

Computer Science



Commentary to support marking

Subject: Computer Science

Paper component: Extended Essay

Language: ENGLISH

Exam session: MAY 2018

Essay: 31A

Criterion	Mark Awarded	Out of	Justification
A	6	6	Both the research question and the topic were clear and appropriate and relevant to current financial dealings. Both context and methodology were present although more detail on double-spending could have been given. A range of appropriate sources were used and the focus maintained throughout the essay.
B	6	6	The student shows a very good understanding of a particularly complex computer science topic that is far beyond the level of an IB course. Terminology is used well and is generally explained in comparatively simple terms
C	11	12	The student has generally made a successful attempt at producing a reasoned argument of a complex topic. He/she has critically analysed the results and the limitations underlying the conclusions, which are to a degree speculative and based in the assumptions of the student's own method of analysis. Critical thinking is deemed to be in the excellent band.
D	3	4	The overall presentation is good with most formal requirements being met, with a clear use of diagrams and a consistent referencing/citation style. There were a couple of issues: the lack of a title in the title page and, more importantly, various redaction errors which could have been rectified with a final proof-reading.
E	5	6	There was good analysis of both research topic and the process itself particularly on the first two reflections. In spite of a very brief

			comment from the supervisor, overall the reflections showed a sense of journey having taken place.
Total:	31	34	

Candidate Marks Report

Series : M18 2018

This candidate's script has been assessed using On-Screen Marking. The marks are therefore not shown on the script itself, but are summarised in the table below.

Centre No :	Assessment Code :	COMPUTER SC. EE EXTENDED ESSAY in ENGLISH
Candidate No :	Component Code :	EE(ENG)TZ0
Candidate Name :		

In the table below 'Total Mark' records the mark scored by this candidate.
'Max Mark' records the Maximum Mark available for the question.

Examiner:	
Paper:	M18comscEEEE0XXXX
Paper Total:	31 / 34
Question	Total / Max Mark Mark
Criterion A	6 / 6
Criterion B	6 / 6
Criterion C	11 / 12
Criterion D	3 / 4
Criterion E	5 / 6

Coursework confirmation

Yes

Hours supervisor spent with candidate

3

Declaration

Yes

Overall:
The student has demonstrated a very good understanding of a difficult topic through a clearly explained exploration of possible Bitcoin attacks. Analysis is sound as are the final conclusions.

Computer Science: To what extent does the 'balance' vulnerability supersede conventional double spend attacks as the most effective threat to Nakamoto's consensus protocol?

D: incomplete title page

Contents

Figures	4
Introduction.....	5
Research Question	5
Bitcoin	5
Methodology	6
Network Structure.....	7
Representation of Network	7
Distributed and Decentralized Networks	8
Pools.....	9
Blockchains.....	9
Merkle Trees.....	11
Block Propagation.....	11
Consensus Algorithms	12
Transaction Confirmation.....	13
Attacks	14
Double Spending on a Blockchain	14
Block-Obliviousness	15
Conventional Double Spend	16
Proof 1.....	16
Balance attack	19
.....	2

D: formatting good

Introduction	19
The Attack	19
Proof 2.....	21
Conclusion.....	23
References	25
Appendices.....	27
Proof 2.....	27

Figures

Figure 1 - (Baran, 1964)	8
Figure 2 - (Bitcoin.org, n.d.)	11
Figure 3 - (Bitcoin.org, n.d.)	12
Figure 4 - (Bitcoin.org, n.d.)	13

Introduction

Research Question

To what extent does the 'balance' vulnerability supersede conventional double-spend attacks as the most effective threat to Nakamoto's consensus protocol?

Bitcoin

Within recent years crypto-currencies have grown from being used by 'geeks', 'criminals' and social libertarians to becoming close to mainstream. However, despite having a completely modern structure, they are still vulnerable to fraudulent activity.

Bitcoin is the largest crypto-currency with a market capitalisation far outstripping any competitor (Coinmarketcap.com, 2017). Satoshi Nakamoto, the pseudonym for the anonymous creator of the first Bitcoin, wrote his consensus protocol to provide a backing for his vision of a decentralized currency. However, Nakamoto's consensus protocol has far larger applications than just in currencies as projects examining its use in housing registries and cross border transactions are currently underway.

However, Bitcoin is not truly decentralized in the way that Nakamoto envisaged. In his seminal paper, he stated that:


"The [Bitcoin] network is robust in its unstructured simplicity. Nodes work all at once with little coordination."

Due to market forces, networks are structured with larger consortiums benefiting from lower marginal costs. As data gathered by BLOCKCHAIN S.A, a leading player in the Bitcoin network, has shown a growing proportion of network power controlled by a relatively small number of 'pools'. (Blockchain.info, 2017) These changes have made new types of attack possible which have a considerably higher chance of success than previous attacks as this essay will show.

