Hybrid architecture model to increase bootstrapping capability on Bitcoin

Richard Dennis
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
Richard.dennis@port.ac.uk

Gareth Owenson
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
Gareth.owenson@port.ac.uk

Abstract— **How to scale Bitcoin is still an open research question, while most of the research currently focuses on increasing the number of transaction Bitcoin can process, this paper takes a different view, and looks at the bootstrapping method. We demonstrate an effective and proven attack on the current DNS protocol which enables a low resourced attacker to partition new nodes joining the network. We then conduct analysis on how well the current DNS model can scale, before suggesting a hybrid P2P architecture model comparing this model with the current protocols, in terms of resistance to the DNS attack and scalability.**

Blockchain, scalability, cryptographic protocols, distrubted networks, peer-to-peer, Bittorent

I. INTRODUCTION

Bitcoin is the first worldwide, mass adopted cryptocurrency and digital payment system to be implemented and deployed without a requirement of a centralized repository system or administrator. It was invented by an unknown programmer, or a group of programmers, under the name Satoshi Nakamoto, published in a white paper in 2008, before being released as open-source software in 2009.

The key invention made by Nakamoto was the blockchain is a novel peer-to-peer approach which links a sequence of transactions or events together in a way that makes them immutable.

The blockchain is a public ledger of all transactions that have ever been completed since the first "genesis" block. Each transaction from the Bitcoin protocol is broadcast to all nodes in the network which are maintaining the blockchain.

A blockchain-node and a miner are two types of nodes on the network, which while can conducted on the same node, is usually separated. A blockchain-node can be classed as node which maintains and updates the blockchain, with valid blocks received from miners on the network.

Each blockchain-node confirms if each transaction is valid and can be added to a block. Each blockchain-node confirms every transaction made on the network, to do so, each blockchain-node will search through the blockchain they store and maintain to see if the user requesting the transaction has got enough funds to process the transaction, and this transaction has not previously been conducted. Only once this process has happened will each node compile a block (a group of transactions) and send this to the miners. There are not incentives to run a blockchain-node.

A miner participates in the process by which transactions are verified and added to the public ledger, known as the block chain, and also the means through which new bitcoin are released. The mining process involves compiling recent transactions into blocks and trying to solve a computationally difficult puzzle. The participant who first solves the puzzle gets to place the next block on the block chain and claim the rewards. The rewards, which incentivize mining, are both the transaction fees associated with the transactions compiled in the block as well as newly released bitcoin

Bitcoin is the most successful blockchain-based network; it has a market cap of over USD 8.5 billion and sees an average of 214,000 transactions being conducted on its network every day.

Blockchain-based networks have not properly addressed the issue of scalability; this causes the original decentralized nature of the blockchain to become increasingly centralized, as only the highest-resourced users are able participate in the network. This is because each node on the network is required to store the entire blockchain, which stores every transaction since its deployment and consequently low-resourced users; such as mobile users – are excluded from the network.

There has been several other peer-to-peer (P2P) and decentralized networks such as BitTorrent which have face similar scalability issues, overcame some of these issues with the use of a Hybrid architecture model, combining both the client-server model and P2P architecture.

This paper proposes a new approach in the way lower resourced nodes can be included in the network. It will examine how trusted "super nodes" can be utilized enabling a global view of the network, while providing monitoring facilities of all nodes on the network. We will conduct a thorough analysis on how the introduction of super nodes can aid in the scalability of the bootstrap process, by first demonstrating a unique attack against the currently implementation of the bootstrap protocol, and then analysing how the introduction of super nodes not only prevents such an attack from occurring, but also allows for a greater number of simultaneous nodes to bootstrap at the same time.

In addition to presenting how super nodes can be utilized, this paper also examines and conducts an in-depth analysis of how pooling of nodes can aid in the reduction of network traffic without introducing any additional security issues into the core data structure.

The proposed additions to the Bitcoin protocol are not solely applicable to Bitcoin, but will aid in the scalability of all blockchain based networks using a similar structure as Bitcoin.

The structure of this paper is as follows; first we examine related work conducted in the field of the Bitcoin, blockchain and also general peer-to-peer networks. Demonstrated next is a DNS attack against the current implementation in addition to calculating the maximum number of nodes currently able to bootstrap into the network at a single time. We then propose a solution to Bitcoin's DNS scalability problem – Authority Nodes – and present an in-depth evaluation against the current implementation of the bootstrap mechanism. We then propose future work for this area of research, before summarizing our findings and concluding the paper.

