On hostile blockchain takeovers

or Goldfinger attacks revisited

Joseph Bonneau

March 2017

Most research modelling Bitcoin-style distributed consensus protocols (sometimes called “Nakamoto consensus”) has focused on attempts to prove incentive compatibility. That is, models attempt to prove that under certain assumptions about attacker motivation a protocol will exhibit desired stability properties such as an exponentially low probability of long chain forks or a distribution of mining rewards that is close the amount of work contributed (called fairness or chain quality). Typically, models assume that the utility function for all participants in the system is the amount of monetary rewards acquired within the protocol (e.g. for Bitcoin, the amount of mining rewards earned denominated in BTC). This leads to the most tractable models.

It is often acknowledged that a more realistic utility function is monetary rewards denominated in an external currency (such as US dollars). For this reason, some mining strategies which deviate from the standard protocol and lead to increased in-system rewards may yield less utility if they affect the exchange rate and therefore provide fewer ex-system rewards. However, modeling the impact of miner behavior on exchange rates is difficult, so this analysis is usually qualitative.

Rarely considered is a miner whose goal is not simply to acquire monetary rewards, but to destabilize a blockchain (even at a financial loss). Such a miner, who can fairly be called an *attacker* to the system, was called a *Goldfinger attacker* by Kroll and Felten (KF 2013). They can also be conceptualized as a *hostile takeover*, a term that is more appropriate for proof-of-stake systems.

I argue that revisiting the dynamics of a Goldfinger-style attack may yield new insights into the stability of blockchain protocols. In particular, it provides an interesting comparison between ASIC-dominated blockchains (such as Bitcoin), commodity hardware-dominated blockchains (such as Ethereum), and proof-of-stake systems.

Methods of obtaining mining capacity

For an attacker aiming to subvert a proof-of-work blockchain, they must obtain control of a large amount of mining capacity. We can consider whether the attacker is obtaining mining capacity permanently or temporarily, and whether they are introducing new capacity into the system or capturing existing mining capacity. This yields four basic attack strategies:

	Obtain new capacity	Obtain existing capacity
Temporary control	Rent	Bribe
Permanent control	Build	Buy out

Note that an immediately that a difference emerges between three types of systems:

- For ASIC-dominated blockchains, such as Bitcoin, the *rent* strategy is not possible because there is a negligible amount of Bitcoin mining hardware that is not already dedicated to Bitcoin mining.
- For pure proof-of-stake blockchains, neither *rent* nor *build* are possible, as the “capacity” in the system is fixed.

We can make some initial observations about each approach.

Rental attacks (on Ethereum)

Rental is only possible for commodity-hardware mined blockchains. Ethereum fits this description today as mining is dominated by graphics cards (GPUs). An attack would consist of renting a large amount of capacity from a system such as Amazon’s Elastic Compute Cloud (EC2). Currently, EC2 rents NVIDIA K80 GPUs for about US$0.20 per hour at spot prices (bulk discounts are available), which can perform about 24 MH/s, a little more than 1 millionth of the Ethereum network hash rate. So, as a very rough estimate for about US$400,000/hr an attacker could rent enough hardware to perform a 51% attack on Ethereum. Presumably only a few hours of such an attack would be sufficient to cause a major loss in value to the system, which has a market cap of over $2.5 billion. Thus, it appears that Ethereum is relatively vulnerable to Goldfinger attacks.

It is worth noting that GPU rental is relatively inefficient. Currently Ethereum miners earn roughly $40,000/hr in block rewards, whereas renting this capacity on EC2 would cost 10x those rewards. This premium means, however, that the attack has no long-term risk for the attacker.

Building attacks

Consider the cost of building enough new mining capacity to subvert Bitcoin. We can take as a representative example the AntMiner S7, a recent ASIC miner. It retails for about US$500 and can perform nearly 5 TH/s. Conveniently, this is about one millionth of the network hash rate-implying an upfront capital cost of about $500 million to obtain enough hash power to perform a 51% attack on Bitcoin. Of course, this figure is very approximate and it would be far cheaper to buy this hardware in bulk. Still, it appears to be roughly 3 orders of magnitude more expensive to perform such an attack on Bitcoin than to perform the rental attack described above on Ethereum. Bitcoin's market cap is a little less than an order of magnitude higher, but this still implies a capacity-building attack is 2 orders of magnitude more expensive (not to mention considerably slower and more complex logistically) to execute. This is an argument in favor of ASIC-friendly mining puzzles.