Methodology

The essay will seek to show the threat which the Balance attack exposes the Bitcoin network as compared to conventional double-spend attacks. The Balance attack, a theoretical method to de-fraud the Ethereum chain, was created by researchers from the University of Sydney and attempts to allow malicious actors with much less than 50% of network power to double-spend with a high probability. (Natoli and Gramoli, 2016) This takes advantage of the problems with consensus algorithms distributing information throughout the system and the ability for malicious actors to easily interrupt communications between nodes in the mining network using commonplace techniques.

To do this, I will consider two thresholds for success, both a double-spend occurring after 3 blocks have passed and a double spend occurring after 6 blocks have passed for reasons that will be shown later. In my scenario, the attacker will be able to control 5% of network hash power. I have used my understanding of how the network operates to calculate the probability that the attacker meets these thresholds. This has been done through a model I created which provides a fair approximation of this probability.



A: content, purpose, worthiness and methodology all addressed

The essay does not seek to show how the attacker might isolate nodes but only seeks to show that if it is possible, an attack is far more likely to occur as a result of the Balance attack as compared to the conventional double-spend attack.

Network Structure

It is impossible to pin down an “authoritative” Bitcoin specification. (Courtois and Bahack, 2014) The protocol is constantly changing and adapting as the result of changes submitted by countless programmers, consortiums and working groups. However, the core principles of the network in the Bitcoin network can be described abstractly.

Representation of Network

The Bitcoin network can be represented as a directed graph G where:

$$G = (V, E)$$

(Natoli and Gramoli, 2016)

In the network, the vertices V are logical nodes. The difference between physical and logical here is distinct as a node could be one of many physical sections of hardware linked together in some form of network. All the nodes are part of the Bitcoin network.

The edges E are fixed communication links.

The Bitcoin network is categorised as distributed and in theory should resemble a mesh network. Full Bitcoin Nodes have degree of 8, meaning that they have 8 logical connections to the outside network. This, in theory, should make them less vulnerable to being isolated from the network as the Balance attack attempts to do.

Distributed and Decentralized Networks

There is limited agreement on the exact definitions of the different types of network as they were never explicitly defined in Baran's original paper (Baran, 1964). The terms are frequently misused by enthusiasts. Furthermore, given that only limited data exists on the degree distribution of the network, making judgements about the exact characteristics are problematic.

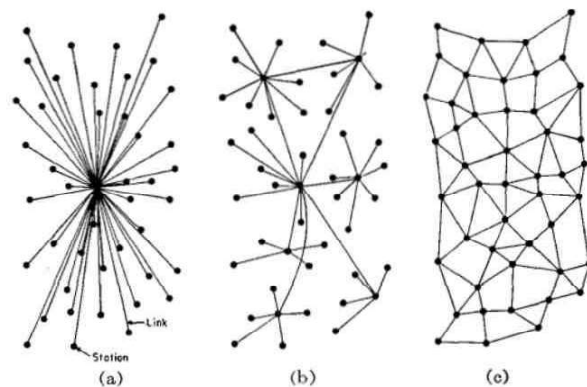


Fig. 1—(a) Centralized. (b) Decentralized. (c) Distributed networks.

Figure 1 - (Baran, 1964)

As Figure 1 shows, the definition of a centralized network is clear; it is one where a single node has individual connections to the rest of the network.

The clearest definitions which match Baran's paper are below:

Decentralised system: system where components operate on local information to accomplish goals, rather than the result of a central ordering influence

Distributed systems: system in which computation is distributed across components, which communicate and coordinate their actions by passing messages. The components interact with each other to achieve a common goal.

(Question on the terms 'distributed' and 'decentralised', 2016)

Bitcoin is both distributed and decentralised. However, it does not resemble the decentralized system (Figure 1 (c) - (Baran, 1964) as Nakamoto hoped. Not all nodes will bother to connect to a full 6 other nodes and in mining 'pools' a single node will act as a front for a far larger amount of computational power. This structuring has the potential to allow the network to resemble a decentralized network (Figure 1 (b) - (Baran, 1964) which is a much easier target as Baran showed.

Pools

As the likelihood of correctly hashing the chain is entirely random there is an incentive to pool together computational resources to provide a stable income for participants. Therefore, a 'pool' is in many ways like a gambling syndicate. The members contribute resources, in this case raw hash power and are rewarded with a share of the proceeds which is proportional to it. (Brezo and Bringas, 2012)

In contrast to the spirit of Nakamoto's paper, pools add some centralization because the administrator of the pool is the only one which connects to the network in some cases. This leads to single vertices which represent an amount of network power which is disproportionate to the number of connections they have to the network.

Blockchains

The basis for all current conventional crypto-currencies is some form of blockchain. The blockchain is frequently described as a distributed ledger system. (Nakamoto, 2008)

The blockchain can be represented mathematically where:

$$\iota = (B, P)$$

(Natoli and Gramoli, 2016)

The blockchain, ι is a directed acyclic graph. This means that the edges link the vertices so that the graph can be transversed only one way. The blockchain is also finite with an identifiable size.