II. Related Work

The blockchain was first described in a self-published research paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System" written under the pseudonym of Satoshi Nakamoto. The blockchain is the underlying gossip protocol of all cryptocurrencies and is a novel peer-to-peer method of linking a sequence of transactions or events together in a way that makes them immutable [9]. McConaghy et al. accurately describe the main characteristics of the blockchain as decentralized control, immutability, and creation & movement of digital assets [8], and Pilkington credits the success of Bitcoin solely to the blockchain [10].

Drainville correctly describes how the blockchain is a collection of every transaction to have ever occurred on the Bitcoin network [2]. On creating a transaction, a user broadcasts this to all peers in the network. Kroll et al. expand on this by explaining how a select group of peers, called miners, collect broadcast transactions and attempt to gather them in a block that satisfies a cryptographic hash function [6]. The block must contain a cryptographic hash of the previous block; this is the meth-od used to cryptographically link every block in the blockchain to its previous block, all the way back to the first or "genesis" block. Producing a block is both computationally intensive and probabilistic. Given a proposed block, each miner has a fixed and independent probability of successfully producing a block which satisfies the hash function for each unit of computation time. Whilst it is difficult to produce a block, it is not difficult to verify a correct block.

Kroll et al. [6] explain that the mining process requires vast computing power as only a "brute force, trial and error" method can be used to calculate the SHA-256 hash. Every two weeks, the complexity of the challenge is adjusted to ensure that, on average, a block is mined every 10 minutes. The financial incentive of 25 bitcoins (USD 14,419.50 [1]) is offered to the first miner to successfully calculate the hash. Barber et al. [2] argue that it is this financial reward that ensures the majority of the miners on the network act honestly and obey the network protocol.

Sompolinsky and Zohar argue that only an attacker controlling more than 51% of the network hashing power would have the ability to change past transactions [12], and demonstrate that the cost of resources required to control 51% would outweigh the potential rewards. Dumas et al. [3] question the 51% vulnerability claim, which was originally presented in Nakamoto's whitepaper [9], suggesting that it is a wide-spread security claim, but no analysis has been conducted to prove or disprove this assumption.

Vulnerability to attacks is not the blockchain's only issue. Poon and Dryja summarise the scalability problem facing all blockchain-based networks as not being a single problem, but rather the combination of multiple issues that ultimately affect the possible scalability of the blockchain [11]. Poon and Dryja [11] reinforce their scalability argument by demonstrating how the maximum theoretical number of transactions per second that Bitcoin's blockchain is able to process is 7, whereas VISA can process 20,000. McConaghy et al. [8] agree with Poon and Dryja and demonstrate that the Bitcoin's blockchain is currently 50GB – having grown by 24GB in 2015 – and also prove that in order to achieve the transaction rate of VISA by only increasing the block size, the blockchain would need to grow by 3.9 GB/day or 1.42 TB/year.

Overall, the blockchain is the most important invention of the original Bitcoin whitepaper. While it has seen impressive growth and now handles an average of 239,138 transactions per day [1], it is not a faultless system. Having been shown to be vulnerable to attacks, such as the 51% attack, and faced with scalability issues which impact on the potential growth, there is plenty of room for further research to solve these issues.

Poon and Dryja [6] describe the Blockchain Scalability Problem as not being a single problem, but rather the combination of multiple issues that ultimately affect the possible scalability of the blockchain.

On average, VISA handles around 2,000 transactions per second (tps), with a recorded daily peak rate of 4,000 tps. It has a peak capacity of around 56,000 transactions per second [13]. By comparison, the maximum number of transactions per second that Bitcoin can currently theoretically achieve with the 1MB block size limit is 7 [4]. Poon and Dryja [6] describe how, whilst it is possible to achieve the tps VISA is capable of on Bitcoin, this would result in 8GB blocks, and a blockchain that would increase in size by over 400 terabytes a year.

Eyal et al. [8] doubt whether an increase in the block size can solve the scalability issue; their research paper demonstrates that, as an increased block size results in additional resources being required, an increased block size would mean fewer home computers would be able to participate in the network, and this would ultimately lead to centralization.

