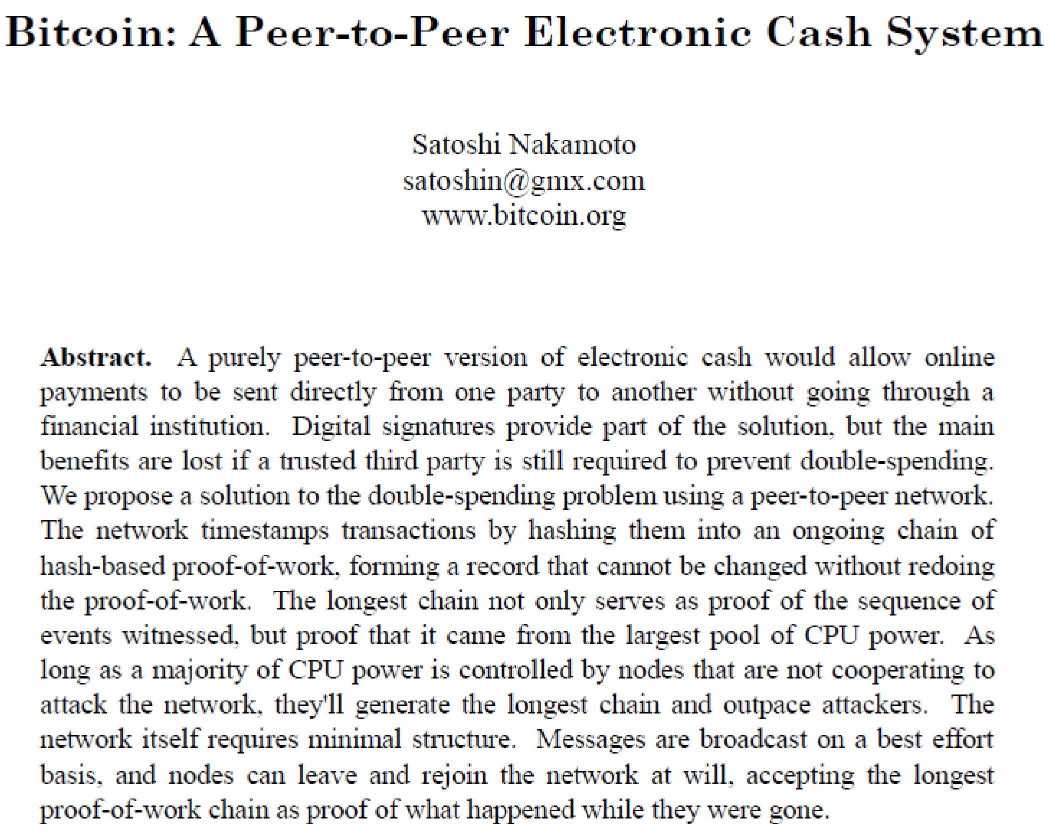
FUQINTRD 697: Innovation and Cryptoventures

51% PoW Attack

Campbell R. Harvey

Duke University and NBER

The warning



Bitcoin: A Peer-to-Peer Electronic Cash System

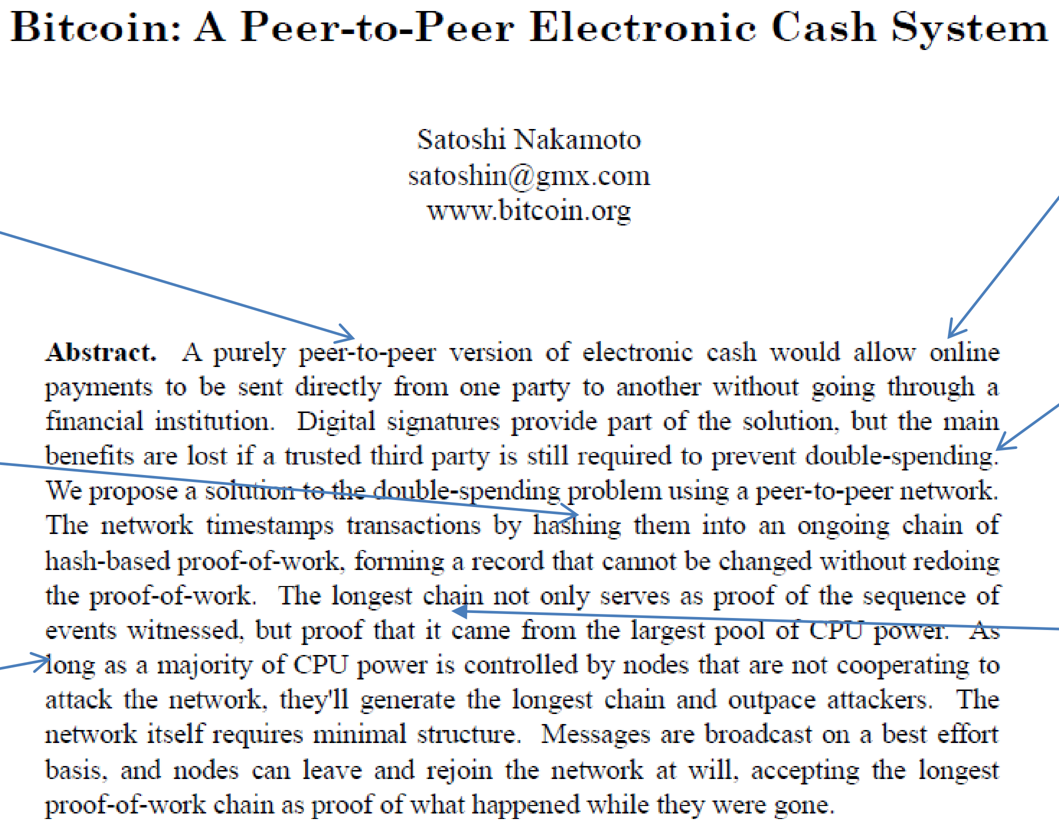
Satoshi Nakamoto
satoshin@gmx.com
www.bitcoin.org

