It will cost you nothing to ”kill” a Proof-of-Stake crypto-currency

Nicolas Houy

GROUPE D'ANALYSE ET DE THÉORIE ÉCONOMIQUE LYON - ST ÉTIENNE

GROUPE D'ANALYSE ET DE THÉORIE ECONOMIQUE
Lyon Saint-Etienne

WP 1404

It will cost you nothing to "kill" a Proof-of-Stake crypto-currency

Nicolas Houy

January 2014

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France
Tel. +33 (0)4 72 86 60 60
Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France
Tel. +33 (0)4 77 42 19 60
Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr
Téléchargement / Download : http://www.gate.cnrs.fr – Publications / Working Papers

It will cost you nothing to "kill" a Proof-of-Stake crypto-currency [v.0.1]

Nicolas HOUY*

February 6, 2014

Abstract

It is a widely spread belief that crypto-currencies implementing a *proof of stake* transaction validation system are less vulnerable to a 51% attack than crypto-currencies implementing a *proof of work* transaction validation system. In this article, we show that it is not the case and that, in fact, if the attacker's motivation is large enough (and this is common knowledge), he will succeed in his attack at no cost.

JEL Classification: G23, Z00.
Keywords: Bitcoin, protocol, proof of work, proof of stake, 51% attack.

1 Introduction

Bitcoin has become increasingly popular in 2013 even though it has been invented in 2008, [Nakamoto, 2008]. It is usually described by laymen as an electronic money even though this definition is much criticized by the computer science community that rather talks about a revolutionary protocol. At its core, Bitcoin allows to secure property rights, in a decentralized peer-to-peer network, on tokens (bitcoins[1]) produced in limited quantity. There exist markets to purchase and exchange bitcoins. At the time this article is written, there are about 12.3 millions bitcoins in circulation and they can be exchanged at about $850 per bitcoin. Whether we consider its theoretical aspects or its use as a currency or as an asset, economists should be interested by this new "Unidentified Financial Object". In this article, we study a particular aspect of the Bitcoin technology that is much debated in the crypto-currency community with tools borrowed from the economic science.

In order to do that, we need to describe a bit how Bitcoin actually works. When an individual sends some bitcoins to another individual, this information is broadcast to the Bitcoin network. However, for technical purposes we won't address here, this transaction, treated in a block with other transactions, needs to be inserted in the blockchain in order to be confirmed and secured. The blockchain is a public ledger that contains the history of all the transactions in bitcoins ever processed. It is the role of the miners to do this work of confirming and securing transactions. Practically, this mining process consists in solving a mathematical problem and the first miner to do so, technically to bring a proof-of-work (POW), can insert a set of transactions in the blockchain. As it requires computational

*Université de Lyon, Lyon, F-69007, France; CNRS, GATE Lyon Saint-Etienne, Ecully, F-69130, France. E-mail: houy@gate.cnrs.fr.

[1] As the norm tends to be, we will write "Bitcoin" for the network or the protocol and "bitcoin" for the tokens that circulate on it.

resources, the successful miner is rewarded in bitcoins for his useful work. In order to control the monetary base, mining is made more complex than it could be. And since the probability for each miner to solve the mining problem depends on his computational power, the mining complexity is made dependent on the total computational power of the miners. To sum up, for POW crypto-currencies, including Bitcoin, miners are in competition to solve a problem needed to confirm and secure transactions. The first miner to solve the problem earns a reward. The problem is made artificially complex in order to control the monetary base. This process is described as brilliant by some but it has been criticized for the inefficiency due to the loss of resources it induces (see [Krugman, 2013] for instance). Indeed, Bitcoin miners have engaged in an arm race to computational power and in the end, much hardware, engineering and power are used in order to solve mathematical problems that are artificially made extremely complex.

As it requires trust in the system to be adopted, Bitcoin is open-source. Hence, many alternative crypto-currencies have been proposed at almost no cost. Each supposedly solves some Bitcoin flaws. Naturally, some of those crypto-currencies try to tackle the problem of the inefficiency due to the POW aspect of Bitcoin. Most of these crypto-currencies are based on another mining process, called proof-of-stake (POS). For the sake of simplicity and with a slight lack of rigor, let us just say that with POS, the expected reward for inserting transactions in the blockchain does not depend on the computational power of miners but on the amount of crypto-currency they already own. Peercoin and Nxtcoin are two alternative crypto-currencies that use POS (the former partially, the later completely[2]).