The blocks B are data structures with multiple fields.

The pointers P are links between the blocks

The genesis block g is the first block in the blockchain and links back to nothing.

Following this model, the block chain is can be represented as,

$(b_1, g), (b_2, b_1) \dots, (b_m, b_{m-1})$ However, for clarity I will use the following notation $g \rightarrow b_1 \rightarrow b_2 \rightarrow \dots \rightarrow b_m$

The ledgers, which are records of transactions that are in groups known as blocks, are cryptographically linked using a hashing function. A hash is a one-way function where data is converted into a string of a fixed length. Every time data is hashed it produces the same output of the same length. However, unlike some other cryptographic methods, a single change will create a completely different hash making it impossible to work towards a solution.

The previous block is referenced in the current block forming an unbroken chain (Dabbs, 2016). This is further shown in Figure 2 below. This is designed to make the record of transactions which is an objective version of the truth designed to prevent fraud. The blockchain is thus a cryptographically backed linked-list where the previous block is referenced using the hash of the previous block's header. There is a continuous link between the blocks back to the genesis block. All the transactions are hashed into a Merkle Tree.

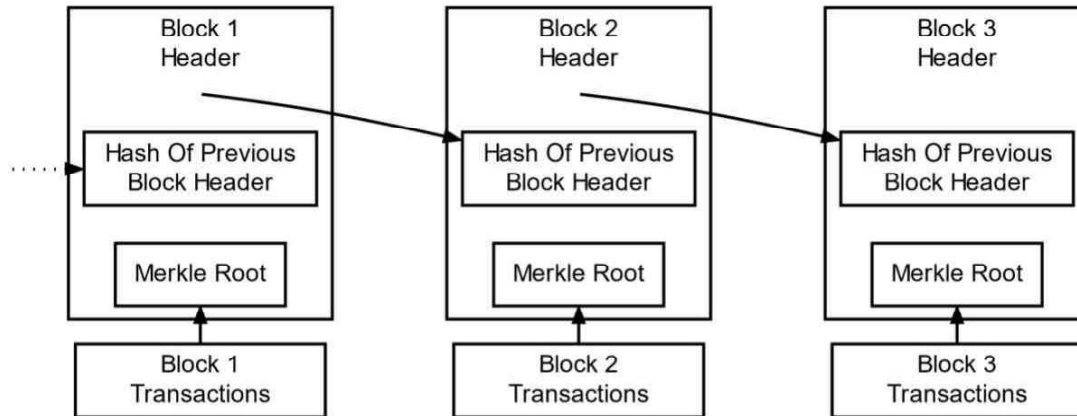


Figure 2 - (Bitcoin.org, n.d.)

Merkle Trees

In the Bitcoin protocol a Merkle tree is used to verify that all the transactions in the block have not been tampered with. Transactions are hashed and paired and the result is hashed together. This process is repeated recursively with the hash building a binary tree where the root node is the cryptographic result of all the transactions. (Chumbley and Moore, n.d.)

Block Propagation

As blockchain based systems are distributed, it is possible for multiple nodes to solve the cryptographic puzzle simultaneously leading to multiple nodes to attempt to propagate their own version of the chain. This is described as a forked chain or a tree (Natoli and Gramoli, 2016) and presents the fundamental problem with this distributed model.

As soon as a miner finds a solution to the hash which satisfies the criteria set by the network it will then broadcast this block outwards onto the network. Miners who are 'honest' will start trying to find the next block in the chain with the hash of the previous block's header. If two miners find solutions then the other miners will work on finding

a block based on the hash of the previous headers of the first complete block they receive.

The consensus algorithm will attempt to quickly stop the chain branching to prevent double spending.

In Nakamoto's protocol the cryptographic difficulty, d , is variable so each 'solution' to the crypto-graphic problem will take approximately 10 minutes. This should allow the block to be propagated throughout the entire network before the next block it created if no other solution has been found to the problem.

Consensus Algorithms

A consensus algorithm¹ underpins most blockchain based systems such as cryptocurrencies. They have two functions:

1. It should ensure that the next block in the blockchain is propagated after it is accepted.
2. It should stop the chain from being influenced by malicious actors.

(CoinDesk, (n.d.))

Although not explicitly mentioned a consensus algorithm will ensure that there is only a single version of the chain. This is fundamental to the ability to safely use a blockchain as it would be impossible otherwise for transactions to be carried out.

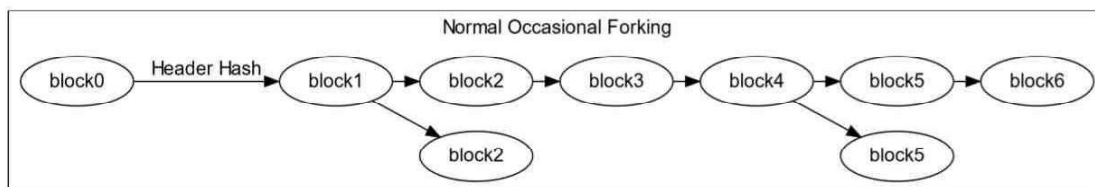


Figure 3 - (Bitcoin.org, n.d.)