In his whitepaper, Nakamoto [5] states that the requirement to store all transaction history since the first transaction will ultimately result in the blockchain protocol failing to scale. Nakamoto foresaw a scalability issue during the design of Bitcoin and proposed a "pruning" mechanism. This mechanism allows a user to remove all spent transactions from their copy

of the blockchain. A spent transaction is a transaction that can no longer be used as an input for a new transaction. However, this method has been criticized for still requiring a user to download the entire blockchain before they can start pruning and still requiring the majority of the nodes on the network to process a complete and unpruned blockchain [5].

III. DNS POISON ATTACK

Currently on the Bitcoin network there are 8 separate DNS seed servers. These are publicly known servers, and are run by volunteers, who gain no reward for providing such a service. From bitcoin core 0.6, the use of DNS nodes replaced the classical IRC model as the method to start the boot strap process.

Over a period of 30 days, the DNS seed nodes was queried hourly, and the up time as well as number of node responses was logged. The results from this data collection can be seen in table 1.

Table 1: Bitcoin DNS queries

DNS seed node	Average online percentage	Average number of nodes received
bitseed.xf2.org	7.57%	18
dnsseed.bitcoin.dashjr.org	96.48%	20
dnsseed.bluematt.me	99.18%	24
seed.bitcoinstats.com	98.77%	23
seed.bitcoin.jonasschnelli.ch	93.77%	12
seed.bitcoin.sipa.be	83.24%	30
seed.bitnodes.io	98.37%	26
seed.btc.petertodd.org	60.68%	20
	79.76%	21.625

As the results from table 1 show, a node has only a 75% chance of querying a DNS node that is only, and receives on average 21 nodes in return.

The nodes received during the DNS query stage were also often not online, during the 30-day period it was observed that an average of 12 nodes per each query was no longer contactable or online.

Due to how the DNS seed mechanism is currently implemented, it is possible with a low amount of resources to conduct a DDOS attack against the 8 seed nodes. Since these nodes are not of high performance, such an attack could be conducted but a low resourced adversary. In addition to the unstable nature of some of the DNS seeds, this attack would easily prevent nodes joining the network.

For low resourced users, the wasting of resources querying nodes which is offline, is an overhead which could be reduced, as this not only wastes time, but also bandwidth which is some areas is very limited.

Since the DNS nodes are publicly known, it could be possible to “poison” the DNS records with malicious nodes instead of genuine nodes. If an adversary was able to poison the DNS records, but constantly announcing to DNS seed nodes they are alive and online, making them more likely to be sent when a DNS request is received, a user looking to join the network would, if the adversary had enough resources be querying all malicious nodes.

With this ability, an adversary could force a new node to mine an incorrect blockchain, effectively partitioning them from the real network. Since the target node would not learn about any additional nodes on the network such an attack would go undetected.

This attack would not only partition the user from the network, but would result in the user having the redownload the blockchain and restart the bootstrap process, wasting resources as there was not a verifiable way of knowing if the nodes was honest, and the blockchain being downloaded was the correct blockchain.

In addition to a partition attack described above, a denial of service (DOS) attack is possible. The current Bitcoin implementation states a response to a block request must be received within 2 seconds, or the connection is dropped. Currently there are 486834 blocks in the Bitcoin blockchain, assuming all 8 connections (the maximum number of outgoing connections) are to malicious nodes conducting this attack, it would take a user 121708.5 seconds to download the blockchain. For a lower resourced user with poor bandwidth, the ability to download a complete 1MB block within the 2 seconds may be impossible, and would delay and prevent a user downloading the chain.

It is clear from the above attacks the current DNS method cannot scale very well, as it would be easy for a low resourced adversary to attack, preventing nodes from joining the network. In addition, with a larger number of nodes joining the network, the currently nodes would be overwhelmed by the demand and would fail as well.

To ensure this attack was not just theoretical but could deployed in a real-world scenario, a simulated version of the bootstrap process following the Bitcoin documentation was created and deployed. Using various number of malicious nodes, with changing online probability, the lowest number of nodes required for the attack to succeed was 5. This is down to the fact a majority out of the 8 nodes queried during the bootstrap process must be compromised, as if not a node would be altered to the malicious node and would drop this from the connection.

However, the probability of a malicious attacker succeeding this attack with just 5 nodes is 8.3%. To increase the success of the attack, the attacker would need 20 nodes compromised and returned to the bootstrapping node when querying a DNS node for a success rate of 92%. The 20 compromised nodes do not need to be unique for each DNS node, which reduces the cost of attack significantly for an attacker.

