A model for Bitcoin's security and the declining block subsidy

Hasu James Prestwich Brandon Curtis

October 2019

We are grateful to Anthony Towns, Arjun Balaji, Brian Venturo, David Vorick, Joe Kendzicky, Lucas Nuzzi, Matthew Hammond, Nic Carter, Philip Daian, Steve Lee, Su Zhu, Tarun Chitra, and Yassine Elmandjra for editing and many valuable ideas. The paper was further inspired by the excellent papers, blog posts, and tweets of Andrew Miller, Arvind Narayanan, Ed Felten, Elaine Ou, Emin Gun Sirer, Eric Budish, Eric Lombrozo, Eric Voskuil, Fernando Nieto, Ittay Eyal, Joseph Bonneau, LaurentMT, Nick Szabo, Paul Sztorc, and Raphael Auer.

Contents

An application or protocol is secure if it realizes its goal in an adversarial environment. In the case of Bitcoin, the goal is to establish a payment system where anyone can participate, only the rightful owner can spend a coin, and all valid transactions make it into the ledger eventually.

For the first ten years of its existence, Bitcoin has successfully held these security properties. At the same time, academia has largely failed to replicate Bitcoin's empirical soundness in their models, giving birth to the meme that "Bitcoin is secure in practice, but not in theory." With this paper, we want to bridge the gap between theory and practice by introducing our model of Bitcoin's security.

We show that Bitcoin can currently tolerate a very high incentive to attack, formalizing the intuition that the incentives of miners are long-term aligned with the system. The key insight is that mining requires large upfront investment whose value is tied to the health of the network. Basically, miners buy half of all coins they expect to mine over the next two years in advance before they can even start mining. Any behavior that hurts the value of these coins before they have been delivered would be highly destructive, showing why many of the attacks feared by academics are indeed irrational in practice.

In the second half, we show that, more than any external attacker, the biggest threat to Bitcoin's security is baked into the protocol itself. The block subsidy schedule, which declines as part of Bitcoin's fixed emission schedule, will lead to lower commitment from miners. If a robust blockspace market doesn't develop, we explain why a decline in block rewards poses a substantial risk for the future. Contrary to popular belief, users can't compensate for this by simply waiting for more confirmations. Finally, we present general ways to think about the problem, including several possible improvement proposals for the community to discuss.

1. Why Bitcoin needs mining

In past payment systems, a trusted central server or group of servers was required to order transactions. That turned out to be a crucial point of failure, as central validators routinely fail or are compelled to exclude certain groups of people or certain types of transactions. Hence, a system aiming to provide permissionless access cannot afford the use of a central party. Satoshi saw the solution in replacing the popular client-server model with the flat peer-to-peer model that had proven its worth in extremely resilient distributed networks like BitTorrent.

Proving and verifying ownership of messages had already been possible, thanks to public-key cryptography. In Bitcoin, the owner of a coin can sign a message with his private key. Other nodes in the network can then verify the message using the sender's hashed public key to prove that it is indeed valid. That satisfies the "safety" requirement in the Bitcoin system. However, public-key cryptography is of little help when a node receives two conflicting messages that are individually valid but can't be both valid at the same time, e.g., when someone tries to spend the same coin twice.

Bitcoin elegantly solves this problem by replacing the signature of a trusted server with a group signature of computational power that nodes can follow to coordinate on a single chain. The nodes can put a high degree of trust in this signature because it is costly to produce, and this cost can be easily verified. When nodes receive two conflicting signatures from miners, they discriminate between them by favoring the more costly one. This "fork-choice-rule" is now known as Nakamoto consensus.

The idea to think about Bitcoin mining as a dynamic multiparty-membership signature (DMMS) was first pioneered by Back, Corallo et al.[1] A DMMS is a signature made by a variable and anonymous set of signers who can enter and leave at any time. Their share of the computational power pointed at the Bitcoin network weights their contribution to the signature. These signatures are cumulative because each block references the previous block, creating a blockchain.

The computational signature is created as follows: first, miners perform dead-weight computations by generating random outputs. When these outputs fall into a specific range, other nodes can take that as proof that the virtual dice must have been rolled a certain number of times on average (similar to how a 1000-sided dice, on average, must be rolled 100 times until a number between 1 and 10 comes up.[2]) Next, a miner publishes her block, including the proof-of-work, to the rest of the network. If it satisfies consensus rules, other nodes add it to their blockchain and compensate the winning miner with the block-reward and all transaction fees in that block.

1.1 The limits of cryptography

While miners have certain freedoms in constructing their blocks, they can't give themselves more coins, steal someone else's coins on the same chain, or even change the proceeds of

[1] https://blockstream.com/sidechains.pdf

[2] Beyond the doomsday economics of "proof-of-work" in cryptocurrencies https://www.bis.org/publ/work765.htm

a block retroactively. Miners have to follow the Bitcoin protocol like any other node, and nodes automatically reject any attempt to break the protocol.

However, there are important aspects the protocol cannot enforce cryptographically. A node doesn't know which of two conflicting transactions is valid, or which of two competing chains to prefer, so users depend on the fork-choice rule to coordinate on a single chain. While the fork-choice rule is required for Bitcoin to stay in consensus, it also gives miners considerable power that is not regulated (and not regulatable!) by the protocol itself.

The most famous "incentive failure" is the double-spend attack, where a majority miner first buys non-BTC goods or services using BTC on the original chain. Once he has taken irreversibly delivery of the goods or services, he produces a longer chain where that transaction never occurred, to end up with both the money and the goods. Nodes, diligently following the most costly signature, automatically switch to the new chain even when it contains off-chain theft or other malicious actions.

