Sakura
Long term UTXO recycling mechanism for Bitcoin

By Joannes Vermorel, April 2nd, 2018

Acknowledgements: based on an original idea of Amaury Séchet

Status: EARLY DRAFT, INCOMPLETE (PROOF MISSING)

Abstract: The very long term viability, centuries ahead, of Bitcoin depends on preventing the runaway growth of the UTXO set (unspent transaction outputs). Also, the ever shrinking monetary mass of Bitcoin is a complication which hinders economic agents from fully relying on perfectly predictable monetary conditions. Here, we propose *Sakura,* a long term recycling mechanism to prune “dead” UTXO entries, defined as entries that have remained untouched for 80 years (defined as 4,200,000 blocks). It allows those “dead” entries to re-enter the pool of mining rewards, on top of the normal halving mechanism which normally occurs every 210,000 blocks. Sakura comes with a trigger condition that “dead” UTXO entries should represent more than 50% of the UTXO set. This condition prevents a premature activation of the change of consensus if there is not enough economic gains to justify the change. Sakura proposes an exponential decay mechanism associated to a half-period of roughly 20 years. The paper also presents a discussion to justify why those seemingly arbitrary choices are made.

Overview

In 2008, Satoshi Nakamoto has introduced Bitcoin, a P2P cash system that distributes the money emission process while rewarding Bitcoin miners for the work they provide in securing the transactions and preventing double-spend from happening. The monetary rewards are following an exponential decay scheme where the miner’s reward for a block is halved every 210000 blocks, which represents about 4 years of time, as mining difficulty is adjusted to keep blocks 10 min apart on average.

There are two reasonable objections to the very long term viability of Bitcoin. First, the UTXO set is growing forever, ultimately bankrupting the miners themselves. Second, the ultimate extinction of the block reward will force Bitcoin to transition toward a fee-market which is a fundamental alteration of how Bitcoin has successfully operated since its inception in 2008.

Concerning the runaway growth of the UTXO set, every time a user loses its private key, this user leaves an entry to remain forever in the UTXO set. While most users will succeed at properly backuping their private keys, hence protective of their money, users should also be expected to lose their keys albeit infrequently, especially if the amounts at risk are very low.

Self-preservation dictates that miners will try not to bankrupt themselves by letting too many entries enter the UTXO set. Yet, it can also be noted that, presently, miners do not have a way to prevent a forever growing UTXO set.

Also, honest but slightly imperfect miners may incorrectly anticipate future hardware improvements that may fail to materialize. While rational miners would not do anything that would bankrupt themselves within a decade or so, there is no reason to think that the rationality of miners is perfect either. Thus, there is plenty of room for subtle (or not so subtle) errors to be made by the Bitcoin miners themselves over the centuries to come which would make them ultimately incapable of remaining a viable economic solution to money, forcing the world to transition to another form of money, Bitcoin v2, where this very problem has been fixed - if only by resetting the UTXO set to its original virgin state.

Concerning the transition of Bitcoin toward the fee-market, it is a fundamental alteration of economic dynamics of Bitcoin because hashpower only plays a much diminished value from an operational perspective. Indeed, the mining competition would essentially be redirected toward who's the best at managing the blockchain itself, which is an *economy-of-scale* game where the biggest player has a natural advantage over its competitors. Each miner has to pay a fixed cost to run a Bitcoin node and manage its copy of the blockchain, while the payouts are proportional to the hashrate. Since the inception of Bitcoin, the operational costs of Bitcoin nodes have remained very small compared to the money emission rewards. However, this may change in the future if the UTXO set becomes extremely large.

The Sakura scheme is a definitive answer to the runaway growth of the UTXO set, and a partial counter to the transition to a pure fee-market. Sakura proposes a change of the Bitcoin consensus to let miners claim for themselves very old UTXO entries, untouched for 80 years. Sakura emphasizes an approach intended to minimize any economic disruption of either Bitcoin or the economy at large, but also an approach intended to minimize the block-to-block impact of the scheme itself on the UTXO set.

Sustainability is highly valued by markets

In the long run, our children live.

Modern markets are given ample proof that a sizeable portion of the population of developed countries deeply care about economic approaches that are perceived as sustainable in the very long run, well beyond our current lifetimes. The point of the author here is not to prove that any of such opinions are scientifically true (ex: danger of climate change, need for organic food,

rejection of civil nuclear power, preservation of super-predators, etc.) but simply that there is ample empirical evidence that individual choices that results from those beliefs are economically massive.

Proving the market interest in sustainable approaches would require lengthy discussions that go well beyond the scope of this paper. Let's simply point out two anecdotal, yet large, pieces of evidence. First, Tesla, in 2017 (at $51 billions, see references), was valued more than GM, while shipping 130 times *less* vehicules. It's relatively self-evident that most of the appeal behind Tesla comes from the fact that its vehicles are *electric*, signaling to the market an expected virtue of sustainability. Second, the worldwide organic food market has surpassed $70 billions in 2017.