Since there are no checks on if the nodes have an entire blockchain or any checks on how long the node has been participating on the network, a very low resourced node can be used for this attack.

Our simulations, demonstrated a raspberry pi zero would have enough computing power and bandwidth to produce such an attack, making an attacker with $100 invested in equipment able to partition the Bitcoin network with a 92% success rate. This is far lower than the 51% of network compromisation that is considered the amount of resources needed for an attack on Bitcoin.

IV. Our approach

A super node (called Authority nodes) in our model, can be classed as a non-mining blockchain node, which in addition to the basic functionality of such a node, also maintains a document containing a list of all nodes on the network.

The number of super nodes required on the network will be directly linked to the current network size. Using the Tor network as an example, a network of 6,000 nodes is able to be serviced by just 8 directory authority nodes.

In addition to maintaining a global view of the network, this node must also monitor all nodes (or a subset depending on network size).

To obtain and maintain the global view of the network, each node when joining the network announce itself to one (or more) of the authority nodes. It is not however a requirement for all nodes to be present on the consensus document, in fact it may be of benefit for a small subset of nodes not be displayed on the document.

By not being a part of the consensus document, a node will still be able to participate in the network, using out of band means for other nodes to find it, but will increase the anti-censorship resistance from a malicious government trying to block Bitcoin. Since by now being public, a censorship of Bitcoin at nation state level would be made simpler, although Bitcoin has never tried to be censorship resistance in its previous design models.

Each super node will continuously monitor all nodes (or a sub set) to ensure they are still online and able to be contactable, this is to ensure the consensus documents remains an accurate representation of the current state of the network. A new consensus document will be updated on an hourly basis.

In addition to querying if a node is still online, a super node will query a node in an attempt to judge their honesty. To achieve this, an authority node will query the node for a particular piece of data, such as a block they should process. An authority node receiving this piece of data will be able to compare it to data they process. The authority node will not have the ability to exclude any node from the network, however they would be able to assign flags to a node if they suspect it behaving malicious.

To prevent a node being malicious but honest when queried by the authority nodes, querying through an anonymity network such as Tor, or a VPN would be required.

In addition to just supplying a global view of the network, the consensus document would also contain check points of blocks contained within the blockchain. This would reduce the likelihood of a malicious chain from being formed and the network partitioned.

The increased in resourced required by the super node to conduct the additional measurement of nodes on the network, requires a more powerful server to act as a super node. There are no incentives to run a super node over a standard blockchain-node, however as can been seen on other peer-to-peer networks, such as Tor the willingness to contribute to the network may be a great enough incentive. However, it could be possible for a change in the coin base award as each block is mined, a small token value could be sent to each super node in order to aid in the running costs of the nodes.

V. Limitations, analysis and solutions

The authority nodes are designed to aid in the bootstrap model. It has been simulated, with the current network size, it would take an average of 334 nodes to be queried, with 1536 packets sent and received to obtain a list of 90% of the nodes on the network, however using the authority node model, a single node would need to be queried with a packet size of 2,171kb based on the current network size.

To conduct a Sybil attack against the previous bootstrap model, an attacker with 30 compromised nodes, positioned correctly in the DNS seed nodes (usually the most recent seen during past experiments), a success rate of 80> % was observed. This is a low amount of resources for such a high probability of success from an attack.

With the newly proposed authority nodes, an attacker would no longer be able to target a bootstrapping node to only connect to blockchain-nodes controlled by the attacker, as instead the bootstrapping node, would have a global view, and would choose nodes based the consensus document.

It is possible for an authority node to become compromised, or run by a malicious actor, however it would be recommended to obtain a consensus document from more than a single authority node, as although this would increase the resources required to bootstrap, the reduction in possible attack from a compromised authority node could be seen as worth the increased bandwidth used.

In addition to having a global view of the network to aid in bootstrapping process, the consensus document, could allow low resources users to easily find other nodes to pool together data.

In a SPV (Simplified Payment Verification) client, when a user wants to confirm a transaction, the client must have the Merkle root in the block header along with a Merkle branch can prove to that the transaction in question is embedded in a block in the block chain. This does not guarantee validity of the transactions that are embedded.