From that, we can see that "hard" protocol rules like cryptographic signatures cannot completely secure transaction ordering - it also depends on "soft" economic incentives for miners to publish updates that serve Bitcoin users.

2. Modeling Bitcoin's security

If users can't trust the protocol to enforce the "right" transaction history, how can they know if a transaction is final or could be reversed by miners in the future? In the traditional financial system, a transaction is final because reversing it is forbidden by law. In Bitcoin, the law has no reach over miners, who can be anonymous, operate from anywhere in the world, and join and leave the network at any time.

If it was profitable for miners, we would expect them to undo transactions all the time, including transactions of other people who pay them to do so. Hence, users should not regard a payment as final unless it is unprofitable to reverse. Folk wisdom has expressed this question as "how many confirmations one has to wait until a payment is final." We show why additional confirmations do not add meaningful to Bitcoin's security. Security, instead, is mainly the result of two simple factors.

2.1 Security assumption

We start by setting up a basic payment system with a block subsidy of 12.5 BTC and no fees. All of the hardware and hashpower required to mine can be rented on-demand, so miners have no long-term commitment to the Bitcoin network. Their actions do not affect the exchange price of BTC, and no users ever ignore the costliest chain chosen by Nakamoto consensus. All models use BTC as the base currency.

We define the BTC value of following the protocol, or “honest mining” as **EV(honest mining)**.

Over an example **duration** of ten blocks, **miner revenue (MR)** would be 125 BTC. Assuming free entry to mining and perfect competition between miners, we can expect the whole of them to expend 125 BTC in **mining cost (MC)** to earn this reward.

Equation 1	**mining revenue (MR)** - **mining cost (MC)** = 0

Equation 2	EV(honest mining) = MR - MC

The EV of honest mining is thus established as a baseline of 0 BTC.

Miner-extractable value (MEV) describes how many BTC a miner hopes to win from his attack. The concept was invented by Daian, Goldfeder et al. to describe value extractable by miners from smart contracts[3], but we expand it to cover any value extractable by miners from manipulating consensus or transaction ordering.

Importantly, MEV does not describe how much an individual user can safely exchange in one block, because the attacker could double-spend many different parties at once. It does not even describe how much all users together can safely exchange in one block, because the attacker could double-spend across several blocks in a row. MEV describes the entire value for the attacker, which, against users who wait for six confirmations, already has a minimum attack duration of seven blocks. Hence, users who compute MEV only based on their own individual transactions would underestimate the real incentive for miners by orders of magnitude.

The final EV of attack mining, e.g., to double-spend, can then be modeled as:

Equation 3	EV(attack mining) = **MEV** + MR - MC

[3] Flash Boys 2.0 https://arxiv.org/pdf/1904.05234.pdf

A rational miner will follow the protocol instead of attacking it as long as *EV(honest mining)* > *EV(attack mining).*

We can thus derive that ***EV(honest mining)* > *EV(attack mining)* is the necessary condition for Bitcoin to be secure against rational attackers.**

It follows then that the difference between *EV(honest mining)* and *EV(attack mining)* describes Bitcoin's tolerance against an irrational ("byzantine") attacker, who is not concerned with profit but will attack Bitcoin for arbitrary reasons. Notably, this tolerance does not have to include value the attacker extracts as a direct consequence of the attack, e.g., from making a large bet against the price of Bitcoin. MEV already captures any such value.

In this simple model, we need not even talk about a byzantine attacker. The system already fails against a rational one, as any *MEV* > *0* is enough to make attacking more attractive than mining honestly. Assuming that a miner can extract 100 MEV from an attack that lasts 10 blocks, we can see that

Example 1: EV(attack mining) = MEV + MR - MC = 100 + 10 - 10 = 100;
100 > 0, thus Bitcoin is insecure

This finding is in line with intuition because attacking the chain has no actual cost for the attacker; it has only a budget requirement of 10 BTC. Any resources he spends on the attack he recovers after the attack has been successful. There are three notable caveats:

1) If the attacker has to invalidate some of his own blocks, the attack starts to have an actual cost, because his effective MR(attack) declines while MC stays the same.
2) If a minority miner ("defender") continues mining the original chain, he can increase the duration of the attack. But as long as the attacker eventually catches up, this doesn't decrease his EV; it only raises the budget requirement. The defender's resources will be wasted.
3) In this model we assume the attacker has either a hashpower majority or coordination between several smaller attackers is costless. In the real world,

coordination has a cost that can increase if miners disagree about the value of MEV or the necessary duration of the attack.

2.2 Market governance

According to a popular saying, as economic actors, we cast votes all the time - by spending money on some things but not others. Blockchains are markets as well, so when users (consumers) buy and sell BTC, they constantly vote for miners (producers or service providers) to act in a certain way. When users are unhappy about the service offered by miners, confidence in the payment system could plummet, and the exchange price of BTC might fall compared to before the attack.

We define **p(postAttackPrice)** as the relative BTCUSD price after the attack, e.g., a postAttackPrice of 95% means that the price fell by 5% from the attack.

Equation 4	EV(attack mining) = **p(postAttackPrice) *** (MEV + MR) - MC

In the updated equation, both the MR (the block rewards + fees) and the MEV get smaller as the BTC price falls as a result of the attack, while MC(attacking mining) stays the same. While using BTC and not fiat as the base unit here may be uncommon, we find it easier to reason about. In reality, a miner does not have less nominal BTC post-attack, but because they lost 5% of their purchasing power, he could only exchange them for 95% pre-attack BTC.

Due to the introduction of market governance, EV(attack mining) is now unprofitable as long MR(honest mining) is larger than p(postAttackPrice) * (MEV + MR(attack mining).

