The Economic Limits of Bitcoin and the Blockchain

Eric Budish
University of Chicago Booth School of Business

June 2018
New Working Paper Series No. 24
The Economic Limits of Bitcoin and the Blockchain\textsuperscript{*†}

Eric Budish\textsuperscript{‡}

June 5, 2018

Abstract

The amount of computational power devoted to anonymous, decentralized blockchains such as Bitcoin’s must simultaneously satisfy two conditions in equilibrium: (1) a zero-profit condition among miners, who engage in a rent-seeking competition for the prize associated with adding the next block to the chain; and (2) an incentive compatibility condition on the system’s vulnerability to a “majority attack”, namely that the computational costs of such an attack must exceed the benefits. Together, these two equations imply that (3) the recurring, “flow”, payments to miners for running the blockchain must be large relative to the one-off, “stock”, benefits of attacking it. This is very expensive! The constraint is softer (i.e., stock versus stock) if both (i) the mining technology used to run the blockchain is both scarce and non-repurposable, and (ii) any majority attack is a “sabotage” in that it causes a collapse in the economic value of the blockchain; however, reliance on non-repurposable technology for security and vulnerability to sabotage each raise their own concerns, and point to specific collapse scenarios. In particular, the model suggests that Bitcoin would be majority attacked if it became sufficiently economically important — e.g., if it became a “store of value” akin to gold — which suggests that there are intrinsic economic limits to how economically important it can become in the first place.

\textsuperscript{*}Project start date: Feb 18, 2018. First public draft: May 3, 2018. For the record, the first large-stakes majority attack of a well-known cryptocurrency, the $18M attack on Bitcoin Gold, occurred a few weeks later in mid-May 2018 (Wilmoth, 2018; Wong, 2018).

\textsuperscript{†}Acknowledgments: thanks are due to Susan Athey, Vitalik Buterin, Alex Frankel, Joshua Gans, Austan Goolsbee, Zhiguo He, Joi Ito, Steve Kaplan, Anil Kashyap, Judd Kessler, Randall Kroszner, Robin Lee, Jacob Leshno, Neale Mahoney, Sendhil Mullainathan, David Parkes, John Shim, Scott Stornetta, Aviv Zohar, and seminar participants at Chicago Booth and the MIT Digital Currency Initiative. Natalia Drozdoff and Matthew O’Keefe have provided excellent research assistance. Disclosure: I do not have any financial interests in blockchain companies or cryptocurrencies, either long or short.

\textsuperscript{‡}University of Chicago Booth School of Business, eric.budish@chicagobooth.edu
1 Bitcoin and the Blockchain: A Critique in 3 Equations

1.1 Rent-Seeking Competition (Among Blockchain Miners)

Bitcoin is an electronic payment system that relies on a combination of cryptography and a large, anonymous, decentralized collection of participants, called miners, to verify transactions, without the need of any trusted third party. The basic details, slightly simplified, are as follows. An owner of Bitcoin may send currency to another user by using a combination of (i) his own public address (an alphanumeric string, somewhat analogous to an account number), (ii) his own private key (i.e., password associated with that address), and (iii) the recipient’s public address to create a transaction with a cryptographically secure signature. This signature has the property that it can only be created by someone who knows the sender’s private key (presumably the sender!), and encodes the amount of Bitcoin to be transferred from the sender to the receiver, but in a way that an observer of the signature cannot invert the information to learn the sender’s private key. So far, while magical to those unfamiliar with modern cryptography, this is completely standard. The innovative idea behind Bitcoin is the way these transactions are publicly recorded, to maintain a public ledger of all transactions, called the “blockchain”. Every so often (currently roughly ten minutes), a large, anonymous, decentralized collection of participants called miners competes in a computational tournament for the right to add a new block of transactions (roughly ten minutes’ worth) to the public ledger. The winner of the computational tournament is the first participant to solve a difficult computational problem based on both the new block and previous block of transactions. This participant reports both the new block of transactions and the solution to the computational problem, and the other participants, to quote from Nakamoto (2008), “express their acceptance of the [new] block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash.”

The incentive issues raised by this system will be the subject of the next section. For the purpose of this section, assume that participants in the Bitcoin system behave as they are supposed to, i.e., “honestly”. Let $P_{\text{block}}$ denote the economic reward to the miner who wins the computational tournament. For the moment think of $P_{\text{block}}$ as exogenous; equation (3) below will place

---

1For an accessible overview of the Bitcoin blockchain system, aimed at economists, highly recommended is the institutional background section of Huberman, Leshno and Moallemi (2017). For empirical facts on Bitcoin usage, good starting points are Athey et al. (2016) and Böhme et al. (2015). For a game-theoretic analysis of Bitcoin mining, a good starting point is Biais et al. (2017). Other helpful references include the original Bitcoin paper, Nakamoto (2008), the website Bitcoin.org (especially its Bitcoin Developer Guide), and, for a textbook length treatment, Narayanan et al. (2016). Huberman, Leshno and Moallemi (2017), Biais et al. (2017) and Athey et al. (2016) contain helpful overviews of the extant academic literature.

2The economic reward consists of two components: a fee associated with each transaction (this is strategically rich in its own right, see Huberman, Leshno and Moallemi, 2017 and Easley, O’Hara and Basu, 2017), and a reward of “freshly minted” Bitcoins, currently 12.5 per block. As of early 2018, with Bitcoin prices of roughly $10,000, the
constraints on what values of $P_{\text{block}}$ are possible in equilibrium. Let $c$ denote the per-block cost of one unit of computational power, including both variable costs such as electricity and a rental cost for capital equipment. More fully, assume that it takes one chip and one unit of electricity to produce one unit of computational power, a chip costs $C$, the per-block cost of capital (including depreciation) is $r$, and the per-block cost of one unit of electricity is $e$; then we have $c = rC + e$. Assume for now that this cost is symmetric across all participants and that the chips are easily repurposeable, so we do not have to worry about sunk costs, adjustment costs, etc. We will revisit this assumption — which I emphasize is not satisfied for Bitcoin at present, due to specialized ASIC chips that are thousands of times more efficient at Bitcoin mining than repurposeable chips, though it does capture the original Nakamoto (2008) vision of “one-CPU-one-vote” — in detail in Section 3. If there are $N$ units of computational power in the network, then each unit has a $\frac{1}{N}$ probability of winning the prize $P_{\text{block}}$. Under standard free entry logic — any entity that likes can add computational power to the network — the equilibrium amount of computational power devoted to blockchain mining, $N^*$, is thus characterized by:

