DOI:
10.1016/j.comnet.2021.108424

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Platt, M., & McBurney, P. (2021). Sybil Attacks on Identity-Augmented Proof-of-Stake. *COMPUTER NETWORKS*, *199*, [108424]. https://doi.org/10.1016/j.comnet.2021.108424

Sybil Attacks on Identity-Augmented Proof-of-Stake[⋆]

Moritz Platt[*], Peter McBurney

King's College London, Department of Informatics, Bush House, 30 Aldwych, London, WC2B 4BG, UK

Abstract

IdAPoS is an identity-based consensus protocol for decentralised Blockchain networks that implements a trustless reputation system by extending Proof-of-Stake to facilitate leader selection in non-economic contexts. Like any protocol operating in a public/permissionless setting, it is vulnerable to Sybil attacks in which byzantine actors interfere with peer sampling by presenting artificially large numbers of identities. This paper demonstrates what influence these attacks have on the stability of member selection of a Blockchain system using the *IdAPoS* protocol and investigates how attacks can be mitigated. As a novel protocol, its vulnerability to this type of attack has not previously been researched. The research question is approached via an agent-based model of an *IdAPoS* system in which both honest and malicious actors are represented as agents. Simulations are run on some reasonable configurations of an *IdAPoS* system that employ different attack mitigation strategies. The results show that a super strategy that combines multiple individual mitigation strategies is more effective for containing Sybil attacks than the unmitigated protocol and any other individual strategies proposed. In the simulation this strategy extended the time until a system was taken over by a malicious entity approximately by a factor of 5. These positive initial results indicate that further research into the practical viability of the protocol is warranted.

Keywords: Blockchain, Consensus, Proof-of-Stake, Sybil Attack, Leader Selection, Self-Governance
2010 MSC: 68M14, 68U20, 68W15

1. Introduction

Distributed Ledger Technology systems, including Blockchains which constitute a specialisation of those, can be considered decentralised record-keeping systems. In any such system, the selection of trustworthy verifiers that adhere to the protocol by verifying and replicating data without interference is crucial. Blockchain technology,

[⋆] We gratefully acknowledge Google Cloud for in-kind contributions through the Google Cloud Research Grant programme.

[*] Corresponding author

Email addresses: moritz.platt@kcl.ac.uk (Moritz Platt), peter.mcburney@kcl.ac.uk (Peter McBurney)

introduced by the cryptocurrency 'Bitcoin' [1] and further popularised by the application platform 'Ethereum' [2], brought with it a variation of the centrally governed systems previously studied in distributed computing. Both of these decisive technologies, along with many contenders, take inspiration from anti-authoritarian concepts [3] interlinked with the 'Cypherpunk' movement and its objective to enable anonymous economic transactions [4], thereby emphasizing decentralisation and the absence of supervision. In a trustless decentralised environment, devoid of hierarchies, the selection of actors to replicate and evolve records is, however, not trivial and comes with risks. Where multiple replicating actors exist, they must form a consensus as to which of the proposed records are valid and thereby permissible. This requires validators to access the proposed records, which depending on privacy requirements, can occur in the clear or in confidential form [5]. In this context, one of the predominant risks is that of malicious actors gaining disproportionate influence over a decentralised system by inflating their perceived legitimacy, through 'Sybil' attacks [6] in which they pretend to represent artificially large numbers of identities. Once an attacker controls a significant portion of identities, without an appropriate consensus protocol, they can disrupt a trustless system and ultimately render it useless. To counteract such attacks, decentralised systems research has produced various consensus protocols that, among other aims, have the objective of limiting the impact a single entity can make. Common protocols such as Proof-of-Work or Proof-of-Stake use scarce resources, such as computing power or cryptocurrency holdings, to approximate legitimacy. While they are not free from criticism (cf. section 2), it can be acknowledged that they form the foundation of the most significant public Blockchains in the market today and have been providing stable environments for those for years. Arguably, the absence of disastrous security events and their high availability [7] is an indication for their effectiveness. This is predominantly a testament to their incentive compatibility in economic contexts, as those who invest or stake resources in the operation of a system are incentivised to comply with the system's rules.

Such consensus protocols are, however, ineffective in non-economic contexts, i.e. in those in which no means of payment or provisions to store value exist. To address this shortcoming, the idea of *IdAPoS*, an identity-augmented Proof-of-Stake protocol to 'implement the democratic ideal of "One Person/One Vote" in membership selection', has been proposed in earlier work [8] and was subsequently formalised [9]. Rather than employing wealth as a determinant in membership selection, *IdAPoS* introduces the concept of democratically elected authorities that supply voting tokens to achieve Proof-of-Stake-like consensus in non-economic contexts. Another aspect of *IdAPoS* is to enable participants to emit reputational signals that control the value voting tokens have on the network, thereby implementing self-governance. This approach targets the problem of integrating disconnected groups of participants and presents an alternative to cross-chain transactions by making the set of validators subject to majority voting.