Equation 5	EV(attack mining) < 0, if	MR > p(postAttackPrice) * (MEV + MR)

From that, we can derive three ways the system could be secure:

1) **MEV could be low**, e.g., because very few people transact in Bitcoin, or users don't consider payments final without additional assurances like knowing a buyer's identity.
2) **p(postAttackPrice) could be low**, meaning that users have very sensitive to what Bitcoin is supposed to do and are willing to switch to a competitor if miners stop doing their job. This is somewhat of a "pick-your-poison" parameter, because if crashing the

price of BTC is easy, other forms of attacks (like sabotage) become more attractive, thereby increasing MEV.[4]

3) **MR could be high**, so the impact from p(postAttackPrice) on MR starts to exceed the potential gain from MEV.

2.3 Miner commitment

Until now, we made the unrealistic assumption that everything needed to mine blocks can be rented on demand (a view which dominates academic reviews of Bitcoin security.) In practice, mining is nothing like that. In a model of strong competition, miners are running on a treadmill. If one miner speeds up and increases his revenue at the same cost, others must follow pace or risk falling off entirely. Mining has very few sustainable moats. As a result, the mining industry has industrialized faster than maybe any other industry in history.

As mining industrializes, the unit cost of finding a block starts to matter more and more. There are several ways to lower the unit cost in a business:

1) If production facilities operate below capacity, the business can sell more units to average their overhead expenses across a greater number of items. In mining, every hash has an automatic buyer in the form of the Bitcoin network, so there is little to optimize here.
2) The business can reduce the ongoing material costs of production. The mining equivalent would be continuous search for cheaper sources of energy, better access to heat dissipation or cooling, and manufacturing optimizations.
3) The business can reduce its overhead costs by specializing its production facilities. In Bitcoin mining, this resulted in hardware becoming ever-more optimized for one job: hashing SHA-256. The moment this hardware can no longer mine Bitcoin, it is effectively worthless. Notably, this is even for large GPU mining networks like Ethereum. Even though Ethereum can be mined using general-purpose hardware, there is not nearly enough demand for GPUs to saturate a sudden large increase in supply. Should the price of Ethereum collapse, Ethereum miner commitments would lose most of their value as well.
4) Miners can also lower their cost per unit of energy by entering into longer-and-longer power purchasing agreements (PPAs.)

Hence, to lower the unit-cost enough to even start mining competitively, a rational miner requires highly specialized hardware and needs to adopt a long-term view of the network.

[4] Although, if the price is known to be vulnerable to attack, derivatives markets should start to price this in and make short-selling more costly.

The more a miner specializes, the more the non-repurposability of his assets and expenses increases. From Equation 1, we know that MR + MC = 0. That means, we can derive the total costs of mining from the total revenue of mining, which is simply the sum of all block rewards.

How much of that cost do miners have to commit in advance? After talking to Bitcoin miners and experts, we came up with a rough estimate that the average miner, and thus the mining industry as a whole, has about 50% of their total costs tied up in such non-repurposable assets. Further, we learned that these assets depreciate, on average, over 24 months.

If we run with this assumption, then the mining industry as a whole has one entire year of block rewards (two years * 50%) committed to mining Bitcoin for the next two years. At a block reward of 12.5 BTC, that amounts to 658,800 BTC.

In other words, **miners have to buy 50% of all coins they expect to mine over two years *in advance* before they can even start mining.**

Anything that jeopardizes the value of these coins before they have been delivered is highly destructive for them.

We can thus say, **miners are strongly committed to mining Bitcoin in a way that maximizes the value of BTC and the utility of the network**.

Equation 6	EV(attack mining) = p(postAttackPrice) * (MEV + MR) - MC - **[1-p(postAttackPrice)]* commitment**

In the first example, where hashrate could still be rented, a p(postAttackPrice) of 95% affected MR only for the attack duration 10 blocks. **Once miners are Bitcoin-committed, the same price drop affects an entire year of revenue - 52,704 blocks!** A 5% price drop would now wipe out the equivalent of 32,940 pre-attack BTC across all miners.

Notably, an attacker needs not to own 100% of the hash power for his attack to succeed. If he attacks with 60% of the hash power, his own commitment would be merely 60% of the total commitment, which is 395,280 BTC.

Example 2: EV(10 block attack with 60% hash power and 100 MEV) = 95% * (100 BTC + 10 * 12.5 BTC) - (100 * 12.5 BTC) - 5% * 395,280 BTC = -19.675 BTC

For an attacker with 60% hashrate, MEV would have to be around ~21,000 BTC, or $187m at today's prices, for the attack to be profitable[5]. The high tolerance for MEV indicates that the Bitcoin network is indeed secure today. These findings can be generalized to all cryptocurrencies that use PoW and show how hugely important it is for security that miner expenses are non-repurposable.

2.4 Suspending Nakamoto consensus

We have shown that the Bitcoin network can tolerate a large amount of MEV today, creating a high barrier for attacks to be profitable. However, to complete our security model of Bitcoin, we need to update the final remaining assumption that Bitcoin users never question Nakamoto consensus.

Users are in the market for a trust-minimized signal that allows them to coordinate on a single chain. They spend great amounts of money for these signals because doing so is cheaper than coordinating any other way - e.g., by talking directly to each other until a similar consensus emerges.

However, that doesn't mean users are bound to follow the signals produced by miners even when a majority of users are unhappy with them. There is plenty of precedent in Bitcoin's history where users have ignored Nakamoto consensus because the resulting chain no longer represented the social contract they had signed up for.