$$N^* c = P_{\text{block}} \tag{1}$$

Equation (1) is the standard characterization of a rent-seeking tournament: the prize in the tournament, $P_{\text{block}}$, is dissipated by expenditures aimed at winning the prize, $N^* c$. That Bitcoin mining can be modeled as a rent-seeking contest is now widely known; see for instance Kroll, Davey and Felten (2013) pg. 8; Huberman, Leshno and Moallemi (2017) Proposition 1; Easley, O’Hara and Basu (2017) equation (1); Chiu and Koepppl (2017) Lemma 1; and Ma, Gans and Tourky (2018) equation (7). An analogous example of a rent-seeking tournament is the high-frequency trading arms race; equation (7) of Budish, Cramton and Shim (2015) is very similar to equation (1) here.

The Bitcoin Wiki acknowledges the rent-seeking competition among miners in detail as well, under the heading “Weaknesses $\rightarrow$ Energy Consumption”:

“... the economic equilibrium for the mining rate is reached when global electricity costs for mining approximate the value of mining reward plus transaction fees. So the higher the value of one bitcoin, the higher the value of mining rewards and transaction fees, the higher the energy consumption of the bitcoin network in the long run. More efficient mining gear does not reduce energy use of the bitcoin network.

latter component of the reward is much larger than the former. In steady state, as the system slowly runs out of new Bitcoins to issue, the fees will need to be the larger component (cf. Huberman, Leshno and Moallemi, 2017 and Easley, O’Hara and Basu, 2017). The number of transactions per block has recently ranged between 1000-2000, so the reward per transaction is roughly $100.
... cheaper energy linearly increases mining energy use ... the same conclusions apply to all proof-of-work based currencies.” (Bitcoin Wiki, 2018c, section “Weaknesses -> Energy Consumption”)

### 1.2 Incentive Compatibility (with Respect to Majority Attack)

From the abstract of Nakamoto (2008):

“The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain serves not only as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. *As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they’ll generate the longest chain and outpace attackers.*” (Emphasis added)

Section 11 of Nakamoto (2008) then analyzes the “scenario of an attacker trying to generate an alternate chain faster than the honest chain” under the assumption that the attacker controls less than a majority of the computational power. The analysis shows that such an attack is analogous to a Gambler’s Ruin problem and that the likelihood of the attacker pulling ahead of the honest participants given a deficit of \( z \) blocks, or the likelihood of a successful double-spending attack (cf. Section 2.1) given an escrow period of \( z \) blocks, is exponentially declining in \( z \). The conclusion is that it “quickly becomes computationally impractical for an attacker to change [the public history of transactions] if honest nodes control a majority of CPU power.”

What about an attacker with a majority of computational power? It is widely acknowledged, including in Nakamoto (2008) and on the Bitcoin Wiki (sections: “Attacker has a lot of computing power” (2018b) and “Majority attack” (2018d)) that such an attack would succeed.\(^3\) From the Wiki’s section on “Majority Attack”: “Bitcoin’s security model relies on no single coalition of miners controlling more than half the mining power.”

What does it cost to gain a majority of computational power? If there are \( N^* \) units of honest computational power devoted to each mining tournament, then the cost to an outside attacker of a simple majority is \( N^*c + \epsilon \) per block. If the attacker already controls some of the honest computational power, the cost of a majority is correspondingly lower; it could be as small as \( \frac{N^*c}{2} + \epsilon \). Super-majorities cost correspondingly more; an expenditure of \( A \cdot N^*c \) per block, for...

\(^3\) For academic analyses of the majority attack see Rosenfeld (2012) and Eyal and Sirer (2014). Eyal and Sirer (2014) also show that the Bitcoin system is vulnerable to a form of minority attack, though the purpose of the Eyal and Sirer (2014) minority attack is more circumscribed in that its purpose is to obtain a disproportionate share of mining rewards, rather than to manipulate the blockchain per se.
\( A > 1 \), yields an \( \frac{A}{A+1} \) super-majority for an outside attacker.\(^4\)

We will discuss two specific possibilities for such a majority attack in detail in Section 2. For the purpose of this section, suppose that there exists a majority attack that yields an expected payoff to the attacker of \( V_{\text{attack}} \), and that has an expected cost to the attacker, net of block rewards, of \( \alpha \cdot N^* c \). By expected cost net of block rewards, we mean, more precisely, that if an attack by an \( \frac{A}{A+1} \) super-majority attacker takes \( t \) blocks worth of time in expectation (measured based on how long it takes the honest chain to solve blocks), and thus yields \( t \) block rewards in expectation, then the total cost net of block rewards is \( At \cdot N^* c - tP_{\text{block}} \) which, using \( P_{\text{block}} = N^* c \) from (1), yields \( \alpha = (A - 1)t \). For the blockchain system to be incentive compatible against such an attack requires:

\[
\alpha \cdot N^* c > V_{\text{attack}} \tag{2}
\]

Equation (2) simply says that the costs of manipulating the blockchain, \( \alpha \cdot N^* c \), must be greater than the benefits of doing so, \( V_{\text{attack}} \). The equation captures that what enables the “decentralized trust” of the blockchain system is the computing power devoted to maintaining it.

Economically, the key thing to note about (2) is that the cost of manipulation on the LHS is related to the flow cost of maintaining the blockchain, i.e., to \( N^* c \). In contrast, consider, e.g., mutually-beneficial cooperation in a relationship and the associated temptation to cheat, or a trusted brand that is tempted to shirk on quality. In such cases, the cost of cheating, to the cheating party, is related to the stock value of the relationship or brand they are destroying, not the flow cost of its maintenance.\(^5\)

From a computer security perspective, the key thing to note about (2) is that the security of the blockchain is linear in the amount of expenditure on mining power, i.e., linear in \( N^* c \) in the LHS of (2). In contrast, in many other contexts investments in computer security yield convex returns (e.g., traditional uses of cryptography) — analogously to how a lock on a door increases the security of a house by more than the cost of the lock.