Then, within Bitcoin, the major fork that happened in 2017, resulting in the creation of Bitcoin Cash and Bitcoin Core was partially driven by fundamentally divergent opinions on the economic viability of large blocks.

The point we are trying to convey here is subtle: there is a need to have consensus in place - decades ahead of time - to counter runaway UTXO growth, *if such growth were to ever materialize*. This opinion does *not* state that the Sakura scheme *must* be implemented, only that there should be a consensus that if runaway UTXO growth were to be observed, *then* the Bitcoin community would react according to plan and not let itself get disrupted through indecision which would invariably result in a highly disruptive fork. This approach gets the best of both worlds: it is provably sustainable for Bitcoin, while effectively requiring nothing to be implemented in the short term.

Trigger condition for Sakura

We propose to have the Sakura scheme (detailed below) triggered on the *second halving event* that follows the first block meeting the condition: *within the UTXO set, over 50% of the entries are over 4,200,000 blocks old*.

First, this condition ensures that Sakura never gets triggered unless there is a massive economic pressure at stake. There is no point in introducing any consensus change unless there is a provable massive gain to be expected from the change itself. If Bitcoin becomes the money of the world, the UTXO set will be massive; thus the economic benefits associated a halving of the UTXO set will be equally massive.

Second, this condition ensures that the Bitcoin community is given, at least, 4 years to effectively roll-out the Sakura scheme. This delay is intentionally long in order to give markets plenty of time to adjust themselves toward the upcoming update of the consensus.

The Sakura scheme

The Sakura scheme proposes to address those two problems at once, without fundamentally altering the *social economic contract* that Bitcoin has offered to its shareholders since 2008. Sakura proposes to put an expiration date on every single UTXO entry approximately equal to 80 years. UXTO entries older than 80 years would be discarded, and their monetary value would be pooled to be ultimately re-emitted by the miners themselves.

The Sakura scheme operates by as follow:

- For each block,
    - Each UTXO entry that is over 4,200,000 block old (inclusive) *must* be pruned from the UTXO set. The UTXO commitment is updated accordingly.
    - The satoshis associated to each pruned UTXO entry is added to M_n , the *Sakura Fund* associated with the current block halving period.
- For each block halving event
    - Let R_n be the new block reward at the n^{th} block halving event.
    - The Sakura reward is computed as $S_n = M_n / (8 * 210{,}000) + S_{n-1} * (1 - 1/8)$ where S_{n-1} is the Sakura reward computed at the previous halving event.
    - The revised per block reward is redefined as $R_n = R_{n-1} / 2 + S_n$.

This scheme introduces many numbers that may appear arbitrary. However, we believe that those numbers are grounded, and will effectively help Bitcoin to preserve all its economic properties over very long periods of times, centuries or more.

The Sakura re-definition of the Bitcoin reward is strictly compatible with all existing Bitcoin implementations until 2088. Indeed, UTXO entries won't be able to expire before 2088, and thus, the value of S_n remains at zero until this point of time is reached.

The mandatory pruning of all UTXO entries, block per block, is intended to put an upper limit to the amount of changes to be brought to the UTXO set itself within the time period of a single block. Indeed, if we were to choose to perform the pruning only at a specific point of time, say the block halving event, we would most certainly create scalability problems to process those "halving" blocks.

Proof: Sakura does not modify the monetary mass of Bitcoin

TODO: do the math, and make sure the constants are correct

Proof-of-work on the user's side

Bitcoin is not intended as substitute for gold with wealth that could sit at the bottom of the ocean for millenia and still retain its value. Bitcoin is cash, and cash is fundamentally attached to "live" economic agents (as opposed to "dead" ones).

Bitcoin is putting the burden of the Proof-of-Work on miners, but why should users be forever dispensed from any proof-of-work? Requesting from users at least one transfer every 80 years can be seen as an ultra-low intensity proof-of-work burden put on the users. This proof-of-work remedies the *tragedy of the commons* where users can otherwise treat the UTXO set as a free data storage, subsidized in full by the Bitcoin miners.

Minimizing the miner reward variance

There are many schemes that can be considered to redistribute the satoshis originating from the pruned UTXO entries. In order to assess which scheme is superior, we propose that, among other criterions, the miner reward variance should be minimized.

It would be possible to implement a rule that lets a miner claim any UTXO entry older than 4,200,000 blocks for himself. However, during the first 4 years of Bitcoin, users were rather careless with their keys. As a result, it's reasonable to assume that UTXO entries carrying over 1000 BCH may expire and be pruned - while the monetary amount they represent in 2088, assuming that Bitcoin as established itself as the world currency, is considerable.