In 2010, an integer overflow bug in block 74,638 caused the creation of a whopping 184 billion BTC, many times more than the 21 million that is supposed ever to exist. Within three hours, Satoshi had published a new Bitcoin client without the bug that "rewound" the hyper-inflated chain.[6]

A second example is the 0.7/0.8 consensus bug in 2013 that split the blockchain in two for several hours. Bitcoind, the most popular Bitcoin implementation at the time, had recently released its 0.8 update. Unbeknownst to the developers, the new software also had a small, unintended change to the consensus rules that caused block 225,430 to be incompatible with older clients. The fork was resolved when Bitcoin developers and mining pools decided to suspend the fork-choice rule temporarily. They manually supported the 0.7

[5] This number represents a lower bound because the fact that in proof-of-work all miners are getting punished collectively can create some interesting dynamics. The other 40% miners who are not part of the attack still have a huge commitment to the network and are incentivized to defend it. However, we can only speculate how exactly that would play out.

[6] https://hackernoon.com/bitcoins-biggest-hack-in-history-184-4-ded46310d4ef

fork and abandoned the 0.8 chain, which required miners to forgo any block rewards from the 0.8 chain to maximize the overall utility of the network.[7]

Finally, the most well-known example may be the UASF movement of 2017. A full year after the code had been released, the majority of miners were still refusing to adopt the Segregated Witness update - possibly because it broke ASICBoost, a patented technology that increases the efficiency of particularly mining hardware[8]. To push this change through anyway, some Bitcoin users installed a client that threatened to, once again, suspend Nakamoto consensus by ignoring blocks from miners who refused SegWit after a certain date. Had miners let this play out, it would have resulted in contentious fork from the main network. The threat to Bitcoin's utility and value was serious enough to the miner's bottom lines that they finally gave up their resistance against the SegWit update.

These examples highlight that ultimately, users lead and miners follow. When they disagree over what governance decisions would maximize overall network utility, users can run custom code like the invalidateblock parameter to temporarily suspend Nakamoto consensus and thereby "disempower" miners.

Attackers must consider the risk that users reject their chain even though it satisfies the protocol rules.

We define p(followNC) as the probability that users coordinate off-chain to suspend Nakamoto consensus. From the attacker's perspective, this further decreases the potential reward while his costs stay the same.

Equation 7	EV(attack mining) = **p(followNC)** * p(postAttackPrice) * (MEV + MR) - MC - [1-p(postAttackPrice)] * commitment

Because it affects only MR and MEV for the duration of the attack but not the miner commitment, NC-suspension adds less to security than market governance. However, users can, in theory, change not just the transaction history but core protocol rules as well. If there was consensus to change the mining algorithm from SHA256 to something else, users could at once invalidate the entire miner commitment, even without the Bitcoin price collapsing to zero. That makes social intervention a very useful defense against attackers who actively try to lower Bitcoin's price or otherwise sabotage the network.

[7] https://bitcoinmagazine.com/articles/bitcoin-network-shaken-by-blockchain-fork-1363144448

[8] It might be interesting to explore the motivations of miners in the context of our commitment model.

2.5 Summary

By building out this model and populating it with real numbers, we were able to gain a couple of key insights.

1) For a high degree of security, honest mining must be more profitable than attack mining over any duration that users consider final.
2) If users want to be able to make large transactions, MEV must be allowed to be high.
3) The ability of the system to tolerate a high MEV depends on the size of the punishment miners take from acting maliciously. Users can punish miners in two main ways:
    a) First, they can sell some or all of their bitcoin. When the exchange price of BTCUSD falls by 10%, miners lose 10% of their commitment value in pre-attack BTC.
    b) Second, users can coordinate off-chain to suspend Nakamoto consensus temporarily.
4) For the potential for punishment to be large, the miner commitment must be large, and the willingness of users to sell coins must be high.
5) The size of the miner commitment is a function of miner revenue (MR), the share of commitment costs of total costs, and their depreciation schedule.
6) If we keep the share of commitment cost, the depreciation schedule, and the willingness to sell coins constant, MR is the determining factor for MEV tolerance and thus how much user activity the network can support.

We invite anyone to download and experiment with our model on their own terms.[9]

3. Mining attacks

Next, we want to know how the most prominent attacks on the Bitcoin system would play out according to our model.

What attacks are possible on the Bitcoin network strongly depends on how much hashpower the attacker has. In theory, a miner with as little as 30% hashrate can engage in

[9] https://docs.google.com/spreadsheets/d/1b6-BtD_sd7x5k3-nDrR-I139nINsotiMH43CAq58YOM/edit?usp=sharing

practices like selfish mining or stubborn mining, which depend on strategically withholding blocks to earn more than the fair share of the miner revenue. To the best of our knowledge, these tactics have not been spotted in Bitcoin to date. Our model suggests that it is indeed irrational for miners to engage in tactics that could decrease the public's trust in Bitcoin, as even a small price decrease would destroy the value of their commitment more than they can hope to gain in MEV.

There is at least one data point in favor of this theory. In 2014, the GHash.io mining pool repeatedly flirted with >50% hashpower (by attracting miners with its zero-fee policy), and allegedly even engaged in double-spending the popular betting site BetCoin Dice[10]. As the news about mining pool centralization propagated through the Bitcoin community, trust in the system was shaken. Several important figures publicly sold parts of their Bitcoin[11].

In the aftermath, individual miners fled the pool in large numbers to protect their investment. After that, no mining pool has dared to come anywhere near this level of hashpower again. It seems miners became aware that any form of market panic can have a highly detrimental effect on their bottom lines.