\(^{4}\)If the attack both (i) uses technology that cannot be easily repurposed, and (ii) is a “sabotage” in the sense that it causes a meaningful decline in the value of Bitcoin / the blockchain, then it is economically appropriate to charge the attacker not just the rental cost of capital embedded in \( c \) (alongside pure variable costs such as electricity), but also some or all of the fixed costs of the equipment he uses. See Section 3 for a detailed analysis.

\(^{5}\)Perhaps the earliest version of this insight is due to Schelling (1956): “What makes many agreements enforceable is only the recognition of future opportunities for agreement that will be eliminated if mutual trust is not created and maintained, and whose value outweighs the momentary gain from cheating in the present instance.” For more references and discussion good starting points are Nobel Prize Committee (2005) and Mailath and Samuelson (2006).
1.3 Economic Limit of the Blockchain I: $P > \frac{V}{\alpha}$

In the hoped-for equilibrium in which participants are honest, the amount of computational power devoted to maintaining the blockchain is characterized by the rent-seeking competition among miners, (1). Combining (1) with the incentive compatibility condition, (2), we have the equilibrium constraint:

$$P_{\text{block}} > \frac{V_{\text{attack}}}{\alpha}$$

In words: the equilibrium per-block payment to miners for running the blockchain must be large relative to the one-off benefits of attacking it. Equation (3) places potentially serious economic constraints on the applicability of the Nakamoto (2008) blockchain innovation. By analogy, imagine if users of the Visa network had to pay fees to Visa, every ten minutes, that were large relative to the value of a successful one-off attack on the Visa network.

2 Two Possibilities for $V_{\text{attack}}$

Before describing two leading possibilities for majority attacks, let us clarify what, technologically, a majority attacker can and cannot do. Because a majority attacker can solve computational puzzles faster, in expectation, than the honest minority, the attacker can create an alternative longest chain of transactions, and replace the honest chain with the alternative chain at a strategically opportune moment. This allows the attacker to control what transactions get added to the public blockchain, and allows the attacker, within computational limits, to remove recent transactions from the public blockchain (by creating an alternative chain starting from the recent past, and catching up). The attacker even earns the blockchain rewards, i.e., the $P_{\text{block}}$ per period, for each period of his alternative chain after he makes it the new longest public chain.

What the attacker cannot do is to create new transactions that spend other participants’ Bitcoins. Creating new transactions that spend other participants’ coins would require not just a majority of computational power, but enough computational power to break modern cryptography:

---

6 As noted above, equation (1) is widely known, as is Bitcoin’s vulnerability to a majority attack, from which equation (2) is essentially tautological; see, for example, the excellent discussion of majority attacks in Bonneau (2018). To the extent this paper makes an intellectual contribution, it is to put (1) and (2) into a common theoretical language, to note that they imply (3), and then to analyze the economic implications of (3). To my knowledge, equation (3) is new to this paper.

7 Blockchain rewards do not vest for 100 periods (see Bitcoin Protocol Rules, section “tx” messages”, item 11; and Bitcoin Developer Guide, section “Transaction Data” (Bitcoin Wiki, 2017; Bitcoin.org, 2018)), so as long as the attacker replaces the honest chain within 100 periods, he earns the blockchain reward for every block he adds. If he replaces the honest chain after 100 periods it is somewhat ambiguous, at least to my current reading of the Bitcoin Protocol, what happens to the block rewards that have already vested on the part of the honest chain that has now been replaced. My best guess is that those Bitcoins would become unusable and that the attacker would get new Bitcoins for all periods of the chain he replaced.
creating a transaction that spends another participant’s coins requires learning their private key. A majority attacker cannot simply “steal all the Bitcoins.”

2.1 “Double Spending” Attack

The most widely discussed manipulation a majority attacker can engage in is known in the literature as “double spending”. An attacker could (i) spend Bitcoins, i.e., engage in a transaction in which he sends his Bitcoins to some merchant in exchange for goods or assets; then (ii) allow that transaction to be added to the public blockchain (i.e., the longest chain); and then subsequently (iii) remove that transaction from the public blockchain, by building an alternative longest chain, which he can do with certainty given his majority of computing power. The merchant, upon seeing the transaction added to the public blockchain in (ii), gives the attacker goods or assets in exchange for the Bitcoins, perhaps after an escrow period. But, when the attacker removes the transaction from the public blockchain in (iii), the merchant effectively loses his Bitcoins, allowing the attacker to “double spend” the coins elsewhere.

As should be clear, while this problem is called the “double spending” problem, the “double” part is a misnomer — the attacker can re-spend his Bitcoins arbitrarily many times.

To translate the “double spending” problem into values for $V_{\text{attack}}$ and $\alpha$, let us make the following assumptions:

1. there are $k$ transactions in a block;
2. the attacker engages in $k$ distinct transactions (relaxing this in either direction is discussed below); each transaction can be interpreted as coming from a separate address he controls; for simplicity all $k$ of the transactions get added to the public blockchain in what we will call block 1, with block 0 referring to the state of the blockchain prior to the attack;

---

8Here is a detailed excerpt on what majority attackers can and cannot do from the Bitcoin Wiki, under “Attacker has a lot of computing power”:

“An attacker that controls more than 50% of the network’s computing power can, for the time that he is in control, exclude and modify the ordering of transactions. This allows him to:

- Reverse transactions that he sends while he’s in control. This has the potential to double-spend transactions that previously had already been seen in the block chain.
- Prevent some or all transactions from gaining any confirmations.
- Prevent some or all other miners from mining any valid blocks.

The attacker can’t:

- Reverse other people’s transactions without their cooperation.
- Prevent transactions from being sent at all (they’ll show as 0/unconfirmed).
- Change the number of coins generated per block.
- Create coins out of thin air.
- Send coins that never belonged to him.” (Bitcoin Wiki, 2018b)
3. the average value of the transactions the attacker engages in is $\bar{v}_{\text{transaction}}$; this can be interpreted as a statistic on the highest-value transactions that are possible on the Bitcoin system;

4. merchants wait for an escrow period of $e$ blocks, counting from the block in which the transaction is first added to the public blockchain, before sending the goods or assets;

5. the honest miners have computational power of $N^*$ and the attacker has power of $AN^*$, with $A > 1$;

6. the attacker obtains block rewards of $P_{\text{block}}$ per block of the attack;

7. the attack does not affect the subsequent value of Bitcoins (this will be discussed in Sections 2.2 and 3).