Increased interest in cross-chain communication shows that there is a need for connecting disjoint user groups that have formed constituencies on different Blockchains. We argue that cross-chain protocols do not constitute the only viable technology to achieve this. Identity-augmented Proof-of-Stake provides an alternative to cross-chain protocols by combining multiple constituencies within a single Blockchain through democratic principles. A likely challenge to the viability of such a system is their vulnerability to Sybil attacks. To test the hypothesis that effective mitigation strategies to these attacks exist, we performed agent-based simulations of such attacks and investigated the time it took for attackers to overwhelm an identity-augmented Proof-of-Stake system.

2. Related Work

The absence of established standards for Blockchains makes a systematic treatment of their properties difficult and, consequently, no canonical taxonomy of Blockchain consensus protocols exists today [10]. Nonetheless, surveys have been undertaken to categorise protocols. Emerging categories are verifier selection approach, block addition protocol and transaction confirmation method [11]. Considering highly capitalised cryptocurrencies as archetypal applications, it is evident that Proof-of-Work and Proof-of-Stake are the most prominent algorithms on public Blockchains. Both of these consensus protocols follow the longest chain principle for transaction confirmation. From a broader application perspective, additionally, Delegated Proof-of-Stake and Practical Byzantine Fault Tolerance are relevant [12]. Private Blockchains often employ Proof-of-Authority or comparable algorithms that rely on a selection of approved validators [13].

Blockchain and Democracy. The intersection of Blockchain and democracy was explored in works in several academic disciplines. Whether Blockchain constitutes a suitable technology for the coordination of democratic processes has been discussed controversially. While some works emphasise the potential of Blockchain technology to reduce electoral fraud [14], others challenge its suitability in the context of public votes, raising concerns around the vulnerability of the technology to serious failures and its potential to be susceptible to yet unknown attacks [15]. Research findings that question the suitability of Blockchain as a tool for critical systems do so by raising that existing consensus protocols lack precautionary rules to mitigate centralisation of the underlying resources, thereby posing unsuspected risks to system stability [16]. This emphasises the need for more rigorous research into the matter.

As discussed in the introduction, it can be argued that Proof-of-Work cannot be considered a democratic approach to distributed systems governance for its reliance on the inherently unequally distributed resources needed to participate in it. This view is supported by findings in critical theory that show that Proof-of-Work Blockchains, exemplified by Bitcoin, lack the ethical justification necessary to be considered democratic [17]. Other work has shown that, while no existing consensus protocol guarantees systems using it are inherently fair, delegated Proof-of-Stake and Practical Byzantine Fault Tolerance can be considered more genuinely democratic than Proof-of-Work since they 'mitigate the political advantages conferred by material wealth' [18]. Consensus protocols constitute a decisive factor in whether a decentralised system can evolve democratically and Proof-of-Stake appears to be the most the most suitable foundational technology for decentralised identity-based governance.

Proof-of-Work. The foundations of today's consensus protocols were laid before the era of Blockchain when mechanisms to control access to shared digital resources through computing moderately hard functions were first proposed [19]. Later extensions of these mechanisms, such as the hash collision based postage scheme 'Hashcash' [20] that was adopted by the cryptocurrency ‘Bitcoin’ [1], were introduced as fundamental parts of the ‘mining’ process that constitutes a key aspect of Proof-of-Work consensus. Proof-of-Work has been shown to allow for the implementation of both small-scale [21] and large-scale [22] voting applications. These applications do, however, not allow participation in consensus based on voter eligibility, and therefore have different goals from identity-augmented Proof-of-Stake. Proof-of-Work, by design reliant on performing computations that serve no purpose beyond establishing consensus, is heavily criticised for its energy demands [12].

Proof-of-Stake. To reduce the dependency on energy consumption, and thereby achieving higher cost-efficiency, ‘PPCoin’ [23] was proposed. This cryptocurrency revisits Proof-of-Stake, a concept discussed in the context of Bitcoin earlier [24]. In this design a stochastic selection process considering the duration and amount of cryptocurrency held determines who can gain the privilege of block generation on the network. This makes affluent participants more influential in record evolvement than those with smaller holdings. Since this influence leads to a higher likelihood of being rewarded, Proof-of-Stake can suffer from a ‘rich getting richer’ effect correlated with the parameters of the underlying incentive system [25]. Therefore, those parameters have to be scrutinised carefully.