Here we can see the divergence between the byzantine and the rational model: under a byzantine model, Bitcoin is insecure once a miner has >50% hashpower. However, the steady state of Bitcoin in a complex world might well be a hashpower monopoly. There could be a monopoly right now, and we have no way to disprove it. Looking at the incentives of all actors can show that Bitcoin doesn't automatically fail in the presence of a majority miner. Users can still shape the incentives of that miner to build the blocks they want,

When a miner has over 50% of the hashpower, he gains the certainty that whatever chain he proposes will eventually become the canonical chain in Nakamoto consensus. That certainty is a precondition for the more severe attacks on Bitcoin users. These attacks can be separated into two categories: double-spend attacks and sabotage attacks.

3.1 Double-spend attacks

In a double-spend attack, an attacker reorganizes a chain where he made a substantial purchase in BTC to replace it with a version where he still owns the goods, but never paid for them.

Our models have shown that a small decrease in the price of BTC can make even large double-spend attacks unfeasible because the gain from MEV must be higher than the

[10] https://bitcoinmagazine.com/articles/mining-2-1403298609
[11] https://www.reddit.com/r/Bitcoin/comments/281ftd/why_i_just_sold_50_of_my_bitcoins_ghashio/

damage to the miner commitment. Further, the miner has to be concerned that users suspend Nakamoto consensus, effectively negating his rewards altogether.

It follows then that a double-spend attacker wants to minimize perceived and actual disruption in the network, so as not to trigger any of the above punishments. He can start by keeping his reorganization shorter than 100 blocks, where the coinbase rewards of the original chain become spendable. A re-org that deep would no longer just affect individual users, but actually destroy coins and their descendants, potentially invalidating more transactions than intended. A surgical attacker would go as far as replaying every single transaction, including coinbase outputs, to recreate the exact same history with only the double-spend transaction changed.

Given all these constraints, it's highly unlikely that isolated double-spend attacks will become an option for rational miners anytime soon.

3.2 Sabotage attacks

Unlike the isolated double-spend attack, a sabotage attacker doesn't intend to make money inside the Bitcoin system. As a result, he is not concerned with user punishments at all. On the contrary, a sabotage attacker could try to crash the price and make users lose faith in Bitcoin. A sabotage attack could be rational for someone with a bet against the price of Bitcoin, or to defend an existing income stream that is threatened by the existence of Bitcoin. Such an income stream could be seigniorage from the fiat money system or the state's ability to tax in a world where Bitcoin allows users to hide their funds from local governments. This kind of scheme has also fittingly been called "Goldfinger attack," after the James Bond villain who planned on tainting all the money in Fort Knox to make his own gold more valuable[12].

To best erode user trust in the system, the attacker would focus on disabling one-by-one the design goals we established for Bitcoin: safety, liveness, and permissionless access.

One way to achieve this would be to establish a mining monopoly and stop processing any transactions at all. Any majority miner can establish a monopoly if he wants, by simply ignoring blocks mined by the minority. Because he is sure to pull ahead eventually, whatever blocks the minority temporarily appends will be reorganized out later. Instead of processing no transactions, the monopoly miner can also extort users by setting a minimum

[12] https://www.semanticscholar.org/paper/The-Economics-of-Bitcoin-Mining-%2C-or-Bitcoin-in-the-Kroll-Davey/7bf78054192d98e999edcdf08971a5eed42518d2

fee or establish his own rules for what transactions will be processed. For example, he could ignore all transactions that don't pass his personal KYC/AML check. There are three basic ways that users can defend against such censorship attacks.

1) We should establish that the damage from censorship is equal to the exit cost from the system for censored users. The more alternatives to Bitcoin exist, the lower the exit cost will be, and the lower the incentive to censor Bitcoin users in the first place. A similar logic applies for on- and off-ramps like decentralized exchanges. Herein lies an interesting conundrum: whereas a strong KYC/AML layer on top of Bitcoin makes theft less attractive (the coins from the Bitfinex hack are blacklisted everywhere), it also makes the system more vulnerable to censorship. On the other hand, a system without any concept of identity presents more incentive to steal but less incentive to censor.
2) When transactions are being censored, the attacker processes fewer transactions, and censored users start increasing the transaction fees on the unprocessed payments. As a result, a spread starts to build between MR(honest mining) and MR(attack mining). Censored users are effectively free-rolling at this point and can raise the fees over time until they consume almost their entire balance. The delta from these transaction fees can turn into a substantial bounty for an honest majority to challenge the existing mining monopoly and possibly overthrow it.
3) Last, users can coordinate to suspend Nakamoto consensus and make rule changes to punish the monopoly miner. One such last-resort change would be changing the proof-of-work algorithm away from SHA256 to an algorithm that is not yet dominated by the attacker. Instead of extending the heaviest chain with useless blocks, a monopoly miner could also repeatedly reorg the chain, but the effect and ways to deal with it would largely be the same.

4. Declining block subsidy

When extrapolating from our model into the future, we have to consider which of the current parameters are going to change, and why. We established that Bitcoin derives the vast majority of its security from a surprisingly small number of factors: the miner commitment, MEV, and user price sensitivity. The ability to suspend Nakamoto consensus rounds out the picture, but it cannot be the basis of security itself. If users knew a cheaper mechanism to coordinate than Nakamoto consensus, we wouldn't need mining in the first place.

Today, Bitcoin's volatility requires miners to have a higher risk tolerance. If price appreciation ever tops out and Bitcoin finds a stable plateau, mining could start to resemble more traditional commodity markets that offer producers low yields and low volatility. Lower volatility naturally allows miners to use higher leverage, making even a small price change more easily felt.

If Bitcoin more seriously threatens the sovereignty of national currencies and the ability of local governments to collect taxes, that would increase the incentive to attack the network by enforcing a censorship regime or other forms of sabotage. The existence of deep derivatives markets can also make it easier to make large bets against the price of Bitcoin, which further adds to possible MEV.