Given these assumptions, the value of the attack is $V_{\text{attack}} = k\bar{v}_{\text{transaction}}$. Define $p_{\text{transaction}} = \frac{P_{\text{block}}}{k}$. Then (3) can be rewritten as:

$$p_{\text{transaction}} > \frac{\bar{v}_{\text{transaction}}}{\alpha}$$

where $p_{\text{transaction}}$ is the per-transaction payment to miners for running the blockchain, $\bar{v}_{\text{transaction}}$ represents the size of the transactions that are possible using the blockchain, and $\alpha$ represents the net cost of the attack.

To compute $\alpha$ we run computational simulations, reported in Table 1, for different values of computational power $A$ and escrow period $e$. Panel A reports the expected duration of the attack (in blocks), Panel B reports the expected gross computational cost of the attack (in units of $N^*c = P_{\text{block}}$), and Panel C reports the expected cost net of block rewards (in units of $N^*c = P_{\text{block}}$), i.e., $\alpha$.

For intuition, first focus on the $e = 0$ column, i.e., no escrow period. In this case, the best-case scenario for an attacker is that they solve two computational puzzles before the honest miners solve two computational puzzles; in such case, after some honest miner reports block 1 which includes the attacker’s transactions, the attacker can report alternative blocks 1’ and 2’ such that the chain $\ldots, 0, 1’, 2’$ is the new longest chain and excludes his transactions. As the attacker’s computational power grows, this best-case scenario becomes increasingly likely, to the point where if $A = 5$, the expected duration of the attack is just 2.12 blocks. With lower computational power, the attack takes longer in expectation, but is usually cheaper, both gross and net. For example, if $A = 1.25$ the attack has expected duration of 6.54 blocks, gross costs of 8.18$P_{\text{block}}$, and net costs of 1.64$P_{\text{block}}$.

Now focus on the $e = 6$ column, a commonly-discussed escrow period, corresponding to about one hour (for example, see the Bitcoin Wiki sections on “Alternative History Attack” and “Attacker has a Lot of Computing Power” (Bitcoin Wiki, 2018a,b)). If, say $A = 1.25$, the expected attack
duration is 13.41 blocks with an expected cost net of block rewards of $α = 3.35$.

Plugging this value of $α$ into (4) yields $p_{\text{transaction}} > \frac{\bar{v}_{\text{transaction}}}{3.35}$. This means that for the blockchain system to be robust against this double-spending attack would require that the per-transaction payment to miners for running the blockchain exceeds $\frac{1}{3.35} \approx 30\%$ of the highest-value transactions that are possible through the system. This can be interpreted as an implicit tax — if $\bar{v}_{\text{transaction}} = $100, then the implicit transactions cost on a $100 purchase must be at least $30$. The tax is larger as a percentage on cheaper transactions, with cheap defined relative to the largest feasible transactions. For example, if the most expensive goods that could be obtained in the system are worth $100, then the implicit tax on a $10 transaction would need to be in excess of 300\%.

Add in some difficulty of coordinating $k$ purchases at once, some inefficiency of the attacker’s technology relative to the marginal honest miners’ technology, or any other form of “slippage” for the attacker of the benefits relative to $k\bar{v}_{\text{transaction}}$ or the costs relative to $\alpha \cdot N^*c$, and it is easy to imagine that $p_{\text{transaction}}$ on the order of $1$-$10$ is enough for the system to be robust to majority attack for purchases of on the order of $100$-$1000$. This may be a reasonable description of Bitcoin in many of its leading early use cases,\(^9\) such as modest-sized purchases of illegal goods and computer equipment, modest-sized international remittances (e.g., Bitcoin as an alternative to Western Union), etc.

If $\bar{v}_{\text{transaction}}$ is much higher, however — and remember, this is a statistic on the largest transaction sizes that are possible — then $p_{\text{transaction}}$ needs to be commensurately higher as well, for the system not to incentivize a majority attack. For example, suppose Bitcoin is used as a “store of value” akin to gold, as has been discussed by Cochrane (2017), Cowen (2017\textit{a}, \textit{b}) and many others. Suppose it is possible to convert $1,000,000$ of Bitcoin into other forms of wealth in the largest transactions in the system. Then, even with an escrow period of $e = 100$ blocks, and an attack period of just one block’s worth of transactions, then, focusing now on $A = 1.05$, you would need $p_{\text{transaction}} > \frac{1}{9.2^2} \approx 1.87$ for the system not to induce a majority attack. Even an escrow period of $e = 1000$ blocks — 1 week at 10 minutes per block — and an attack period of just one block’s worth of transactions would require $p_{\text{transaction}} \approx 18.7$. If the attacker instead attacks say 2 blocks’ worth of transactions (perhaps spread out over time), then even with $e = 1000$ the system requires $p_{\text{transaction}} \approx 37.4$ not to be vulnerable to this majority attack.

Essentially, for the system to be usable for large transactions requires implicit tax rates that

\(^9\)See, for instance, Böhme et al. (2015). The authors of Soska and Christin (2015) report in a private communication that in their data, based on 16 different online anonymous marketplaces over the period 2013-2015, that the average transaction they observe is around $100$-$150 depending on how the data is cut. They report that there are zero $1M+$ transactions and that $10,000+$ transactions are extremely rare. See also the section “Bulk Items” in Christin (2017) for details on the small number of $10,000+$ transactions they do observe.
likely render it unusable for smaller transactions.