Let us now explain a weakness of all crypto-currencies. Roughly speaking, regardless on it using POW or POS, any crypto-currency cannot be trusted if one individual can mine too many blocks in expectation (see [Kroll *et al.*, 2013], [Eyal and Sirer, 2013]). In a POW crypto-currency, the condition of what is called a ”51% attack” and that would totally undermine the value of the money, is that an individual owns strictly more than 50% of the total computational power of the network. In a POS crypto-currency, the same attack would happen if an individual owns strictly more than 50% of the monetary base.[3] It is believed in the crypto-currency community that a 51% attack is less likely to occur in a POS system than in a POW system because it would be more expensive (in direct and opportunity costs) for a malicious agent to buy 50% of a POS crypto-currency than 50% of the computational power of a POW network (see [Bitcoin Wiki, 2014] for instance). In this article, we show that not only this is not the case under some conditions but even that it would cost nothing for a malicious agent to buy 50% of a POS crypto-currency monetary base.

2 Model

Let us consider a set of $N + 1$ agents with $N > 2$. Each agent is indexed by an integer in $\{0, ..., N\}$. There are two goods in the economy, a crypto-currency (CC) and money. There is no money liquidity constraint. Each agent is initially endowed with one unit of CC. CC yields

[2] As Nxtcoin is a 100% POS protocol, for reasons that would bring us too far, its monetary base could not be controlled if it was working exactly as we describe in this article. This is why, its creators have fixed the number of nxtcoins since its launch and all nxtcoins have been premined. This technical details do not invalidate our study and can be easily ignored.

[3] We use the simplification that the fatal threshold remains 50% for a POS crypto-currency. In fact, this depends on some rule that we don't describe here. However, our argument remains valid even if this assumption is not made.

a monetary interest r for each unit of time. This interest rate embodies the utility that can be extracted from using CC as a mean of exchange. The time discount factor of all agents is β. CC loses all its utility whenever an agent holds strictly more than half ($(N+1)/2$) of the CC units. Agent 0 has a special interest in killing the CC we study. Hence, he earns U if an agent holds strictly more than $(N+1)/2$ units of CC.

CC can be exchanged on a market. We are especially interested in situations where one agent (specifically agent 0) may be willing to hold more than half of the CC quantity. Hence, we cannot just use the usual supply-demand model that makes the assumption that agents are atomistic. We need to go further in the description of the market. At each time step, agent 0 is matched with the same probability with any other agent that holds some CC unit, say i. Agent 0 makes a "take or leave" price offer to i in order to buy his unit of CC. i accepts or not the offer. Exchange takes place or not depending on the offer by 0 and the acceptance decision by i. The time step between two matching is dt, arbitrarily short. We will denote $V(n)$ the expected discounted future flow of money earned by any agent $i > 0$ holding one unit of CC where n is the number of CC units held by 0. $V_0(n)$ is the expected discounted future flow of money earned by agent 0 where n is the number of CC units he holds.

Obviously, the step "offer by 0, take or leave by i" has a simple outcome: either 0 makes the cheapest offer that will be accepted or he makes an offer that will be rejected. In the first case, i's unit of CC changes hand for $(1-\beta dt)V(n)$.

Once this step outcome is computed, we can simply write the dynamics of V and V_0. Precisely,

$$V_0(n) = \begin{cases} U & \text{if } n \geq \frac{n+1}{2} \\ n.r.dt + (1-\beta.dt)\max\left\{V_0(n+1) - V(n), V_0(n)\right\} & \text{otherwise} \end{cases}$$

and

$$V(n) = \begin{cases} 0 & \text{if } n \geq \frac{n+1}{2} \\ r.dt + (1-\beta.dt)\begin{pmatrix} p(n)V(n)+ \\ (1-p(n))((1-P^e(n))V(n) + P^e(n)V(n+1)) \end{pmatrix} & \text{otherwise} \end{cases}$$

where $p(n) = 1/(N-n)$ is the probability for any agent holding some CC to be matched with agent 0 when the latter holds n units of CC and $P^e(n)$ is the belief that an agent different from 0 has that agent 0 will buy one more unit of CC when he already holds n.