The largest change, however, is programmed into the Bitcoin protocol itself. All miner revenue, the determining factor for the strength of miner commitment, comes from block rewards, consisting of

1) a **block subsidy in the form of newly minted coins**, and
2) **transaction fees**.

That block subsidy, which currently makes up 99% of the total block reward, is being phased out according to Bitcoin's fixed emission schedule. In 2020, Bitcoin's annual issuance will fall to 1.8%. By 2028 it has halved twice more to 0.5%.

As a result, **the most important source of miner revenue, the block subsidy, will have to be replaced by an entirely new source of revenue**. So far, Bitcoin has derived its security from the value of Bitcoin itself. Going forward, it will derive its security from a secondary market that does not yet exist.

Whether the transition can succeed or not is a substantial source of uncertainty over Bitcoin's future. Today, fees serve the purpose of arbitrating priority for the fixed supply of block space. To create a sufficient miner revenue, the demand for blockspace has to exceed the supply of blockspace at a meaningful price level to create a constant backlog of transactions waiting to be processed.

While it is possible that future demand for block space will be both high and little volatile, there are also scenarios where the market finds Bitcoin useful and transaction fees remain low anyway. This would be the case if most people used Bitcoin simply by holding it, and most transactions happen either on centralized exchanges or various off-chain solutions (there is no reason why large exchanges should settle with each other more than once a day or once a week.)

4.1 The impact of confirmations on security

Folk wisdom suggests that the declining block subsidy does not pose a significant risk because users can make up for it by waiting for more than a few confirmations. Our model shows that the relationship between security and confirmations is a little more complex than that.

First, we consider the impact of additional confirmations on the total miner commitment. Recall that miners have already committed to 50% of the coins they expect to mine over two years. Their total commitment is 658,800 BTC, or 6.25 BTC per block. At every block, miners combine 6.25 BTC of operating costs with 6.25 BTC of commitment costs into the total MC per block of 12.5, which is equal to the block reward.

If users consider a payment final after 6 blocks, then the minimum attack duration for a double-spending attacker becomes 7 blocks. To mine these 7 blocks, the attacker only has to spend an additional 7 * 6.25 BTC = 43.75 BTC.

In a 7-block attack, he now risks 658,800 BTC from his commitment plus 43.75 BTC from his operating costs . In a 70-block (~12 hour) attack , he risks 658,800 BTC plus 437.5 BTC. And with 700 (~5 days) blocks, he risks 658,800 BTC plus 4,375 BTC. Hence we can see that if users were willing to wait for an entire week, the total miner commitment would have increased by less than 1%. The bottom line is that waiting for more confirmations does not add any substantially to miner commitment, and there would be no demand to transact at the point where it does (months+). The same logic applies if block rewards become smaller in the future. Confirmations contribute to the effective miner commitment relative to MR, with every additional confirmation adding exactly 50% of the current block reward. As MR decreases, the value of each confirmation decreases in lockstep.

This logic changes considerably in the low confirmation numbers. While longer reorgs assume that the attacker has a hashpower majority, and hence a large commitment, shallower reorgs don't require the attacker to have this majority. The chance that a miner with 10% hashpower successfully reorganizes two blocks is 17%, and he still succeeds 1% in reorganizing 6 blocks[13]. These probabilistic attacks could be profitable if miner commitment is low and MEV very high. Therefore, waiting for the first couple confirmations is not a defense against a majority attacker but against the probabilistic attacks of a minority miner.

[13] https://arxiv.org/pdf/1702.02867.pdf

While there may be no significant difference between waiting six confirmations and 60 confirmations, after 100 confirmations the security benefit starts adding up again. As discussed in chapter 3.1, this is the threshold where the block reward becomes spendable, making reorgs beyond this depth significantly more disruptive to the network. The greater the disruption for users, the lower is the coordination cost for users to strike back back by selling coins or suspending Nakamoto consensus.

Hence, when receiving money without extra-protocol recourse, a good rule could be to wait >100 confirmations for only the largest of transactions, and 6 confirmations for anything else. In between, we find that confirmations add very little to the miner commitment.

5. Considerations for long-term security

If a robust blockspace market fails to materialize, Bitcoin does not become unusable overnight. Instead, the block subsidy decreases steadily, and over a long period of time. Any problems that can arise as a result of lower MR will appear in weaker form first, before becoming more severe over time, giving users ample time to react and coordinate on possible solutions.

We find it important to note that even if these problems materialize, our outlook on Bitcoin remains optimistic. Bitcoin has the largest user base, the most widely respected supply distribution, and is increasingly integrated into financial infrastructure. Over the course of its short lifetime, Bitcoin has evolved from a technology to a socio-political movement with an ideological following and bitcoin as its currency. It is hard to imagine that Bitcoin could completely die from anything other than a total lack of demand.

With all the talk about Bitcoin's immutability to unwanted change, the examples from chapter 2.4 show that Bitcoin can change if the health of the system is on the line. Suggestions to increase security in the future generally fall into three categories: they can seek to increase MR, decrease MEV, or improve the ability to punish miners.

5.1 Improving blockspace

First, Bitcoin developers can try to increase demand for Bitcoin blockspace. This can be accomplished by protocol-level changes that make Bitcoin blockspace more attractive and useful, and by the development of profitable business processes that consume Bitcoin blockspace as an input.

Demand for Bitcoin blockspace consists of a component of demand to transact bitcoins and a component of demand to store arbitrary data within the chain. Innovations that have increased the power and flexibility of Bitcoin transactions have included the addition of timelocks and the construction of the Bitcoin Lightning Network. Arbitrary data storage can be used to implement non-consensus asset ledgers like USDT or colored coins, or to anchor an attestation to the state of another system as in Factom or Veriblock.