¹ NB. For this essay, when consensus algorithms are referred to they are all proof of work.

In Nakamoto's consensus protocol, the Bitcoin Network is forced to accept the longest version of the chain. As figure 3 shows, forking for one block is expected occasionally and the next block mined easily decides the which chain is mined by the network.

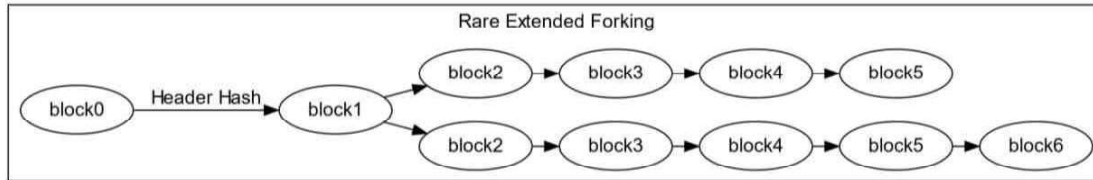


Figure 4 - (Bitcoin.org, n.d.)

In some rare situations, an extended fork may occur as shown in Figure 4 where the chain will split for a period of time, however as long as one fork has a larger share of computational power then it will eventually merge as one chain overtakes the other. According to data from Blockchain.info there has never been a fork of greater than four blocks apart from the 'Value Overflow' bug in August 2010. (Blockchain.info, 2017)

Transaction Confirmation

As there is no central authority to enforce or even give standards which must be followed, there is no clear number of blocks which must be confirmed before the transaction takes place. It is best practice within the Bitcoin Network to wait for up to 6 transactions before the transaction is counted as confirmed to limit the possibility of double-spend attacks occurring. (Bonneau, 2015) However, this practice is not ubiquitous within the network and is only considered best practice. Blockchain.info implies that transactions are confirmed after 3 blocks are in the chain. The longer the chain after the block the lower the chance of a transaction being reversed as it becomes exponentially more difficult to edit the chain. This shows the reasoning behind my dual threshold test used in this essay's analysis.

Attacks

Conventional double-spend attacks² attempt to exploit the nature of a consensus protocol by disrupting the integrity of the block chain. (Barber et al., 2014) Most attacks on the blockchain which do not directly target the hosts in the peer to peer network involve history revision attacks.

Double Spending on a Blockchain

To reverse a past transaction in order to double-spend a Bitcoin, the malicious actor must catch up and overtake the chain which is being created by honest miners.

This attack involves:

1. The victim, a merchant who will send something of value, usually another crypto-currency or electronic money which can be quickly transferred.
2. The attacker, probably some form of hacker, will try to get something of value from the merchant without having to pay for it. For any significant chance of success, they will need control a large amount of hashing power.

The attack happens in several stages:

1. A transaction $tx1$ is carried out and correctly signed by the necessary cryptographic keys. This is all honestly done, and the attacker and the victim do not know each other's private keys nor does the victim know that he is about to be defrauded.
2. $tx1$ is inserted onto block b_v and its hash added to the Merkle tree.

² Conventional double spend attacks are sometimes referred to as a history revision attack by some academic writing but as the Balance attack also involves history revision I have chosen to use this phrasing.

3. The attacker is secretly holding his block b_a which does not have a record of him paying the vendor for the goods. He proceeds to 'race' the network waiting for the goods from the vendor to be transferred.
4. After x blocks the transaction occurs and he gets his goods from the merchant. At this time, the chain is $g \rightarrow \dots \rightarrow b_v \rightarrow b_{v+1} \rightarrow \dots \rightarrow b_{v+x}$
5. He then broadcasts his chain out to the other nodes within the network. $g \rightarrow \dots \rightarrow b_a \rightarrow b_{a+1} \rightarrow \dots \rightarrow b_{a+x+1}$ which as it is 1 longer should be accepted by the network

This results in the blockchain being $g \rightarrow \dots \rightarrow b_a \rightarrow b_{a+1} \rightarrow \dots \rightarrow b_{a+x+1} \rightarrow \dots$ Where tx_1 is not present on the blockchain. This leads to the attacker having not spent any Bitcoins while receiving the goods which the vendor thought that he paid for. The attacker is now free to spend the Bitcoins again.

The system in this case has been rendered block-oblivious



Block-Obliviousness

Natoli and Gramoli provide a definition for a successful double spend attack.

A blockchain system is considered block oblivious if an attacker can:

1. Make the recipient of a transaction tx observe that tx is committed and
2. Later remove the transaction tx with a probability of success of $1 - \epsilon$ where ϵ is a small positive constant.

