CSDS-325 Computer Networks, proj1b

Nicolas Slavonia, njs140

November 2022

Summary

This article studies how current blockchain protocols may not be able to maintain stability or fail to scale when multiple peers participate in the network. The article then provides insight into modified versions of the blockchain that do provide scalability and stability with a main focus the Bitcoin network.

What is a blockchain?
The blockchain enables distributed consensus over a peer to peer network (P2P). Each peer mines (collects) new information for the blockchain. For cryptocurrencies, this information would consist of transactions on the network. The peer would then bundle this information together into a 'block', and send it across the network to other peers. Each block also has references to previous blocks in the network, which creates a 'chain' of blocks as each block references the previous block. To provide security, every peer agrees to trust a single chain of blocks. These blocks are referenced as 'confirmed', and every transaction in these blocks becomes official.

To define stability: "The blockchain system is stable if there are infinitely many times of consistency with at least one block arrival between subsequent times of consistency." This means as t $\rightarrow \infty$, there are infinitely many points in time where every peer in the network possesses the same blockchain.

To define scalability: "A blockchain system is scalable on H_k if there exists a non-zero block arrival rate $\lambda > 0$ such that, for every $k \in N$, the blockchain system is stable with rate λ on the P2P network H_k." This means that as the number of peers participating in the network grows, the P2P network maintains stability for $\lambda > 0$ (maintains stability as blocks are continuously mined).

The issue with stability is there is no guarantee of blocks being mined or everyone being fully up to date. No blocks being mined means all recent transactions have yet to be confirmed, halting the network. Scalability refers to the guarantee that the performance of the blockchain does not degrade as the number of peers grows. Since every peer needs to be updated on the most recent block, large delays in communication will confuse and slow the network down.

The article describes two methods of maintaining a distributed ledger which are both members of the Longest Chain Policy (the trusted ledger is longest chain). Tree Policy: Every newly mined block references the most recently confirmed block on the blockchain. This can be represented as a linked list where every new block points to the head of list. Once confirmed, the new block becomes the head of the list and the cycle continues. Both Bitcoin (BTC) and Ethereum (ETH) use the tree policy.

A confirmed block is a block on the network that 1) every peer on the network has acknowledged and 2) every block mined after it possess a path to it (i.e. in the linked list of blocks). A trusted block is a confirmed block that has X number of confirmed blocks mined after it. The amount of blocks required to produce a trusted block is up to the implementer. For BTC, 6 confirmed blocks are required to make a trusted block (BTC's λ equals 10 minutes, i.e. a block is mined every 10 minutes on average).

The article then emphasizes the importance of one-ended blockchain protocols. A one-ended network is a network that is able to produce infinitely many blocks as t $\rightarrow \infty$. Basically, the network is stable and scalable for all t. The article explains how both the tree policy and throughput optimal policy are one-ended.

The article then analysis the bounds for a stable network as a function of the block arrival rate. They provide mathematical insights on calculated a networks conductance. The higher the conductance, the less information is held up in bottlenecks.

The article then explores the scalability of different types of network structures using the knowledge gained so far. Using analysis for conductance and the previously discussed bounds for stability, the article proves regular grids, regular trees and random geometric networks are not scalable while Erdös-Rényi networks, Random d-regular networks and Preferential Attachment networks cannot be determined.

We then arrive to the system insights. Using the knowledge presented so far, the article provides five different metrics to analysis a blockchain network. The first is "Time to Consistency". This is the time it takes for a newly mined block to be acknowledged by everyone on the network.
"Cycle Length" is the mean of the time a consistent network remains consistent and the time it takes for an inconsistent network to become consistent.
"Consistency Fraction" is the expected fraction of peers who have acknowledged every block currently on the blockchain. These peers can be seen as consistent between each other because they each share the fully updated ledge. Because of the longest chain policy mentioned above, any peer who is not fully updated will be wasting resources because they can not reference the newest block.
"Growth Rate of the Distinguished Path" characterizes the rate at which blocks are confirmed and/or trusted.
"Age of Information" This represents how far off peers are from being fully consistent. An value of 0 indicates a peer is fully updated, while a value of 2 means

that peer is missing the 2 most recent blocks.