Non-Economic Protocols. While not used widely, numerous other consensus protocols have been proposed in the context of Blockchain [26]. Notable protocols that are appropriate for non-economic contexts include Proof-of-Personhood [27], a protocol that requires participants on a network to mutually attest to their physical presence to prevent Sybil attacks, and Proof-of-Witness Presence [28], similarly requiring participants to provide proof of being physically present at a certain location, albeit automatically. A non-economic protocol that does not rely on physical presence or spatial information is Reputation-Based Consensus [29] in which randomly-selected judges reward cooperative behaviour. Despite these efforts, there is a lack of protocols that allow multiple constituencies, governed by independent identity authorities, to jointly run a decentralised system.

Attacks on IdAPoS. Several attacks on identity-augmented Proof-of-Stake consensus, as described in section 3.1, are conceivable (cf. figure 1). As it constitutes an augmentation of a general longest chain Proof-of-Stake protocol, all attacks on such systems have to be considered. This includes both long-range attacks [30, 31] that are orchestrated over lengthy durations to exploit the wasteless nature of Proof-of-Stake, as well as short-range attacks applicable to common longest chain Proof-of-Stake protocols [32, 33, 34, 35, 36, 37]. A more detailed discussion of these attacks is out of scope within the context of this paper since it focuses exclusively on Sybil attacks on identity-augmented Proof-of-Stake. These attacks, known in the context of Proof-of-Work but absent in common Proof-of-Stake protocol designs, are reintroduced by giving issuers on a network control over the supply of voting tokens and are discussed in a dedicated section later (cf. section 3.2).

Cross-Chain Transactions. A fundamental question that arises when an environment in which multiple constituencies exist is whether these constituencies should share a common technological foundation or whether they should run disparate systems with small,

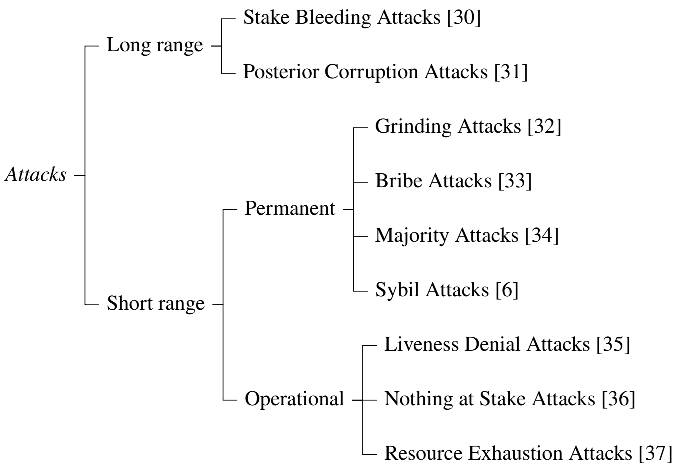
Attacks
Long range
Stake Bleeding Attacks [30]
Posterior Corruption Attacks [31]
Short range
Permanent
Grinding Attacks [32]
Bribe Attacks [33]
Majority Attacks [34]
Sybil Attacks [6]
Operational
Liveness Denial Attacks [35]
Nothing at Stake Attacks [36]
Resource Exhaustion Attacks [37]

Figure 1: Categorisation of attacks on identity-augmented Proof-of-Stake systems.

well-defined integration points. Regardless of approach, it should be considered crucial for systems spanning multiple constituencies to make the reputation of participants portable between constituencies [38]. While *IdAPoS* approaches the problem of multiple constituencies by representing them on the same Blockchain, the alternative approach, in which constituencies operate on disjoint but loosely connected Blockchains, has received significant academic attention in the past. Connecting multiple Blockchains with potentially different approaches to governance, different goals and different technology is known as 'chain interoperability'. Buterin [39] names three strategies to implement chain interoperation: first, an explicit strategy in which dedicated participants agree to carry out an operation on target chain B when an event occurs on the source chain A. Second, a sidechain strategy in which a third system S is employed to process events occurring on A to inform B. Third, a lock-based strategy, in which an operation on B is only carried out if some form of cryptographic proof of an event on A is presented. While strategies that rely on trust between participants can be implemented through sophisticated protocols that do not require locking [40], lock-based strategies, commonly implemented via 'hash time locked contracts' [41], are the only known mechanism to connect two disjoint Blockchains without necessitating trust [42]. Even though recent advances in cross-chain transactions make synchronous inter-blockchain function calls more easily implementable [43], connecting two disjoint chains remains subject to strong assumptions [42]. This is particularly prevalent as lock-based techniques pose the risk of arbitrage in scenarios in which a participant can gain monetary advantage from aborting a cross-chain protocol prematurely [44]. Identity-augmented Proof-of-Stake avoids lock-based cross-chain communication by combining multiple constituencies on the same Blockchain, allowing them to transact atomically.