2.2 “Sabotage” Attack

One seemingly obvious response to the logic in the previous section is that the majority attack would be “noticed” by Bitcoin users, perhaps after a period of initial confusion. As a result, the argument goes, while the attack would indeed work in the sense of obtaining the hoped-for goods or assets, there is an additional cost to consider in that the attack will harm the subsequent value of the attacker’s own Bitcoin holdings — which the attacker must have to engage in the attack in the first place. The Bitcoin Wiki classifies the majority attack into its “Probably Not a Problem” category for this reason, making the following argument:

“A miner with more than 50% hash power is incentived to reduce their mining power and refrain from attacking in order for their mining equipment and bitcoin income to retain its value.” (Bitcoin Wiki, 2018d, section “Majority Attack”)

Formally, let us assume that the double-spending attack analyzed in Section 2.1 causes a proportional decline in the value of Bitcoin of \( \Delta_{\text{attack}} \), and that the attacker holds the minimum amount of Bitcoin necessary to conduct the attack, namely \( k\bar{v}_{\text{transaction}} \) worth. For this section, we maintain the assumption that the technology used for the attack can be repurposed; we will consider blockchain-specific mining equipment in detail (as mentioned in the Wiki quote above) in the next section. The \( \Delta_{\text{attack}} \) decline in the value of Bitcoin modifies equation (4) to be:

\[
10p_{\text{transaction}} > \frac{(1 - \Delta_{\text{attack}})}{(A - 1 + \Delta_{\text{attack}})t}\bar{v}_{\text{transaction}} \tag{4'}
\]

The larger is \( \Delta_{\text{attack}} \), the smaller is the implicit tax on the system necessary to deter the majority attack, i.e., the level of \( p_{\text{transaction}} \) necessary to support a given level of \( \bar{v}_{\text{transaction}} \). For example, if \( \Delta_{\text{attack}} = 1 \), i.e., if the attack causes a total collapse of the value of Bitcoin, the attacker loses exactly as much in Bitcoin value as he gains from double spending; in effect, there is no chance to “double” spend after all. In this respect, the argument on the Bitcoin Wiki is correct. However, \( \Delta_{\text{attack}} \) is something of a “pick your poison” parameter. If \( \Delta_{\text{attack}} \) is small, then the system is vulnerable to the double-spending attack described in Section 2.1, and the implicit transactions tax on economic activity using the blockchain has to be high. If \( \Delta_{\text{attack}} \) is large, then a short time period of access to a large amount of computing power can sabotage the blockchain.

\[\text{The attacker gets a benefit of } k\bar{v}_{\text{transaction}}, \text{ but to realize this benefit has to hold Bitcoins worth this amount, so the net benefit of the attack is } (1 - \Delta_{\text{attack}})k\bar{v}_{\text{transaction}}. \text{ The attacker pays gross computational costs of } At\cdot N^c, \text{ where } t \text{ is the expected duration of the attack. He gets block rewards of } t\cdot N^c \text{ which then decline in value by proportion } \Delta_{\text{attack}}, \text{ so the net cost is } (A - 1 + \Delta_{\text{attack}})t\cdot N^c. \text{ Substituting } kp_{\text{transaction}} = P_{\text{block}} = N^c \text{ and rearranging yields } (4').\]
If $\Delta_{\text{attack}}$ is large, one must then consider the possibility of an attacker who is motivated by sabotage per se, rather than double spending; call this value $V_{\text{sabotage}}$. A well-known early paper on the double spending problem, Rosenfeld (2012), notes exactly this possibility:

“In this section we will assume $q < p$ [i.e., that the attacker does not have a majority]. Otherwise, all bets are off with the current Bitcoin protocol ... The honest miners, who no longer receive any rewards, would quit due to lack of incentive; this will make it even easier for the attacker to maintain his dominance. This will cause either the collapse of Bitcoin or a move to a modified protocol. As such, this attack is best seen as an attempt to destroy Bitcoin, motivated not by the desire to obtain Bitcoin value, but rather wishing to maintain entrenched economical systems or obtain speculative profits from holding a short position.” (Emphasis Added)

What is the value of $V_{\text{sabotage}}$? It is hard to say of course, but easy to imagine that the magnitudes are already large, and would be larger still if Bitcoin and/or the blockchain live up to their hype. Open interest on CME and CBOE Bitcoin futures — which gives a sense of magnitudes for what could be made at present from a short-selling attack — is about $160 million as of March 2018.\(^{11}\) In comparison, open interest on CME Gold Futures — which may give a more appropriate sense of magnitudes for the hypothetical scenario in which Bitcoin is used as a “store of value” — is about $65 billion.\(^{12}\) The market capitalization of Bitcoin — which perhaps gives another sense of magnitudes for the amount of economic harm a bad actor could cause by sabotaging the system — is presently on the order of $100-$200 billion. The market capitalization of the gold stock is on the order of $7.5 trillion (Cowen, 2017\(^a\)).

More broadly, many have argued that the Nakamoto (2008) blockchain innovation could be useful for domains including global supply chains, land provenance, identity management, medical records, and even voting (cf. Wikipedia (2018) for a variety of references). For example Goldman Sachs (2018) describes the “blockchain technology [that] was originally developed as part of the digital currency Bitcoin” as “The New Technology of Trust”, and describes applications such as “An international ID blockchain, accessible anywhere in the world, [that] allows people to prove their identity, connect with family members and even receive money without a bank account.” While in some cases this use of the word “blockchain” appears to be marketing for older ideas from computer science (e.g., distributed ledgers or databases with known, trusted parties), to the

\(^{11}\)On March 5, 2018, open interest in CME Group Bitcoin Futures was 1622 contracts (CME Group, 2018\(^a\)), each of which tracks the value of 5 Bitcoins, and open interest in CBOE Bitcoin Futures was 5938 contracts (Choe, 2018), each of which tracks the value of 1 Bitcoin. At the March 5, 2018 Bitcoin price of about $11,500, this open interest is worth about $93 million on CME and $68 million on CBOE.

\(^{12}\)March 5 open interest in the CME Gold Futures was 499,602 contracts, each worth 100 troy ounces of gold or $130,000 at the current gold price of roughly $1300 per ounce CME Group (2018\(^b\)).
extent that the Nakamoto (2008) blockchain is used in these wider domains, one should really worry about the value of \( V_{\text{sabotage}} \).

3 Blockchain-Specific Mining Technology

The analysis in Sections 1-2 assumed that the attacker’s cost of waging the majority attack was proportional to the per-block “flow” cost of mining the block chain. Formally, we assumed that the per-block cost of one unit of computational power was \( c = rC + e \), i.e., the rental cost of a chip \( rC \) plus the per-block per-chip cost of electricity \( e \), and that the attacker’s cost of waging the majority attack was \( \alpha \cdot N^*c \), with \( \alpha \) representing the duration of the attack net of block rewards. However, if both (i) the technology necessary for mining the blockchain is specific (i.e., non-repurposable), and (ii) the attack harms the subsequent value of that technology, then it may be appropriate to charge the attacker a stock cost rather than a flow cost. Importantly, (i) and (ii) seem likely to hold for the Bitcoin blockchain at present.