(Natoli and Gramoli, 2016)

Conventional Double Spend

The conventional double spend attack is as old as the Bitcoin Network but is difficult to complete successfully. Nakamoto showed that this is with synonymous with the 'gambler's ruin' problem and that the probability of success drops exponentially as the time since the block was propagated increases. (Nakamoto, 2008)

If the network power belonging to the honest miners is greater than the network power belonging to the malicious miners then the probability of this type of attack occurring can be shown to be mathematically negligible.

Proof 1

To show the probability of success of the attack, I have used Nakamoto's method and applied it to my conditions set out in my methodology.

Keeping with Nakamoto's notation we can state that:

p = probability an honest node finds the next block

q = probability the attacker finds the next block

q_z = probability the attacker will ever catch up from z blocks behind

It is assumed that $z = 6$ or $z = 3$ as per the methodology.

In this p and q represent the network power exerted by the malicious and honest nodes respectively. If q controlled 5% of the network power the probability of hashing the block first is 0.05. This is because the chance of solving the cryptographic 'puzzle'

is entirely random and the only factors effecting the chance of a successful hash is the hashing power.

The probability that the attacker 'wins' this binomial random walk can be expressed with the following probability density function.

$$q_z = \begin{cases} 1 & \text{if } p \leq q \\ (q/p)^z & \text{if } p > q \end{cases}$$

For the attack to be successful the attacker will need to control the chain for 3 or 6 blocks. The longer the chain the lower the chance of a successful attack.

$$P(x) = q_z \rightarrow 0 \text{ as } z \rightarrow \infty$$

As the probability of is small we can approximate the likelihood of this occurring using the Poisson distribution and the Poisson limit theorem. ³

$$\lim_{n \rightarrow \infty} P(q_z) = \frac{e^{-\lambda} \lambda^z}{z!}$$

Where

$$\lambda = z \frac{p}{q}$$

We can therefore multiply $P(x)$, the probability density function of q_z by the Poisson distribution to calculate results.

³ For the full proof see <http://www.oxfordmathcenter.com/drupal7/node/297> Accessed June 2017.

$$\sum_{z=0}^{\infty} \frac{e^{-\lambda} \lambda^z}{z!} \cdot P(x)$$

This can be rearranged to prevent summing the infinite end to the distribution.

$$1 - \sum_{z=0}^z \frac{e^{-\lambda} \lambda^z}{z!} \cdot (1 - P(x))$$

Therefore, I calculated that:

For $z = 3$

The probability q_z is 0.000125

For $z = 6$

The probability q_z is 0.15625e-8 using exponent notation

Both these probabilities have shown that conventional attacks simply are not effective with a small amount of hash power.

This shows why this double spend is considered the '51% attack'; it is not worth attempting to attack the network unless you control more than 50% of the network power. The network is inherently resistant to malicious actors if a full 6 transactions elapse before goods are moved out of escrow. The attack is also statistically very unlikely with 3 blocks used for confirmation with 5% network power.

Balance attack

Introduction

The Balance attack addresses the problem encountered by the conventional attack that proof of work blockchains require a large amount of hash power to have a high probability of editing the blockchain.

The Balance attack involves delaying communication within the network to allow a double to spend to occur. This relies on the fact, as noted previously, that network power is not evenly distributed within the network among many independent parties but a small number of pools of computational power. If these pools can be identified and be made to be block oblivious then the hash power which they control is in effect working for the attacker.

Application to the Nakamoto's Protocol

The Balance attack was created to attack the GHOST⁴ protocol which forms consensus on the Ethereum network. It is still applicable to the Bitcoin network as instead of adding uncle blocks to weight the sub-tree, the attacker instead mines on the top of the sub-group which does not contain his transactions.

The Attack

For simplicity, the number of mining sub-groups, k , has been fixed at 2 in my working.

The transaction subgroup, G_1 contains an approximate 45% of network power and will be isolated from the rest of the network by the attacker. This is made easier

⁴ GHOST is the abbreviation for Greedy Heaviest Observed Sub-Tree, the protocol used to provide consensus in Ethereum.

by the fact that some pools, while having many physical nodes, will only have 6 logical connections to the network. This subgroup need not be a single pool but could be two or more pools which would remain connected but still isolated from the rest of the network.

The mining subgroup, G_2 has another proportion of network power which ideally will be balanced in terms of hash power with G_1 .

The attacker A controls 5% of network power.

The rest of the network will be isolated from both sub-groups and for the simplicity of my essay will be considered to be de-facto nonexistent.

Execution of the Attack

1. The attacker using some form of malware creates two subgroups of roughly equal hash power.
2. A transaction $tx1$ is issued in G_1 sending several Bitcoins to a merchant in return for some goods.
3. The communication must be delayed by τ seconds so that the chain created by G_1 becomes long enough for the merchant to assume that the transaction is confirmed and the product is delivered.
4. Communication is restored.
5. The chain created by G_2 which does not contain the transaction should be longer and will replace G_1 chain meaning $tx1$ never took place.

The protocol has been manipulated and can be deemed block-oblivious.