As for building attacks on Ethereum, the most efficient GPU hardware available currently costs about $10/MH/s, suggesting a cost of about $200 million to build towards 51% capacity. Note that this is several times more expensive (relative to the market cap) than for Bitcoin. Perhaps more capacity has been built for Ethereum as it is recyclable.

Bribery attacks

Mechanisms for bribery attacks were considered by Bonneau (Bonneau 2016). There are several approaches, including direct bribery, running a mining pool that pays excess rewards, smart contracts which deliver payment, or leaving bribes available on an attacker's fork. It is difficult to estimate the premium an attacker would need to pay to break miner loyalty and convince them to work on a fork that would be highly detrimental to the system. With negligible premiums, bribery is very cheap, requiring only half of the rate of network rewards (about $40,000/hr for Ethereum or $100,000/hr for Bitcoin). Presumably, similar economics apply to proof-of-stake systems.

Buy-out attacks

Buy-out attacks would involve either purchasing mining capacity from current owners or purchasing currency (in a proof-of-stake system). The cost of such an attack appears straightforward. For proof-of-work systems, it should cost about half of the net present value of all future mining rewards (with a steepened discount rate due to reflect likely future growth in network capacity). For proof-of-stake systems, half of the value of the system must be bought up.

It appears that proof-of-stake systems are much more secure here, as the attacker must buy half of all value of the system, whereas with proof-of-work the attacker must only buy half of the future mining rewards (which must be strictly less valuable than the entire market cap). In this case the term "hostile takeover" seems appropriate.

In either case, there is an interesting possibility of a *race to the door* among current capacity owners. Imagine that an attacker credibly announces they will buy out half of all capacity and then use it to destroy the system. Current capacity owners will have a strong incentive to sell to avoid being left in the 49% which does not sell and hence loses everything. As they being to sell and the attack appears more likely to succeed (which is easy for the attacker to signal as the amount of capacity grows) this could lead to a death spiral of lowered prices and increased confidence the attack will succeed.

Commodity proof-of-work systems appear less likely to suffer from a race-to-the-door, since capacity owners who do not sell to the attacker can still sell their hardware even if the attack succeeds.

Countermeasures

For all of the attack models, there is the possibility of countermeasures by current capacity owners. Current owners can respond in kind to building, renting, or bribing. With buy-outs, they can attempt to set a market floor by offering to buy more capacity themselves. This may be a profitable strategy-if a race-to-the-door is in progress which has lowered the value of capacity, it may be profitable to buy if the attack fails and prices rebound.

Note that an attacker may respond to a buy-out attack by building (or renting) new hardware. This may be a wise strategy for a coalition of miners who would otherwise be stuck with a worthless 49% mining share after a successful attack. This countermeasure is not possible for proof-of-stake systems, in which a successful buy-out attack will be permanent.

Comparison

At first glance, proof-of-stake systems appear less vulnerable to Goldfinger attacks. They are not vulnerable to rental or building attacks. Bribery attacks appear similar, while buy-out attacks appear strictly more difficult. However, proof-of-stake is more fragile in that building new capacity is not available as a countermeasure.

Commodity proof-of-work systems appear more resilient to buy-out attacks as a race to the door is less likely to develop. However, ASIC proof-of-work systems are not vulnerable to rental attacks.

Open questions

- Is the cost of Goldfinger attacks a useful lower bound on the security of a given system?
- Is there a strict ordering between the three main types of system considered here in terms of resilience to Goldfinger attacks? Or are they incomparable?
- Which attack strategy is the most plausible in practice?
- Is there a minimum amount of reward miners should receive (relative to the market cap of the coin) for security purposes, to ensure the disincentive to sell is high? Or will this simply cause more capacity to be built?