A full node can simply lie by omission, leading an SPV client to believe a transaction has not occurred. The implementation of authority nodes, would allow for these authority nodes to query full nodes, while behaving as a SPV client. This would

allow for greater trust to be placed in the honesty of the network.

The current bootstrap model cannot scale. A python simulator of the current bootstrap process showed 896 new nodes can bootstrap at the same time. Assuming the blockchain is 140GB in size, and the nodes used to download the blockchain are used exclusively during the download.

With the proposed model of using, the maximum number of simultaneous nodes able to bootstrap is 75176. This is a massive increase, increasing the capacity by a factor of 84, and shows the network is able to be more scalable than current implementation allows.

With the observed network churn of 85 nodes per day on the current Bitcoin network, the probability of a node being selected from the consensus and it being offline is 0.036%, the currently DNS model it was observed to be 37%

A low resourced node when bootstrapped into the network, does not have to download the entire blockchain, the Merkle root of all block headers would be sufficient to act as a SPV client.

VI. Conclusion

In this paper, we demonstrated a unique attack on the current DNS protocol which has never been described previously in literature.

From our analysis of this attack it was clear the current DNS protocol is not only vulnerable to attack from a low resourced adversary, but in addition does not scale, with a large number of nodes wishing to join the network.

We proposed a hybrid server model, to aid in the bootstrap process by allowing a global view of the network to obtained from a single document. In addition to this we theorized how this authority node could play further roles in the network and reducing the likelihood of malicious nodes being able to be undetected.

Overall this paper has set out a blueprint model for how a hybrid architecture model can be implemented on the blockchain and allows the foundations for much more in depth research into the impact such a model will have on the general use of the network in addition to the security.

VII. Future work

This paper has intended to be the foundations for further work to be conducted on how hybrid architecture can play a role in Bitcoin, and in fact all blockchain based networks, as a way to greater decentralized nodes, by utilizing the abilities of a semi trusted central node in order to create node pools, reducing of malicious nodes on the network, and allow for a fully auditable record of nodes on the network.

Further research is being untaken, which aims to advance this paper into greater detail and how the authority nodes would be implemented and a full specification of there uses.

References

[1] CoinDesk, http://www.coindesk.com/data/bitcoin-daily-transactions/

[2] Drainville, D.: An Analysis of the Bitcoin Electronic Cash System, https://uwaterloo.ca/combinatorics-and-optimization/sites/ca.combinatorics-and-optimization/files/uploads/files/drainville_danielle.pdf

[3] Dumas, J., Sygnet, P., Xuereb, V.: Bitcoin a Peer-to-Peer Payment Solution, https://www.semanticscholar.org/paper/Bitcoin-a-Peer-to-peer-Payment-Solution-security-Dumas-Joseph/7a1e2a9e0fa3b9e64d09c0587ce302dfe7a32ee3/pdf

[4] Hashing It. (2014). 7 Transactions Per Second? Really?[Online]. Available: http://hashingit.com/analysis/33-7-transactions-per-second

[5] I. Eyal, A. Gencer, E. Sirer and R. van Renesse. (2015). Bitcoin-NG: A Scalable Blockchain Protocol[Online]. Available: http://arxiv.org/abs/1510.02037

[6] J. Poon and T. Dryja. (2016). The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments [Online]. Available: https://lightning.network/lightning-network-paper.pdf

[7] Kroll, J., Davey, I., Felten, E.: The Economics of Bitcoin Mining or Bitcoin in the Presence of Adversaries, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.5595&rep=rep1&type=pdf

[8] McConaghy, T., Marques, R., Müller, A., De Jonghe, D., McConaghy, T., McMullen, G., Henderson, R., Bellemare, S., Granzotto, A.: Bigchain DB: A Scalable Blockchain Database, https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf

[9] Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System, https://bitcoin.org/bitcoin.pdf

[10] Pilkington, M.: Blockchain Technology: Principles and Applications, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2662660

[11] Poon, J., Dryja, T.: The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments, https://lightning.network/lightning-network-paper.pdf

[12] Sompolinsky, Y., Zohar, A.: Secure High-Rate Transaction Processing in Bitcoin, http://fc15.ifca.ai/preproceedings/paper_30.pdf

[13] Visa. (2015). Visa Inc. at a Glance [Online]. Available: https://usa.visa.com/dam/VCOM/download/corporate/media/visa-fact-sheet-Jun2015.pdf