For the attack to work the chain from G_2 will have to be greater than the chain from G_1 so that Nakamoto's protocol chooses it as the valid chain once connection within the network has been restored.

In the 'race' between the two chains, the attacker needs to be sufficiently certain that the transaction chain made by the network subgroup G_1 is shorter than G_2 .

Proof 2

To show the probability of success I have used my own method based on modelling the propagation of Bitcoins using the Poisson distribution

To show that the attack is intuitively more effective than the conventional double spend attack let us consider a network running Nakamoto's protocol where the subgroups remain the same as above. As we know, with the same network power, G_1 and G_2 will perform the same number of Bernoulli trials to find a solution to the crypto-puzzle.

If two random variables X_1 and X_2 represent the sum of the successes from, G_1 and G_2 respectively.

For the attack to fail, the chain G_2 , represented by X_2 will have to be longer than G_1 by an amount greater than the number of blocks A which the attacker can mine.

$$\text{mod}(X_1 - X_2) - A < 0$$

This is not an exact equation but is in my opinion a fair approximation which I have created of the probability the attack fails.

The probability of finding a block is Poissonly distributed (Onies, Olayinka and Daniele, 2017) with an expected value of 10 minutes in the Bitcoin network. Therefore, in an hour $X \sim Po(6)$. However, given that only half the hash power available $X \sim Po(3)$ for G_1 and G_2 .

Given that the attacker gives a time delay t , of enough time so that it is expected that the required number of blocks are mined, he will mine one block with a certain probability.

Therefore, using a Graphical Display Calculator (GDC) I have worked out the Poisson values.

For where $z = 3$

0.259181779318282

For where $z = 6$

0.451188363905973

G_1 and G_2 are expected to mine for where $z = 3$

Number of Blocks	Probability
1	0.149361205103592
2	0.224041807655388
3	0.224041807655388
4	0.168031355741541
5	0.100818813444924

Therefore, after the possibilities are cross referenced⁵ the chance is 0.100136273 of a double-spend occurring.

⁵ For full method see appendix

G1 and G2 are expected to mine for where $z = 6$

Number of Blocks	Probability
4	0.133852617539983
5	0.160623141047988
6	0.160623141047988
7	0.137676978041126
8	0.103257733530844

Therefore, after the possibilities are cross referenced⁶ the chance is 0.077919365 of a double-spend occurring.

In neither case is the network deemed block-oblivious with the statistical methods I have used. It is interesting to note that the probability of a successful attack where the time delay t is statistically large rises to over 40%. Despite this, with a 3-block verification time a 10% probability of de-frauding the network while holding only 5% of network power is something that cyber-criminals may take seriously given the amount of money flowing through the network

Conclusion

To what extent does the 'balance' vulnerability supersede conventional double-spend attacks as the most effective threat to Nakamoto's consensus protocol?

The attacks cannot be cleanly compared as they do not have common variables which directly affect the success or failure of the attacks. If hash power is kept

⁶ For full method see appendix

constant then as the estimates show the Balance vulnerability has a far higher chance of success than conventional double-spend attacks when comparing the results of proof 1 and proof 2 for both thresholds.

As data taken from BLOCKCHAIN S.A shows, the hash power within the network is highly concentrated which in theory could lead to a Balance attack being executed with a higher probability as compared to other methods of attack. Further concentration amongst mining groups along with slackening of security measures in attempt to speed up confirmation could be a toxic combination.

The threat posed by smart attackers who have knowledge of how the network works along with the known existence of large bot-nets with huge hash power leaves blockchains in a position which is far from secure. With the continuing faults being found, I agree with Natoli and Gramoli's analysis that proof of work consensus algorithms are inappropriate for use in secure systems.