Chapter 8 in the article provides synthetic and real data showing how the metrics above can be used to help blockchain protocols. The synthetic data is produced by creating three P2P networks while analyzing the traffic across them. They very the block arrival rate to monitor the 5 metrics. Through graphs, we can see the following outcomes: Time to consistency grows monotonically with block arrival rate and the lower the number of peers, the smaller the time to consistency.

The cycle length goes to ∞ at both 0 and μ (the upper and lower bounds for λ). The cycle appears to be convex and is flat in the middle. This allows the implementer to adjust the arrival rate without consequence.
For the consistency fraction, the more the peers the lower the fraction. Around a fraction of 1/2, the functions changes from concave to convex.
The growth rate of the distinguished path appears to have a maximum in the graph while both ends approach to 0.
Lastly, the age of information graph grows monotonically with block arrival rate.
The real data is measured from the BTC network. These measurements use the same metrics (except for the growth rate metric). This data is able to show the correlation between the BTC network and the Poisson distribution. The data also shows that BTC is currently stable, but runs at risk for issues with scalability.

They conclude by referencing other related work to explore similar ideals.

Analysis

This article provides a lot of mathematical logic to prove their work.
They provide a lot of derivations to prove the limits for λ. They start off with definition 5.1 which states: "Let H be an undirected network on a vertex set V and let $S \subseteq V$ be a subset of vertices. The conductance $\phi(S)$ of the set S is given by:"

$$\phi_H^{(S)} = \frac{\Sigma_{p \in S, q \in S^c} \frac{1}{d(p)} \mathbf{1}_{pq}}{\frac{1}{N}|S||S^c|}$$

They use this equation to analyze how much information propagation is affected by bottlenecks. They then expand the the equation to represent the limits of λ.

$$\frac{inf_{S \subseteq V} \phi_H^{(S)}}{2log(N)} \ \leq \ \mu \ \leq inf_{S \subseteq V} \phi_H^{(S)}$$

for:

$$0 \ < \ \lambda \ < \ \mu \ < \ \infty$$

This shows that a guaranteed stability condition is $\lambda < \phi_H$ and that the true stability region is bounded by μ. This is really important for this article because they analyze the limits of their own model. They show where there own model fails and provides us with their proof. This provides information to the reader that can be used to compare different models.

They also provide proofs for every single claim they make. Page 29 to the end are all proofs that what they claimed is true. This is important because this shows the model is based off of claims that are true which minimizes possible future errors. They want to show that the model they designed is mathematical and logically true.

On top of all their proofs, they even provide the data from synthetic and real experiments. The articles goal is provide new methods of analysis for blockchain models. They show how their models work in the field to provide examples to the reader of their work.

They provide two experiments. The first is synthetic. They created their own P2P network and modify the block arrival rate (λ) for different numbers of peers (N). They show how their 5 methods perform in the 5 graphs given. They show their models to the reader. This allows the reader to understand what their models provide and how to use them. The data also shows where the models fail, which again gives information to the reader of the limits of the models.

The real data is an analysis of the different topologies. This data is taken from direct measurements from the Bitcoin network. This data provides three very important things. The goal here is to provide the reader with an understanding of how their models behave under different topologies. This is very important for this article because they proved above how the limits on λ are a function of number of peers N, and the topology of the network H. Having two sets of data that show how they model functions under all parameters is extraordinarily important for the reader as we can now know what to expect in the field.

The second important thing here is they show how under their models, the BTC network behaves similarly to the Poisson distribution. This ties directly into future work (which I explain later) because there is a lot of work studied on the Poisson distribution for probability (I even studied this in my Math 380 class). If their models make BTC resemble the distribution, all that work can be applied to the BTC network.

Finally, the article provides the valuable information of showing how their models work in a controlled environment vs an uncontrolled environment. Although the data measured is different for the two experiments, combined they help the reader get a larger sense of their models.