The Bitcoin system is highly optimized for the transfer of bitcoins, but there are limits to the extent to which the storage of arbitrary data can be discouraged. Because this arbitrary data can represent unbounded value outside of the Bitcoin network, a business process that consumes blockspace in this way may have inflexible demand and an extremely high willingness to pay, repurposing Bitcoin transaction structures—inefficiently, if necessary—to accomplish its goals. While this arbitrary data demand may create stable demand for Bitcoin blockspace, consistently elevating fees and increasing MR even when the demand to transfer bitcoins experiences large transients, it also injects potentially limitless MEV and adds incentive to attack the chain. Bitcoin users will have to consider the relative value and risks posed by blockspace usage for this purpose and reckon with the incentives that Bitcoin creates to modulate this aspect of blockspace demand given that there are limits on the attributability and technical disincentives to arbitrary data storage.

5.2 Perpetual issuance

A second mechanism could be to fork in perpetual issuance of new coins. While we expect the topic to be highly contentious with the Bitcoin community, we want to talk about it anyway to clean up some popular misconceptions. If we accept that some level of MR is necessary to make Bitcoin work, then MR must be paid for by users one way or the other. If the necessary MR was 1% annually, all Bitcoin users together would already lose 1% of their purchasing power every year to power the Bitcoin system. The bottom line is that while Bitcoin can be a nominal fixed-supply asset, it can’t be a fixed-purchasing power asset.

Further, it is false to talk about perpetual issuance as inflationary. If Bitcoin requires users to lose 1% of their purchasing power anyway, then paying these costs via perpetual issuance presents no more of a loss of purchasing power than paying for them via transaction fees. In fact, the purchasing power of BTC could be higher in a Bitcoin system with 1% perpetual issuance per year and higher security than in a Bitcoin system with 0% perpetual issuance and lower security.

Instead, we should ask who should pay for MR, and using what mechanism? In an idealized system, users would pay for the operating costs based on how much value they receive from it. That would maximize revenue and hence security, because all users pay according to their utility. It further ensures fairness and longevity in the system. A system that is perceived as unfair by some of its members is unlikely to sustain itself for very long - too great is the incentive for the “suckers” to fork into a system of their own and leave the free riders behind.

In practice, the designers of a system may not know in advance who the highest-value users will be. Once established, all users might agree that changing the original parameters to something more optimal would be more costly than simply living with them.

Conceptually there are two main users in the Bitcoin system: holders and transactors. There is no clean cut between them, as any transactor must have owned bitcoin, at least for a short period of time, and any holder must plan on transacting his bitcoin eventually (though not necessarily on-chain).

An adversarial should be resilient to external shocks in soft parameters like the demand to hold bitcoin or demand to use block space. In a perpetual issuance paradigm, MR would be unaffected by events in the block space market, whereas a demand shock for block space in the current paradigm would send the security of the entire system plummeting.

Intuition here says that ownership of the good we want to monetize matters. If we want to monetize block space from transactors, we must ensure that most units of block space are owned by someone all the time. Charging holders eliminates this friction entirely, as every bitcoin has an owner all the time.

Last, it should be noted that the contributions of holders can be less visible than those of transactors, but are nonetheless real. When the system is under attack, holders have more skin in the game and are more likely to pay the costs of social coordination. It is important to have a holistic view of the Bitcoin system under all conditions when evaluating how much either use case contributes to security.

While perpetual issuance of coins would reduce uncertainty over miner revenue, others have argued that a zero-issuance policy is an eternal Schelling point in cryptocurrency.[14] If users really hate the kind of implicit taxation that comes with perpetual issuance, the gamble on the less secure zero-issuance architecture could pay off by creating higher perpetual demand than a comparable low-issuance asset.

[14] http://www.truthcoin.info/blog/deflation-the-last-word/

5.3 Crowdfunding

A less controversial way for Bitcoin holders to chip into MR under the blockspace market paradigm would be the use of crowdfunding. Large holders and businesses with a strong interest to preserve Bitcoin's security could pay into a fund that creates "anyone-can-spend-transactions" (maybe in the form of a Bitcoin-DAO.) These transactions could be claimed by miners at a certain block height and hence serve as a privately-financed block subsidy. The benefit of this solution is that no changes to the protocol are required. The downside is that you end up in a classical free-rider scenario: many people want Bitcoin to be secure, but nobody wants to be the sucker who pays the whole bill for everyone else.

A solution to the free-rider problem could come in the form of dominant assurance contracts (DACs), a variant of the crowdfunding contract that attempts to make contributing the dominant strategy over waiting for others to contribute[15]. In the DAC, one party must take the role of the entrepreneur who wants to have a certain public good (in this case, MR) funded. He defines a target sum to be raised and encourages other people to contribute by paying them a small sum in case the fundraiser misses its target. This small detail is said to make contributing more attractive because contributors now win in both cases - they either get the good or their money back with a profit.

5.4 Adapting the supply of blockspace

Finally, a solution to raising MR can be found in altering the supply of block space. The biggest drawback with the fixed supply system is that whenever demand is even marginally lower than supply, fees immediately go to zero. All users in a block could be willing to pay 5 BTC of transaction fees collectively, but if there is excess supply, they all end up paying nothing because there is no congestion.

Even when the total demand exceeds the available supply, it is not guaranteed to maximize revenue. Assume, for example, that there is 1 MB demand willing to pay 15 BTC, with another 1 MB willing to pay 5 BTC. If the available supply is anything between 1 MB and 2 MB, the total sum of fees would be slightly above 10 BTC because the group that wants to pay the least sets the price for everyone else (group one pays 5.01 whereas group two pays 5.00). If the supply was lowered to below 1 MB, then group one would have to pay 15 BTC, leading to much higher MR even though group two is no longer served at all.