References

- Baran, P. (1964). Introduction to distributed communication networks. On distributed communications. [online] RAND Corporation. Available at: https://www.rand.org/content/dam/rand/pubs/research_memoranda/2006/RM3420.pdf [Accessed 16 Aug. 2017].
- Barber, S., Boyen, X., Shi, E. and Uzun, E. (2014). "Bitter to Better - How to Make Bitcoin a Better Currency.." Financial Cryptography, [online] 7397(399-414). Available at: <http://elaineshi.com/docs/bitcoin.pdf> [Accessed 13 Jul. 2017].
- Bitcoin.org (n.d.). Blockchain Forks. [image] Available at: <https://bitcoin.org/en/developer-guide#proof-of-work> [Accessed 9 Aug. 2017].
- Bitcoin.org (n.d.). Simplified Bitcoin Block Chain. [image] Available at: <https://bitcoin.org/en/developer-guide#block-chain> [Accessed 9 Aug. 2017].
- Blockchain.info. (2017). Hashrate Distribution. [online] Available at: <https://blockchain.info/pools?timespan=4days> [Accessed 20 Jul. 2017].
- Blockchain.info. (2017). Orphaned Blocks - Blockchain.info. [online] Available at: <https://blockchain.info/orphaned-blocks> [Accessed 11 Aug. 2017].
- Bonadonna, E. (2013). Bitcoin and the double Spending Problem. [Blog] Available at: <https://blogs.cornell.edu/info4220/2013/03/29/bitcoin-and-the-double-spending-problem/> [Accessed 23 Feb. 2017].
- Bonneau, J. (2015). How long does it take for a Bitcoin transaction to be confirmed? | Coin Center. [online] Coin Center. Available at: <https://coincenter.org/entry/how-long-does-it-take-for-a-bitcoin-transaction-to-be-confirmed> [Accessed 21 Jul. 2017].
- Brezo, F. and Bringas, P. (2012). Issues and Risks Associated with Cryptocurrencies such as Bitcoin. The Second International Conference on Social Eco-Informatics. [online] Available at: https://www.researchgate.net/publication/234845612_Issues_and_Risks_Associated_with_Cryptocurrencies_such_as_Bitcoin [Accessed 16 Aug. 2017].
- Chumbley, A. and Moore, K. (n.d.). Merkle Tree | Brilliant Math & Science Wiki. [online] Brilliant.org. Available at: <https://brilliant.org/wiki/merkle-tree/> [Accessed 17 Aug. 2017].
- CoinDesk. (n.d.). A (Short) Guide to Blockchain Consensus Protocols - CoinDesk. [online] Available at: <http://www.coindesk.com/short-guide-blockchain-consensus-protocols/> [Accessed 6 Jul. 2017].
- Coinmarketcap.com. (2017). Bitcoin (BTC) price, charts, and info | Crypto-Currency Market Capitalizations. [online] Available at: <https://coinmarketcap.com/currencies/bitcoin/> [Accessed 12 Feb. 2017].
- Courtois, N. and Bahack, L. (2014). On Subversive Miner Strategies and Block Withholding Attack in Bitcoin Digital Currency. [online] Available at: <https://arxiv.org/pdf/1402.1718v4.pdf> [Accessed 9 Aug. 2017].

- Dabbs, A. (2016). What is a blockchain, and why is it growing in popularity?. [online] Ars Technica. Available at: <https://arstechnica.com/information-technology/2016/11/what-is-blockchain/> [Accessed 11 Jul. 2017].
- Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. 1st ed. [ebook] Satoshi Nakamoto. Available at: <https://bitcoin.org/bitcoin.pdf> [Accessed 12 Feb. 2017].
- Natoli, C. and Gramoli, V. (2016). The Balance Attack Against Proof-Of-Work Blockchains. [online] University of Sydney, pp.1-13. Available at: <https://arxiv.org/abs/1612.09426> [Accessed 7 Feb. 2017].
- Onies, A., Olayinka, T. and Daniele, G. (2017). Technology | Bitcoin. [online] Cs.stanford.edu. Available at: <http://cs.stanford.edu/people/eroberts/cs181/projects/2010-11/DigitalCurrencies/technology/index.html> [Accessed 14 Sep. 2017].
- Question on the terms 'distributed' and 'decentralised'. (2016). [Blog] Stack Exchange. Available at: <https://ethereum.stackexchange.com/questions/7812/question-on-the-terms-distributed-and-decentralised> [Accessed 16 Aug. 2017].

Appendices

Proof 2

To calculate the results for proof 2 I used the following method:

If we start with this equation:

$$\text{mod}(X_1 - X_2) - A \leq 0$$

We need to calculate the probability that it is greater than 0. However, because this represents the interaction between a large number of probabilities I have had to simplify the calculation. Had the λ of the Poisson distribution been large I could of approximated with the normal distribution but in this case it was not possible.

Firstly, for the following tables I am working with a time delay of the expected value which would allow the transaction to be authenticated. I have worked out the probability that $G2 < G1$ or that probability that when $G2$ and $A < G1$. All the conditions when its not possible have been left blank or counted as fail.

As stated previously, this is an estimation and an accurate figure would require calculating an infinite number of probabilities.

For $z = 3$

	G2	1	2	3	4	5
G1		0.149361205	0.224041808	0.224041808	0.168031356	0.100818813
1	0.1494	0.005782027	0.033463154	0.033463154	0.025097366	0.015058419
2	0.224		0.01300956	0.050194732	0.037646049	0.022587629
3	0.224			0.01300956	0.037646049	0.022587629
4	0.168				0.007317877	0.016940722
5	0.1008					0.002634436
						0.336438363

For z
= 6

	G2	4	5	6	7	8
G1		0.133852618	0.160623141	0.160623141	0.137676978	0.103257734
4	0.1339	0.008083727	0.021499828	0.021499828	0.018428424	0.013821318
5	0.1606		0.011640567	0.025799793	0.022114109	0.016585581
6	0.1606			0.011640567	0.022114109	0.016585581
7	0.1377				0.008552253	0.014216213
8	0.1033					0.004810642
						0.237392539

For the following two the time delay statistically large. The expected value is now x .