Let us first solve the problem for n the greatest integer smaller than $(N+1)/2$, $n = \lfloor (N+1)/2 \rfloor$. There exist two possible equilibria. The first one is with $P^e(n) = 0$, $V(n) = r/\beta$ and $V_0(n) = nr/\beta$. This equilibrium is subgame perfect if and only if $U \leq r(n+1)/\beta$. The second equilibrium is with $P^e(n) = 1$, $V(n)$ arbitrarily close to 0 when dt tends to 0 and $V_0(n)$ arbitrarily close to U. This equilibrium is subgame perfect if and only if $U > r(n+1)/\beta$. Let us now solve our game one time step before. Again, there exist two possible equilibria and these are the same as above. The first one is with any $P^e(n)$, $V(n) = r/\beta$ and $V_0(n) = nr/\beta$. This equilibrium is subgame perfect if and only if $U \leq r(n+1)/\beta$. The second equilibrium is with $P^e(n) = 1$, $V(n)$ arbitrarily close to 0 when dt tends to 0 and $V_0(n)$ arbitrarily close to U. This equilibrium is subgame perfect if and only if $U > r(n+1)/\beta$. The same reasoning can be made for all preceding steps.

Then, there are two equilibria for our game. In the first one, when $U > r(N+1)/\beta$, agent 0 buys strictly more than half of the coins and actually kills the CC. Since this is anticipated by all the other agents, the latter are in competition to sell to agent 0 their coins, who they know have already no value. The attack can be undertaken at no cost.

In the second equilibrium, even if 0 accumulates enough CC coins, he will have no incentive to cross the 50% threshold because it is better for him to keep the coins and receive the interest flow that goes with it rather than kill the CC at the expense of this flow. Anticipating this, the other agents are not ready to sell the CC units below their value, r/β.

3 Discussion

With a simple (one could say simplistic) model, we showed that the belief, widely spread in the computer science community, that POS crypto-currencies are immune to a 51% attack because of the supposedly too high cost to buy half of the coins is flawed. Indeed, the underlying reasoning does not take into account the fact that if the attack is undertaken by someone credibly willing to really kill the crypto-currency, agents should anticipate that their coins are worthless since the start and should practically sell them for nothing to the attacker.

A more realistic model would take into account differentiated beliefs about the attacker's motivations (U) and hence Bayesian updating of this, liquidity constraints, different beliefs about the future value of the crypto-currency without attack... We chose our market model with a special care for simplicity. We checked that results are unchanged for other market structures. In particular, our results would be unchanged if the potential attacker was the Stackelberg follower in the "take or leave" step. The basic requirement needed to get our results is simply that sellers are in competition in front of the attacker. Whenever the latter is credible, the CC has already lost its value, there is no need to wait for the attacker to actually buy the CC.

We believe that, in the first approximation at least, we should consider that POS implies high vulnerability to 51% attacks and not see POS as a viable alternative to POW at least in this regard. Notice that our model cannot be applied to POW. Indeed, with POW, agents invest a high fixed cost in computational power and only suffer a very marginal cost to mine. In this case, an attacker would have to actually spend a high fixed cost to gain more than 50% of the network computational power. The announcement of the attacker's motivation, even if credible, would not be enough for other agents to give up their resources.

References

[Bitcoin Wiki, 2014] Bitcoin Wiki "Proof of Stake" page. https://en.bitcoin.it/wiki/Proof_of_Stake. Retrieved on 02/05/2014.

[Eyal and Sirer, 2013] Eyal I. and Sirer E.G. (2013) "Majority is not enough: Bitcoin mining is vulnerable", arXiv: 1311.0243.

[Kroll *et al.*, 2013] Kroll J.A., Davey I.C. and Felten E.W. (2013) "The economics of Bitcoin mining, or Bitcoin in the presence of adversaries", *Mimeo.*

[Krugman, 2013] Krugman P. (2013) "Adam Smith hates Bitcoin". NYTimes blog. http://krugman.blogs.nytimes.com/2013/04/12/adam-smith-hates-bitcoin/

[Nakamoto, 2008] Nakamoto S. (2009) "Bitcoin: A peer-to-peer electronic cash system".