3.1 Flow vs. Stock Costs of Attack

More fully, the flow cost approach used in Sections 1-2 is appropriate under the following four scenarios:

Case #1: The most efficient chips for mining the blockchain in question are repurposable for other uses. This corresponds to the original vision of Nakamoto (2008), who described the blockchain consensus protocol as “one-CPU-one-vote”. However, it does not correspond to Bitcoin mining at present; the most efficient chips are ASICs (application specific integrated circuits) useful only for Bitcoin mining. For Ethereum, the second largest cryptocurrency by market capitalization, the most efficient chips at present are GPUs (graphics processing units), which are repurposable; however, ASICs for the Ethereum blockchain were recently announced, with availability in July 2018 (O’Leary, 2018).

Notably, many cryptocurrency blockchains, including Ethereum’s, are specifically designed to be “ASIC resistant”.13 Roughly, this means that the proof-of-work function is designed so that

---

13The Ethereum Wiki’s main white paper (Ethereum Wiki, 2018a) laments that “the [Bitcoin] mining algorithm is vulnerable to two forms of centralization. First, the mining ecosystem has come to be dominated by ASICs .... this means that Bitcoin mining is no longer a highly decentralized and egalitarian pursuit”. It then argues that in the event that ASICs are designed for Ethereum, that “Ethereum contracts can include any kind of computation, so an Ethereum ASIC would essentially be an ASIC for general computation — i.e., a better CPU.” The recently announced Ethereum ASICs are controversial within the Ethereum developer community, which was reported to have considered modifying the Ethereum protocol to block the use of ASICs (Huillet, 2018). It remains to be seen how repurposable these Ethereum ASICs will be.
ASICS do not meaningfully improve upon general-purpose chips. If a blockchain is in fact ASIC resistant, then the repurposability assumption is appropriate.

Case #2: The most efficient chips are specialized, but there are repurposable chips that are efficient enough for the purpose of an attack. Formally, denote by $c^*$ the per-block per-computational unit cost of the most efficient specialized chip, and denote by $\tilde{c}$ the per-block per-computational unit cost of the best repurposable chip. The cost of an attack using repurposable chips would be $\alpha \cdot N^* \tilde{c}$. If $\tilde{c}$ is within a reasonable factor of $c^*$ then the cost of the attack can reasonably be modeled as a flow cost, not a stock. This is not currently the case for Bitcoin mining, as the best specialized chips are on the order of 1000-10000 times more economically efficient than the best repurposable chips.\(^\text{14}\)

Case #3: The most efficient chips are specialized, and there exist previous-generation specialized chips that are not economically efficient for mining but are efficient enough for the purpose of an attack, and exist in large quantity. Suppose that the most efficient specialized chip costs $C^*$ and uses energy $e^*$, for per-block per-computational unit cost of $c^* = rC^* + e^*$, and that there exists a previous generation specialized chip with per-block per-computational unit energy cost of $\tilde{e}$ such that $\tilde{e} > rC^* + e^*$. That is, the new chip improved on the energy efficiency of the old chip by enough that it would be inefficient to use the old chip for mining even if it were free. The market price of the old chips will therefore be negligible — they are specialized and no longer economically useful for their one purpose. The cost of an attack using such chips would be $\alpha \cdot N^* \tilde{e}$. If $\tilde{e}$ is within a reasonable factor of $e^*$ and there are a large enough number of the previous-generation chips available to amass $N^*$ of computational power, then the flow cost approach is appropriate.

Case #4: The most efficient chips are specialized, there are neither reasonably efficient repurposable chips nor older generation specialized chips, but the attack does not cause a decline in the value of mining equipment, i.e., it is not a sabotage ($\Delta_{\text{attack}} = 0$). In this scenario, flow costs are appropriate because an attacker (especially an insider) would pay a flow cost for the attack, and then could resume mining as usual. Or, the attacker (especially an outsider) could attack the blockchain repeatedly, paying a flow cost each time.

\(^{14}\)The Bitmain Antminer S9 rig, described in more detail in footnote 16 and one of the most efficient Bitcoin ASIC rigs, has a cost of $1160 and a hash rate of 13TH/s, for a capital cost per hash rate of $0.089 $/GH/s. In contrast, most GPUs have hash rates of below 1GH/s (Bitcoin Wiki, 2015; Taylor, 2017) and costs in excess of $100 per chip, for a capital cost per hash rate of more than $100 $/GH/s, a difference of at least 1000x on capital costs. ASICS are also more energy efficient than GPUs.
However, the flow cost approach is not appropriate if:

Case #5: The most efficient chips are specialized, there are neither reasonably efficient repurposable chips nor older generation specialized chips, and the attack is a sabotage. Importantly, case #5 seems to be the accurate case for Bitcoin circa spring 2018. Bitcoin ASIC chips are currently thousands of times more economically efficient for the Bitcoin hashing problem than repurposable chips, and, with the dramatic rise of Bitcoin’s value over the past several years, the Bitcoin ASIC market seems to mostly have been trying to catch up with demand (e.g., Samsung recently announced it is entering (Russell, 2018)), so there is not a glut of previous-generation chips that could be cheaply deployed for an attack.

3.2 Economic Limit of the Blockchain II: $N^*C > V_{sabotage}$

To analyze case #5, consider the extreme in which the attack causes a total collapse of the economic value of the blockchain, including the specialized equipment; this is the case for which the incentive constraint against the attack is least constraining. The incentive constraint (2) can be rewritten, approximately,\(^{15}\) as

$$N^*C > V_{sabotage} \quad (2')$$

In comparison to (2), now there is a stock value on the LHS of the constraint as opposed to a flow value. Whereas the LHS of (2) was likely on the order of a few million dollars (potentially even less), the LHS of (2'), given current Bitcoin mining technology, may be on the order of $1.5Bn-$2Bn.\(^{16}\) Still, equation (2') is a serious constraint on the blockchain. First, the blockchain’s security is still linear in computation expense, as in (2). Second, for (2') rather than (2) to be the relevant incentive constraint, one has to concede both (i) the possibility of sabotage, and (ii) that the security of the blockchain relies on the use of highly specialized equipment.

\(^{15}\)The expression drops the non-capital costs of the attack (i.e., electricity), which are $N^* \cdot A \cdot e$. It also assumes that the benefits of double spending cancel out the cost of the decline of the value of the attacker’s Bitcoin holdings, as explained in footnote 10.

