data from [https://btc.com/block](https://btc.com/block), estimate the standard deviation of the inter-block mining time. Is it close to what your model in part (a) predicts? Explain if there is any significant discrepancy. State carefully how you perform the estimation and justify why you estimate this way.

Answer:

From [https://btc.com/block?date=2021-01-18](https://btc.com/block?date=2021-01-18) we can extract the mining time of 145 blocks. Taking the difference of successive mining times, we get $n = 144$
samples for the time between mining events: $T_1, \ldots, T_n$. First calculate the sample mean $\overline{T} = \frac{1}{n} \sum_{i=1}^{n} T_i$. The sample standard deviation can be calculated as

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (T_i - \overline{T})^2}.$$ 

For the blocks from 2021-01-18, the sample mean was 596.31 s, and the sample standard deviation was 629.99 s. Both the values are very close to 600 s, indicating that the mining times fit the exponential distribution of part a) very well.

2. (20%) The total hashrate of the Bitcoin network on January 1, 2018 was 14.4 EH/s.

a) Estimate the threshold in the hash inequality on that day from this fact. Compare this with the true threshold. Why might there be a discrepancy?

Answer:

14.4 \times 10^{18} \text{ hashes per second corresponds to } 8.64 \times 10^{21} \text{ hashes per 10 minutes. Since only one block on an average solves the hash inequality in 10 minutes, the threshold should be } 2^{-\log_2(8.64 \times 10^{21})} \approx 2^{-73}. \text{ This corresponds to } \approx 73 \text{ leading 0 bits or } \approx 18 \text{ leading 0 hex digits.}


A discrepancy could result from a mismatch in difficulty, or by chance the block’s hash is significantly smaller than required by the difficulty. In this case, the heuristic of estimating the ‘true threshold’ by inspecting block hashes fails. (For the proper way of obtaining the true threshold from the difficulty, see https://en.bitcoin.it/wiki/Difficulty. However, all that was expected for this problem was the heuristic used in class.)

b) Assume all mining was conducted using back then state-of-the-art Antminer S9 hardware, which delivers 14 TH/s at a power consumption of 1372 W. What was the power consumption of the Bitcoin network? Find a comparable country in https://en.wikipedia.org/wiki/List_of_countries_by_electricity_consumption

Answer:
Under the specified circumstances, the Bitcoin network would have used a power of \(1.4 \times 10^{18} \times 1.372 = 1.4 \times 10^9\) Watt. (This is comparable to a nuclear power station, \(https://en.wikipedia.org/wiki/List_of_nuclear_power_stations\))

From the referenced Wikipedia list, 1 kW-h/yr = 0.11408 Watt, so that \(1.4 \times 10^9\) Watt \(\approx 1.26 \times 10^{10}\) kW-h/yr. This corresponds roughly to the electricity consumption of North Korea.

Remark: As of Jan 2021, the energy consumption of Bitcoin is estimated (\(https://digiconomist.net/bitcoin-energy-consumption\)) to be \(\approx 7.5 \times 10^{10}\) kW-h/yr, which is comparable to countries such as Finland, Belgium, Venezuela or Austria.

3. (35%) According to Figure 7 of lecture note #2, the mining pool \(F2Pool\) has about 20% of the total hash rate of the current Bitcoin network.

   a) What is the distribution of the time we need to wait until a block is mined by \(F2Pool\)? Justify your answer.

   **Answer:**

   The rate at which \(F2Pool\) computes hashes is 20% of the hash rate of the total network. \(F2Pool\)'s mining successes are independent of the other mining that goes on in the network. Thus the time taken for \(F2Pool\) to mine their next block has an exponential distribution with rate \(\lambda = 0.2 \cdot \frac{1}{1600} = \frac{1}{8000}\). The probability density function is \(\frac{1}{8000}e^{-t/8000}\) and the mean is 8000s.

   b) What is the probability that the next block mined is from \(F2Pool\)? Justify your answer.