As a final note in the article, they mention other peer's related work. They quickly explain what others have published, and briefly explain the differences between the models. They reference 4 different topics, which can summarize the work done in this article. They talk about different P2P networks, different methods of making a blockchain, other blockchain models and measurements and statistical information. They provide the reader with a broader perspective of how their work compares with others. This provides the reader with even more information on how to modify a network to make it more stable. This shows the authors are focused on providing the reader with an in-depth understanding of how to make a network safe.

These models are very useful to the real world. The article does not provide the reader with an ideal network, rather information to help a given network scale. This means these models can be applied to every blockchain system. They will provide the implementer with information on what may need to change in their system. Bitcoin currently has a market cap of over $300 billion so these models can act as assurances that the network can grow. They can also show the limits of the model so that people can prioritize using a network that is safe.

Blockchain Networks, usually implemented for cyrptocurrencies, need to maintain stability otherwise investors may lose their money. Unfortunately, most cryptocurrencies fail due to fraudulent activities, most recently, FTX. However, this does not mean a coin won't fail due to network failures. BTC recently implemented "Bitcoin Lighting" which aims to minimize network congestion and lower delays. These models can easily be useful to the implementers to help understand the limits of the network. This can provide safety and trustworthiness to the investors which may make the coin more valuable. If a network begins to approach the limit, the implementers can be aware and work to fix the network.

Some of the limitations of the article are how they mention that this model is limited to bandwidth (it is good they mentioned this for the reader). They do not analyze their models over different bandwidths. This may pose an issue because the reader may be limited by the bandwidth over the network rather than the block arrival rate. This article does not mention how to adjust the models for cases like these. Although this is an issue, the article does mention how bandwidth is not their focus. This just means more experiments should be conducted on top of these models.

The last skeptical point is how the data generated may not represent every blockchain. The synthetic data monitored different block arrival rates for 3 different P2P networks. The three networks had 10, 20 and 30 peers. Although they analyze the trend as N grows, 30 nodes may not be able to represent larger networks. For example, BTC is estimated to have over 1 million miners. They don't inform the user on how the models may change for much larger networks. Although they study BTC in the real data, they don't conduct the same experiments.

The real data is only focused on BTC which has a fixed block arrival rate of 10 minutes (averages 10 minutes) regardless of the number of users. Because of this, they weren't able to measure their 4th model (growth rate) on BTC. This forces them to run different experiments (although the new experiments do have many benefits). The reader is left in the dark on how to fully connect the two experiments. The authors should have provided more experiments or context to show the correlation.

Future Work

The article does mention a possible branch of future work that someone could do. As mentioned above, blockchain technology wraps a bunch of transactions into a block and passes blocks around the network. This article only analysis the impact of blocks on the network. They mentioned how future work could be used to run similar sets of analysis at the transaction level. This would provide information on the limits number of transactions per second (tps). The analysis could show how the limits of the transactions affect the limits of block arrival rate and vice versa. For example, this could be very useful for Bitcoin because since the block arrival rate is 10 minutes, the transactions per second stable up to 7 tps. This, combined with the analysis above, would provide lots of information regarding the scalability of a network.

Another point they brought up was the Poisson distribution. The Poisson distribution is a discrete probability distribution that gives the probability of a given number of independent events occurring in a fixed interval of time. The models show how BTC can resemble such a distribution. Future work can take the analysis on the Poisson distribution and apply it to BTC and other blockchain networks. Such as, what is the probability the block rate increases or decreases? Number of nodes, tps, λ, increases/decreases? What would the models show then? We could find the 95% confidence intervals for these networks and models. These values can tell the implementers what to expect and what to prepare for.

All these models are trying to show the limits of a blockchain network. I feel that more work should be done on accounting for more variables. Such as temporal congestion in a network, latency adjustments, servers crashing, spikes in the network, bandwidth and more. I think adjustments should be made on the bounds of $\lambda$to take all these different things into account. This would provide the reader with bounds that can be applied to all sorts of unexpected situations.