\(^{16}\)The Bitmain Antminer S9, T9+, and V9 are three of the most economically efficient mining rigs we could find among a larger set of rigs with data available on both the rig’s hash rate and its cost. At the March 2018 average level of the Total Network Hash Rate (TNHR) of about 25 million TH/s, one would need to purchase about 1.9 million S9’s, 2.4 million T9+’s, or 6.2 million V9’s to achieve TNHR, which at current consumer posted prices would come to $2.2bn, $1.6bn, and $1.5bn, respectively. (Please see Blockchain.info (2018) for the TNHR and Bitmain (2018) for prices and hash rates for their Antminer rigs). Presumably a large-scale purchaser could obtain Bitcoin ASICs at prices that are lower than the prices that are available to consumers. More importantly, as Bitcoin ASIC technology becomes more mature — e.g., Samsung is now entering Bitcoin ASIC manufacturing (Russell, 2018) — it seems likely that prices will fall, and/or one could cheaply repurpose older ASICs that are economically inefficient for mining (as discussed in case #3) to more cheaply conduct an attack.
3.3 Collapse Scenarios

Suppose, for the purpose of discussion, that the Bitcoin blockchain currently does satisfy constraint \((2')\) but does not satisfy constraint \((2)\), and that the difference between these two constraints helps explain why the Bitcoin blockchain has not had a major attack. What does this analysis then imply about Bitcoin’s security in the future?

The model suggests 3 possible scenarios to worry about:

1. Ultra-cheap specialized ASICs.
   (a) As Bitcoin ASIC technology matures, Bitcoin ASICs become plentiful and very cheap, especially versions that are not at the frontier in terms of energy efficiency. Then we are in case \#3 above.
   (b) If Bitcoin’s value were to fall (for other reasons), that would lead to a glut of specialized ASICs relative to the amount needed per equation \((1)\), even holding ASIC technology per se fixed.

2. Efficient-enough repurposable chips.
   (a) If excitement about the blockchain continues to grow, then it seems plausible that repurposable chips will get more efficient at hashing. They would never catch up with specialized ASICs in terms of economic efficiency, but perhaps the gap closes from a factor of several thousand to something meaningfully smaller. This is case \#2 above.
   (b) Advances in nanotechnology could lead to meaningful improvements in FPGA-like technology, i.e., innovations that aim at the goal of general-purpose thermodynamically optimal computation. This would also be case \#2 above.

3. Economic sabotage becomes sufficiently tempting.
   (a) Bitcoin futures markets grow in economic importance, inducing a short-seller attack.
   (b) Bitcoin grows in economic importance, inducing a sabotage to harm Bitcoin per se.

4 Conclusion

The anonymous, decentralized trust enabled by the Nakamoto (2008) blockchain, while ingenious, is expensive. Equation \((3)\) says that for the trust to be meaningful requires that the flow cost of running the blockchain is large relative to the one-shot value of attacking it. In the double-spending attack considered in Section 2.1, the implication is that the transaction costs of the blockchain must be large in relation to the largest-possible economic uses of the blockchain, which can be interpreted as a large implicit tax. The argument that an attack is actually more expensive
than this flow cost, considered in Section 3, requires one to concede both (i) that the security of the blockchain actually relies on its use of scarce, non-repurposable technology (counter to the Nakamoto (2008) vision of “one-CPU-one-vote”), and (ii) that the blockchain is vulnerable to sabotage, and at a cost that is linear in the amount of specialized computational equipment devoted to its maintenance. These concessions leave the blockchain vulnerable to collapse if either conditions change in the specialized chip market or if the Bitcoin blockchain becomes economically important enough to tempt a saboteur. Overall, the results place potentially serious economic constraints on the use of the Nakamoto (2008) blockchain innovation.

It bears emphasis that the earliest use cases of Bitcoin—black-market transactions, purchases by computer hobbyists, intra-family international transfers, etc., all of relatively modest value—are completely consistent with the model in this paper. In the language of the model, $\tilde{v}_{\text{transaction}}$ is low relative to the acceptable levels of $p_{\text{transaction}}$ for such transactions. Rather, this paper suggests skepticism and caution about larger-scale uses of this technology, such as Bitcoin as a “store of value” akin to gold, or the use of the Nakamoto (2008) blockchain by businesses and governments. Most businesses and governments presumably have access to cheaper forms of data security, e.g., distributed ledgers or databases that require a trusted party (e.g., the business or businesses themselves), rather than having to pay the high costs of the trust that is emergent from a large network of untrusted computers coordinating on maximum proof-of-work.

Relatedly to this last point, an important clarification: as interest in Bitcoin and its blockchain have surged, some have started to use the phrase “blockchain” to refer as well to the use of distributed ledgers or databases among known, trusted parties—that is, without the anonymous, decentralized trust innovation of Nakamoto (2008). An example is the uses by Walmart and British Airways described in Nash (2018)—essentially, the use of well-architected databases, strong version control (possibly utilizing one-way hash functions), and allowing multiple interested parties to easily search or update the data in accordance with prescribed business practices. As one financial columnist astutely observed: “If you announce that you are updating the database software used by a consortium of banks to track derivatives trades, the New York Times will not write an article about it. If you say that you are blockchaining the blockchain software used by a blockchain of blockchains to blockchain blockchain blockchains, the New York Times will blockchain a blockchain about it.”(Levine, 2017) The WSJ reports that “Companies that have taken an ‘If it ain’t broke, don’t fix it’ attitude toward back-office processes and logistics IT might be ready to spend big on updating those systems when they hear the buzzword ‘blockchain.’”(Mims, 2018) As should be quite clear, this paper’s critique is about blockchain in the sense of Nakamoto (2008), not about the use of distributed databases more broadly. Indeed, what this paper highlights is that it is exactly the aspect of Bitcoin and Nakamoto (2008) that is so innovative relative to traditional
An interesting open question raised by this paper — perhaps more for computer scientists than for economists, or perhaps requiring both perspectives — is whether there is some other approach to generating anonymous, decentralized trust in a public ledger that is less economically constrained by the possibility of an attack. More precisely, allowing that some version of equations (1)-(3) seems intrinsic to any anonymous, decentralized blockchain protocol, is there an alternative to Nakamoto (2008) that either reduces $V_{\text{attack}}$ or raises $\alpha$, relative to a given level of payment for maintenance of the ledger, $P_{\text{block}}$. Within the proof-of-work paradigm, the most natural idea is to find a modification to the longest-chain convention that utilizes the fact that, in the event of an attack, it will be widely “noticed”. Or, perhaps one can prove a theorem that shows that no such modification can exist while preserving anonymity and decentralization, suitably defined. Another interesting idea in this regard is proof-of-stake (cf., Buterin and Griffith (2017), Ethereum Wiki (2018b)). The usual motivation for proof-of-stake over proof-of-work — the deadweight loss and environmental harm associated with proof-of-work mining, currently estimated to utilize over 0.30% of global electricity consumption (Digiconomist, 2018; Vries, 2018; Saleh, 2018) — is in fact completely orthogonal to the concerns raised in this paper. Just conceptualize $c$ as the per-block opportunity cost of holding one unit of stake, and versions of equations (1)-(3) obtain immediately. But, the use of “stakes” instead of computational work may open new possibilities for thwarting attacks, e.g., confiscation of an attacker’s stake, or building some limited forms of reputation (e.g., Buterin, 2016). It will be interesting to watch this research develop, and see whether or not it constitutes a valid response to the critique in this paper.
References