   **Answer:**

   From 3.a), the time taken for \(F2Pool\) to mine their next block has an exponential distribution with rate \(\frac{1}{8000}\). Since the rest of the network has 80% of the hash rate, the time taken for the rest of the network (except \(F2Pool\)) to mine their next block has an exponential distribution with rate \(0.8 \cdot \frac{1}{600} = \frac{1}{750}\). Define two random variables \(T_{F2} \sim \text{Exponential}\left(\frac{1}{8000}\right)\), and \(T_{\text{other}} \sim \text{Exponential}\left(\frac{1}{750}\right)\). Then
the probability that the next block is mined by $F2Pool$ is

$$Pr(T_{F2} < T_{\text{other}}) = \int_0^\infty \Pr(T_{F2} < t) f_{T_{\text{other}}}(t) dt$$

$$= \int_0^\infty \left(1 - e^{-t/3000}\right) \frac{1}{750} e^{-t/750} dt$$

$$= 1 - \frac{1}{750} \cdot \frac{1}{\frac{1}{750} + \frac{1}{3000}} = 0.2$$

Alternatively, every miner tries nonces randomly to mine a block, so each attempted nonce is independently and equally likely to mine the next block. Since $F2Pool$ performs 20% of the hashes, the probability that the next block is mined by $F2Pool$ is 20%.

c) At the end of Lecture 2, we consider an attack on the confirmation rule where a transaction enters the ledger as soon as its block enters the longest chain. Alice’s transaction is in a block $B$ and she is trying to mine two blocks in private before the honest miners are able to mine a new block on $B$, so as to remove the transaction from the ledger. Suppose $F2Pool$ is trying to perform this attack. What is the probability it will succeed? You can assume the attacker starts mining the private blocks as soon as the transaction enters the mempool, and that the transaction will be in the next block mined by the network. Make explicit any other assumptions you are making in deriving the answer.

**Answer:**

Consider that the block $B$ was mined by an honest miner (in this case, someone other than $F2Pool$). Out of the 3 blocks mined right after the transaction enters the mempool, one of them ($B$) is mined by the honest miners. If the attacker mines the other 2 blocks, the attack will be successful. This is because the attacker can build a private chain using these 2 blocks which will then replace $B$. The attack is shown in Fig. 1. The attack will be successful no matter in which order these 3 blocks are mined, as long as no other honest block is mined before the 2 adversarial blocks. Since there is a 20% probability that the next block will be mined by $F2Pool$, the probability that the attack will succeed is $3 \cdot (0.2)^2 \cdot 0.8 = 0.096$.

More generally, the adversary can replace their transaction even if $B$ was mined by the adversary. They first add the transaction to $B$ and make $B$ public to confirm the transaction, then make the private chain public, to replace the transaction. In this case, the adversary mines all 3 blocks after the transaction enters the mempool. The probability that the attack is successful (including this scenario) is $3 \cdot (0.2)^2 \cdot 0.8 + (0.2)^3 = 0.104$.  

5
4. (15 %) Using the data in https://www.blockchain.com/charts/ or elsewhere, estimate
the following quantities, in the past month on the Bitcoin network:

a) The average size of a transaction, in bytes;

b) the average size of a block in Mbytes;

c) the throughput in transactions per seconds (Tps);

d) compare this to the throughput of a system like Visa.

e) Roughly how many transactions are waiting to be confirmed on average?

f) Any changes in the charts that could be explained by the recent price fluctuations?

Provide a short description of how you derived the estimates and state any assumptions
you have made.

Answer:

The answers are based on 30-day averages up to 18-Jan-2021.

a) The average block size is 1.305 MB and the average number of transactions per block
is 2135. This means average transaction size is about $1.305 \cdot 2^{20}/2135 \approx 640$ bytes.
The average transaction size is reported as 763 bytes on https://charts.bitcoin.com/btc/chart/transaction-size#69zg. We have neglected the size of the block
headers, since they are only a tiny fraction of the block size.

b) Average block size: 1.305 MB

c) 323,972k transactions were confirmed in a day, which corresponds to 3.75 Tps. The
rate of transactions added to the mempool is 3.77 Tps, which is similar to the
confirmed Tps so that the mempool size remains somewhat stable.
d) The average throughput of Visa is around 1,700 transactions per second (based on a calculation derived from the official claim of over 150 million transactions per day in 2010\textsuperscript{1}. Different estimates of Visa's average throughput range from 1000 to 4000 Tps. (However, Visa is capable of handling even up to 65k Tps.)\textsuperscript{2}

e) The size of the mempool is about 56 million bytes. Based on the typical size of a transaction (and transactions do not vary widely in size), this corresponds to about 87,500 transaction in the mempool waiting to be confirmed.

f) The Bitcoin market price has risen rapidly since December 2020. This is accompanied by an increase in the total hash rate, indicating that the high prices may have motivated more miners to join the network. There is also a simultaneous increase in the network difficulty, in order to maintain the block mining rate at one every 10 minutes. This can be seen from the charts for market price, total hash rate, and network difficulty on https://www.blockchain.com/charts/.