[15] A concept from mechanism design that was first applied to Bitcoin by Mike Hearn https://en.bitcoin.it/wiki/Dominant_Assurance_Contracts

This value could be captured by lowering the block size slightly below demand to create permanent congestion. Such changes could be made manually by developers or automatically by the Bitcoin protocol itself. **One such idea is the adaptive block size**: the system looks at the MR produced from fees and compares it to the targetMR required to make the system secure. If MR < targetMR, it lowers the maximum block size to create artificial congestion. If MR > targetMR, users overpay for security, and some of the artificial congestion can be removed, thus increasing the block size, up to a community-chosen hard limit (currently 2.3 MB).

Other proposals, where miners are given control of the block size, are not robust because miners are incentivized to game the system and make blocks as large as possible. The reason is that the largest, best-connected, miners gain a competitive advantage against smaller, or more poorly-connected, miners as the propagation time of blocks increases. We don’t need to be concerned about that here, since the low block size cap ensures that propagation times always stay short.

5.5 Decreasing miner-extractable value

In addition to increasing MR, Bitcoin users can also consider various ideas to decrease MEV. A good starting point is to consider potential sources of MEV on the Bitcoin blockchain.

As discussed in chapters 2 and 3, the incentive to censor the system would decrease as the exit cost from a system decreases. When a miner can’t distinguish between different transactions, he can’t censor any individual users. Hence, a flourishing space of competing cryptocurrencies with private transactions and permissionless exchange between them would make any one of them individually more robust against censorship.

If users lower the barrier to ignore Nakamoto consensus by embracing strategies such as the USAF, they could lower the MEV from some attacks at the risk of making the system less socially scalable. As the Bitcoin system harbors more and more people with diverging or even opposing political views, it seems likely that coming to social consensus without proof-of-work will only become more difficult.

Maybe technical solutions can be discovered in the meantime to restrict the options available to miners further and make attacks less attractive that way. One such suggestion is to have Bitcoin transactions commit to a certain block outside of which they become invalidated. That would make it impossible for miners to replay transactions in a reorg, which has two significant benefits.

1) It makes the attack more costly because the miner loses access to previous transactions and their transaction fees.
2) It makes coordination around suspending Nakamoto consensus easier, because a miner can no longer attack individual users in isolation. He now has to choose between reorganizing many transactions at once, or none at all.

Further, we could improve automatic detection of malicious miner behavior. Dealing with attacks first requires all users to learn about them. The better we can monitor the state of the Bitcoin system, the less can a miner hope to get away with not following the protocol, including non-consensus attacks like selfish mining.

More education around Bitcoin's trust model can help lower the potential for theft as well. Not every transaction a user receives comes from a miner, or someone bribing a miner, and is at risk of being double-spent. Using the traditional legal system in addition to Bitcoin, where possible, can greatly amplify its viability in commerce. Whenever the buyer has a legal relationship with the seller, the seller can treat that as an external commitment via the traditional legal system, and thus gain additional confidence that the payment will not be reversed.

5.6 Improving miner punishment

Low tolerance of Bitcoin users to malicious behavior from miners is a powerful check on their behavior. When the price moves more strongly in response to attacks, Bitcoin can afford the same level of MEV but with less committed miners. If the price is very robust, the commitment must be larger.

The sensitivity of the Bitcoin price to the utility of the system is once again a function of the exit cost from the system. Walking away is much easier when leaving is cheap, possibly because Bitcoin is not the only game in town, and there exists competition between many cryptocurrencies with similar assurances. In fact, the concept of cryptocurrencies is at its most robust when there are many "micro-chains" that are more fragile but allow for fluid exchange between them. The reason is that smaller blockchains are easier for users to walk away from, resulting in a scorched-earth defense against the attacker.[16]

[16] A concept pioneered by David Vorick.

6. Omissions and future research

Our model of Bitcoin's security could be expanded in several ways. First, one could look at the ability for miners to "un-commit" from the system. So far, we consider this implicitly - if a miner had a large bet against the price of BTC, we can increase his MEV to reflect that. With unlimited capital, a miner could fully hedge his commitment while maintaining the same level of hashpower - and hence potential MEV. A follow-up analysis could focus on the capital costs of hedging, its impact on costs and MEV, and how the existence of deep derivatives markets changes the incentives of all actors.

Second, previous security analyses may have dramatically underestimated the incentive for a possible hashpower minority to fight back during, or immediately after, an attack to defend the value of their commitment. Because the defenders are effectively free-rolling, hashing at much higher unit costs becomes profitable again, and old hardware could rejoin the network en masse. Further, existing hardware can be overclocked to increase their effectiveness in the short-term at the cost of faster depreciation. In general, users and minority miners should start seeing each other as allies against outside attackers. Due to the miner commitment, they both sit in the same boat. The dynamics between attacks and counter-attacks deserve more exploration, as they could greatly increase the cost of an attack.

Finally, even if a robust blockspace market does develop, Bitcoin's security model will change in several ways. These changes affect the optimal behavior of both miners and users. For example, block withholding tactics are attractive if individual blocks have very low fees attached to them. Increased competition between miners will further develop around particularly "wealthy" blocks, leading to fee sniping and gaps in block production[17]. We strongly encourage a systemization of all the ways that a fee-based system is different from an issuance-based system.

[17] http://randomwalker.info/publications/mining_CCS.pdf and https://arxiv.org/abs/1805.05288