### Table 1

**Majority Attack Computational Simulations**

<table>
<thead>
<tr>
<th></th>
<th>(e = 0)</th>
<th>(e = 1)</th>
<th>(e = 6)</th>
<th>(e = 10)</th>
<th>(e = 100)</th>
<th>(e = 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Expected Duration of Attack</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A = 1.05)</td>
<td>26.50</td>
<td>30.80</td>
<td>46.02</td>
<td>55.48</td>
<td>182.4</td>
<td>1,068.8</td>
</tr>
<tr>
<td>(A = 1.1)</td>
<td>14.02</td>
<td>16.43</td>
<td>25.50</td>
<td>31.35</td>
<td>126.8</td>
<td>1,004.9</td>
</tr>
<tr>
<td>(A = 1.2)</td>
<td>7.79</td>
<td>9.28</td>
<td>15.37</td>
<td>19.66</td>
<td>106.1</td>
<td>1,002.0</td>
</tr>
<tr>
<td>(A = 1.25)</td>
<td>6.54</td>
<td>7.86</td>
<td>13.41</td>
<td>17.43</td>
<td>103.8</td>
<td>1,002.0</td>
</tr>
<tr>
<td>(A = 1.33)</td>
<td>5.34</td>
<td>6.49</td>
<td>11.58</td>
<td>15.40</td>
<td>102.5</td>
<td>1,002.0</td>
</tr>
<tr>
<td>(A = 1.5)</td>
<td>4.08</td>
<td>5.07</td>
<td>9.77</td>
<td>13.49</td>
<td>102.0</td>
<td>1,002.0</td>
</tr>
<tr>
<td>(A = 2)</td>
<td>2.89</td>
<td>3.78</td>
<td>8.39</td>
<td>12.23</td>
<td>102.0</td>
<td>1,002.0</td>
</tr>
<tr>
<td>(A = 5)</td>
<td>2.12</td>
<td>3.06</td>
<td>8.00</td>
<td>12.00</td>
<td>102.0</td>
<td>1,002.0</td>
</tr>
</tbody>
</table>

|                     |            |            |            |            |             |             |
| **B. Expected Computational Cost of Attack** |            |            |            |            |             |             |
| \(A = 1.05\)       | 27.83      | 32.34      | 48.32      | 58.26      | 191.5       | 1,122.2     |
| \(A = 1.1\)        | 15.42      | 18.07      | 28.05      | 34.48      | 139.5       | 1,105.4     |
| \(A = 1.2\)        | 9.34       | 11.13      | 18.44      | 23.59      | 127.4       | 1,202.4     |
| \(A = 1.25\)       | 8.18       | 9.82       | 16.76      | 21.79      | 129.7       | 1,252.5     |
| \(A = 1.33\)       | 7.11       | 8.64       | 15.40      | 20.48      | 136.3       | 1,332.7     |
| \(A = 1.5\)        | 6.12       | 7.61       | 14.65      | 20.24      | 153.0       | 1,503.0     |
| \(A = 2\)          | 5.78       | 7.56       | 16.78      | 24.45      | 204.0       | 2,004.0     |
| \(A = 5\)          | 10.59      | 15.29      | 40.01      | 60.00      | 510.0       | 5,010.0     |

|                     |            |            |            |            |             |             |
| **C. Expected Cost of Attack, Net of Block Rewards** |            |            |            |            |             |             |
| \(A = 1.05\)       | 1.34       | 1.55       | 2.33       | 2.80       | 9.2         | 53.5        |
| \(A = 1.1\)        | 1.40       | 1.64       | 2.55       | 3.13       | 12.7        | 100.5       |
| \(A = 1.2\)        | 1.56       | 1.86       | 3.07       | 3.93       | 21.2        | 200.4       |
| \(A = 1.25\)       | 1.64       | 1.96       | 3.35       | 4.36       | 25.9        | 250.5       |
| \(A = 1.33\)       | 1.76       | 2.14       | 3.82       | 5.08       | 33.8        | 330.7       |
| \(A = 1.5\)        | 2.04       | 2.54       | 4.88       | 6.75       | 51.0        | 501.0       |
| \(A = 2\)          | 2.89       | 3.78       | 8.39       | 12.23      | 102.0       | 1,002.0     |
| \(A = 5\)          | 8.47       | 12.23      | 32.01      | 48.00      | 408.0       | 4,008.0     |

*Notes:* Panel A is reported in number of blocks. Panels B and C are reported in units of \(P_{\text{block}} = N^c\). The probability that the attacker wins in the k-th period is estimated by simulating 100M attacks. For each set of parameters, the simulation ends after the larger of 10,000 periods or the period at which 99.999% of attacks have become successful; unresolved attacks are abandoned, incurring compute costs without block rewards. Simulation assumptions: (1) purchased goods/assets are held in escrow for \(e\) periods; (2) attacker solves blocks in \(\text{Exponential}(A)\) time at computation cost \(A\) per block, honest CPUs solve blocks in \(\text{Exponential}(1)\) time; (3) an attack is successful after both (i) all goods/assets have been released from escrow and (ii) attacker has solved more blocks than honest CPUs.