	G2	x-2	x-1	x	x+1	x+2
G1		0.14936121	0.22404181	0.22404181	0.16803136	0.10081881
x-2	0.1494	0.00578203	0.03346315	0.03346315	0.02509737	0.01505842
x-1	0.224		0.01300956	0.05019473	0.03764605	0.02258763
x	0.224			0.01300956	0.03764605	0.02258763
x+1	0.168				0.00731788	0.01694072
x+2	0.1008					0.00263444
						0.33643836

	G2	x-2	x-1	x	x+1	x+2
G1		0.13385262	0.16062314	0.16062314	0.13767698	0.10325773
x-2	0.1339	0.00808373	0.02149983	0.02149983	0.01842842	0.01382132
x-1	0.1606		0.01164057	0.02579979	0.02211411	0.01658558
x	0.1606			0.01164057	0.02211411	0.01658558
x+1	0.1377				0.00855225	0.01421621
x+2	0.1033					0.00481064
						0.23739254



Candidate personal code:

Extended essay - Reflections on planning and progress form

Candidate: This form is to be completed by the candidate during the course and completion of their EE. This document records reflections on your planning and progress, and the nature of your discussions with your supervisor. You must undertake three formal reflection sessions with your supervisor: The first formal reflection session should focus on your initial ideas and how you plan to undertake your research; the interim reflection session is once a significant amount of your research has been completed, and the final session will be in the form of a viva voce once you have completed and handed in your EE. This document acts as a record in supporting the authenticity of your work. The three reflections combined must amount to no more than 500 words.

The completion of this form is a mandatory requirement of the EE for first assessment May 2018. It must be submitted together with the completed EE for assessment under Criterion E.

Supervisor: You must have three reflection sessions with each candidate, one early on in the process, an interim meeting and then the final viva voce. Other check-in sessions are permitted but do not need to be recorded on this sheet. After each reflection session candidates must record their reflections and as the supervisor you must sign and date this form.

First reflection session

Candidate comments:

I chose to focus my research question away from the general topic of distributed ledger systems towards two specific cryptocurrencies (Ethereum and Bitcoin) on the advice of my supervisor. Although, I knew that I would look at vulnerabilities, in particular the 'balance' vulnerability, I did not envisage having an essay solely based on it. We discussed the new peer reviewed paper on the vulnerability and decided on applying the findings of this paper to bitcoin thus allowing me to explore this protocol with much greater depth and rigor. My supervisor and I decided that alternative research questions focusing on just a single vulnerability may be too limited and similar to previously published papers so we thought a better question would be, "to what extent does the 'balance' vulnerability supersede 'subversive mining' as the most effective threat to Nakamoto's consensus protocol?" I am seeking to use this research question to both improve my understanding of cryptocurrencies and their vulnerabilities as well as to improve my cyber analytical skills which will be useful in later life. Personally, although I was nervous when I started my personal research finding, adequate research material has given me confidence in completing the extended essay.

good analysis if the early stages

Date: February 1, 2017

Supervisor initials:

Interim reflection

Candidate comments:

The extended essay is proving to be challenging yet rewarding experience. I am on track to answer my research question in around 4000 words although I expect that it will be difficult to write fast. I decided to continue to not attempt to collect or create data as this would prove challenging to even a PhD level student. I am instead applying the research conducted by Natoli and Gramoli, and applying it to the Bitcoin as compared to an ethereum based system. Due to the relative infancy of Bitcoin and its study, I have struggled to define terms such as "Block Obliviousness" which are used in academic research but are not 'textbook' material. However, my sources and methodology are all rooted in high quality research and I will be able to use them to answer my research question. I think they have sufficient breadth and depth because I have researched a combination of broad overviews and in depth research material. I am happy that I learnt to use Microsoft Word's citation tool as this has allowed me to keep on top of my sources.

more analysis in both the process and the research focus

Date: July 10, 2017

Supervisor initials:

Final reflection - Viva voce

Candidate comments:

Personally I think I made moderate decisions throughout the extended essay process. Despite me picking a topic which was inherently challenging I managed to complete the essay. I found part way through the essay that a key part of my mathematical arguments would be impossible to complete in the way that I envisaged so I had to modify my argument away from an algebraic proof. If I were to do an extended piece of research again I think I would write the mathematical argument first as really that is the key part. I would love to do more on Nakamoto's Consensus Protocol in the future as it has been a great experience for me.

weak evaluation mainly on the topic focus but has highlighted the journey through the 3 reflections

Date: October 30, 2017

Supervisor initials:

Supervisor comments:

Supervisor: *By submitting this candidate work for assessment, you are taking responsibility for its authenticity. No piece of candidate work should be uploaded/submitted to the e-Coursework system if its authenticity is in doubt or if contradictory comments are added to this form. If your text in the box below raises any doubt on the authenticity of the work, this component will not be assessed.*

Throughout the extended essay processes, _____ has consistently demonstrated excellent organisation and motivation and an unflinching determination to fully understand this complex topic. An impressive piece of work.