Abstract. A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. 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 not only serves 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. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone.

https://bitcoin.org/bitcoin.pdf

Published on Internet November 2008

The warning



Bitcoin: A Peer-to-Peer Electronic Cash System

Satoshi Nakamoto
satoshin@gmx.com
www.bitcoin.org

Abstract. A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. 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 not only serves 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. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone.

P2P

Electronic payment system

No double spending

Secure via hash

Longest chain

Warning about 51% attack

https://bitcoin.org/bitcoin.pdf

Published November 2008

Setting

Attacker (A) amasses 51% of hashing power in PoW blockchain

- Attacker acquires something and pays 100 BTC to B
- This transaction is valid and is eventually confirmed by entering block 10
- Important that the attacker has collected the goods from B

Rewriting history

Attacker uses 51% of hashing power to rewrite history

- Suppose we are now on block 12
- Attacker redoes block 10 by changing the transaction: instead of paying 100BTC to B, attacker pays 100BTC to herself (at the time of block 10, this would be a valid transaction). “Coinbase” transaction (reward to miner) is also changed.
- Attacker also must redo block 11 because of the hash link is different. Attacker may keep all of the transactions in block 11 (change the coinbase transaction)
- Attacker is now working on their version of block 12

Rewriting history

There are two chains now

- There is the original chain 1-11 (we are working on 12)
- There is the attacker's chain 1-11 (blocks 1-9 are common with the original chain)
- How does the network know what one to use?
- That is, what triggers the replacement of blocks 10-12 (assuming the attacker finishes block 12 before the rest of the miners)?

Rewriting history

Cumulative proof of work best chain rule

- Any node that receives a different VALID chain of blocks (that don't break any protocol rules) will reorganize its perspective of the blockchain to use the new chain of blocks if and only if the cumulative proof of work securing that chain of blocks is greater than the cumulative proof of work securing the chain it currently believes to be the best.*

Thanks to bitcoin developer Jameson Lopp for these details

Achieving the hack

Bitcoin's difficulty only adjusts every 2,016 blocks

- The attacker has 51% of the mining power whereas other miners have only 49%
- This means that the attacker will be able to add blocks faster than rest of miners
- Eventually, attacker will have one extra block than the rest:
  - For example, attacker has a chain 1-20 whereas the rest have a chain 1-19

Achieving the hack

Longest chain

- Once you have the longest chain, you broadcast your newest block (say block 20) to the nodes
- Nodes will see it is valid and automatically request the missing blocks 10-19 to backfill them
- The nodes will recognize the attacker’s version of the blockchain because it is the higher cumulative proof of work (because it has one extra block)
- In this particular case (where difficulty is the same for a long stretch of blocks), the longest chain wins

Other issues

Empty blocks

- Attacker posts blocks that include no transactions in the memory pool of pending transactions
- The only transaction is the coinbase transaction to the attacker
- This means that the memory pool will get large and it will take a long time to confirm transactions
- It might be that new transactions start to come in at a faster rate than they can be put in blocks
- This would lead people to lose confidence in the system

Other issues

Spamming mempool

- Flood mempool with millions of 1 satoshi transactions
- This is less feasible today because these transactions would be ignored because they don’t include transactions fees

Other issues

Involuntary promotion of mempool spamming

- Coinbase, Blockchain, and Genesis have been causing significant congestions by not batching their transactions
- Coinbase.com is working on a solution (moving to SegWit)

Blockchain vs. exchanges

Hacking an exchange is completely different

- Hacks of CoinCheck, Bitfinex and Mt Gox do not involve blockchain hacking – they simply reflect the lack of security on these exchangers

Other ways to attack

There are many other ways to attack

- https://en.bitcoin.it/wiki/Irreversible_Transactions presents a list of attack vectors (that includes the “majority attack”)

Recommended Reading

There are many other ways to attack

- https://en.bitcoin.it/wiki/Majority_attack
- https://medium.com/@fhansmann/demystifying-the-51-consensus-attack-942252090b33