Thus, letting miners claim for themselves - one way or another - hyper-blocks, possibly offering 10,000x the reward of a regular block may incite them to apply bad behaviors and collusion in particular. One the surface, this problem seems similar to the double-spend problem. However, it is not exactly the case.

In the case of a double-spend, one person or organization gets defrauded. If we are considering a considerably large payment (say 100M USD or more), then, while a miner might theoretically pull off a large double-spend attack acting as the buyer, it appears implausible, in real life, that the seller would not seek remedy in court. All the seller has to do is simply to sue to invalidate his part of the transaction, for example the transfer of ownership of a large company to the buyer. While Bitcoin transactions can be made largely anonymous, the transfer of ownership of pretty much any considerable asset (bitcoins aside) requires the two participants to identify themselves to each other. Then, even assuming that double-spends could not ever be secured for the Bitcoin network when very large amounts are involved, the seller could simply request the payments to be fragmented in many transactions spreading over time. This option is not available if a miner can proceed and claim an expired UTXO entry for himself.

The Sakura scheme provides a mechanism that minimizes the amount of variance in the miner reward originating from the Sakura scheme itself. It is also aligned with the original Bitcoin emission mechanism itself.

Why not a reward adjustment at every block

The satoshis collected from the expired UTXO entries should be redistributed over time in order to minimize the miner's variance as outlined in the previous section. Any series associated with a convergent partial sum would fulfill this requirement. However, as the Bitcoin money emission mechanism is using an exponential decay scheme, it makes sense to align Sakura with this scheme.

It would be possible to consider variants where the Sakura reward is updated at every block, instead of relying on one large update every 4 years. However, achieving numerical stability with such schemes is a more complicated problem. Indeed, we are fundamentally applying a series of multiplications in order to properly distribute the satoshis. It matters little if the Sakura scheme is off by 1 satoshi every 4 years. If matters more if the scheme is off by 1 satoshi per block.

Countering the miner's incentive to self-censor transactions

Any scheme that offers the miner the possibility to claim UTXO entries as their own gives them an incentive to do just that. From this perspective, miners pursuing self-interest could effectively censor transactions that are associated with very old UTXO entries, e.g. 79 years old entries, as they would have more to gain from the censorship than from the transaction fee.

However, such a censorship is not only economically inefficient, but there are also further approaches that can be used to counter it altogether.

Censoring transactions by miners is economically inefficient because it's fundamentally a *prisoner's dilemma*: if the user who seeks to broadcast its transaction is insistent enough, iteratively reaching every single Bitcoin node of the network, then any miner can decide to betray its fellow miners and put the user's transaction in a block of his own making taking the fee for himself. All it takes is a single honest miner, which is a rather low requirement.

Then, in order to provably not censor transactions, a miner might be offering a paid service where transactions are first submitted in encrypted form to the miner, signed by the miner, and then decrypted by user and passed again to the miner to be included in a block of its own making. This mechanism would provide a way for the user to seek remedies in court from a miner who would fail to act on his end of the bargain. Some miners, skin-in-the-game miners, would be willing to offer such a service precisely because of the higher fees they could extract from users by making themselves vulnerable to an action in court.

Why 80 years

The value of 80 years is chosen because it approximately matches the human longevity at present time. Putting an expiration date on UTXO entries will force people who want to preserve their wealth to move their bitcoins, but no more than once in a lifetime. If Bitcoin is correctly designed, the transaction fees should remain low forever. Thus, the economic burden that the Sakura scheme puts on Bitcoin users is very low.

Once in a lifetime, typically from parent to children, a money transfer would occur. This money transfer would also ensure some degree of renewed security - as any private key redundantly stored might ultimately get compromised, given a few centuries of storage over a necessarily imperfect physical storage medium.

Why a 20 years half-life

Bitcoin started with block halving at 4 years. By 2088, if Bitcoin has succeeded at becoming the money of the world, then the monetary system will probably have been stabilized decades earlier. Re-introducing massive liquidity is creating the risk of disrupting this stability, which would go against what Bitcoin is trying to achieve in the first place.

The 4 year period was a proper duration suitable for the launch of Bitcoin. However, we are now presently contemplating the very long term economics of Bitcoin. Re-injecting 1M bitcoins into the economy in 4 years would not be overwhelming - about 1.25% of extra monetary mass per year, but doing the same in 20 years decrease this number to 0.25%; which feels sufficiently low to be near harmless to the economy at large.

References

Tesla is valued as high as Ford and GM — but that has nothing to do with what it's done so far, Business Insider, April 2017,
http://www.businessinsider.fr/us/tesla-value-vs-ford-gm-chart-2017-4

Organic Food & Beverage Market Size Worth $320.5 Billion By 2025, Grand View Research, April 2017,
https://www.grandviewresearch.com/press-release/global-organic-food-beverages-market