3. Problem

To motivate the problem of Sybil attacks on identity-augmented Proof-of-Stake, the underlying principles have to be understood. Therefore, this section will summarise *IdAPoS* briefly, showing the parallels to Proof-of-Stake.

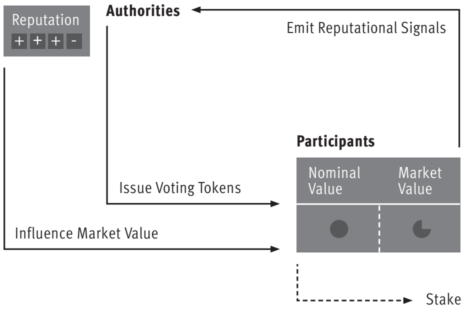

Figure 2: The circular nature of *IdAPoS* [9]: Authorities issue voting tokens with a given nominal value (0..1] to participants. In turn, these become eligible to express their trust towards authorities via reputational signals. Authorities' reputations influence the market value of voting tokens they have issued. Ultimately, the market value is used as stake in the underlying Proof-of-Stake protocol.

3.1. The Identity-Augmented Proof-of-Stake Protocol

The identity-augmented Proof-of-Stake protocol evaluated here, and more extensively discussed in previous work [8, 9], aims to provide a quantifiable approximation of unique identity on a Proof-of-Stake network. Signals forming this approximation are provided by the participant's in the network themselves. As such, its goals are similar to those of reputation systems that 'collect, distribute, and aggregate feedback about participants' past behavior'[45] to allow an informed decision of who can be considered trustworthy on an otherwise virtually anonymous network.

As shown in figure 2, *IdAPoS* considers two types of entities: Regular *participants*, who can engage with the common functionality of a Proof-of-Stake Blockchain (i.e. transact with others) and express their trust in authorities via reputational messages and *authorities*, who can issue voting tokens to those participants they believe to represent unique identities. Voting tokens replace cryptocurrency holdings as they are common in Proof-of-Stake. They cannot be traded between participants but exclusively serve as a marker of an authority's trust. Voting tokens have a nominal value $nv \in (0, 1]$ that represents the trust the issuing authority has in the recipient of the voting tokens. Reputational signals: endorsements r_+ and discouragements r_-, express a participant's trust or mistrust in an authority. These signals influence the trustworthiness $t_a \in [0, 1]$ of an authority. This value in turn determines the market value $mv \in [0, 1]$ of voting tokens issued by an authority. The average of market values of voting tokens a participant holds represents a numerical approximation of whether this participant represents a unique identity. It is then used as a basis for Proof-of-Stake miner selection. The identity-augmented Proof-of-Stake protocol thus roughly follows the following steps:

1. Discover new authorities based on endorsements

2. Evaluate reputational signals to calculate trustworthiness of authorities using nominal values
3. Calculate nominal voting token balances for participants
4. Calculate the market value of all participants' holdings
5. Perform regular Proof-of-Stake protocol [1]

Algorithm 1 describes the market value calculation that occurs in the mining phase in more detail. It takes the inputs described in table 1.

Variable	Description
B	The 'block'; a set of all mined transactions
H	A map of maps containing the nominal holdings of voting tokens of a participant (Participant $\mapsto$ (Authority $\mapsto$ Nominal voting tokens value)) as established in previous blocks
R	A map of maps containing the trust scores expressed by participants (Authority $\mapsto$ (Participant $\mapsto$ Nominal reputational value)) as established in previous blocks

Table 1: Inputs and backing data structures for the identity-augmented Proof-of-Stake protocol

The algorithm returns a map containing the market values of tokens per holder (Participant $\mapsto$ Voting tokens market value) which forms the base for Proof-of-Stake miner selection.

Issuing Voting tokens with Nominal Values. An identity authority can issue tokens of value $(0, 1]$ to any participant. An authority can issue at most one token per participant. This value is called the nominal value. The average value of voting tokens issued to a beneficiary approximates the trust the network has in them representing a single, trustworthy identity. In most cases, identity authorities will issue a static number of tokens (e.g. 1) to all participants they support, except in cases where they use sophisticated assessment methods that allow them to quantify trust levels on a per-participant basis.

Calculating Trustworthiness Scores. Identity authorities are assigned trustworthiness scores. These scores are based on reputational signals emitted by participants towards the authority in question. They are positive in the case of endorsements and negative in the case of discouragements. They are calculated by summing the average nominal values of voting tokens holdings of the participant issuing the reputational signal (with a negative sign in case of discouragement). Should the sum be negative, the trustworthiness score is 0. Should the sum be positive, an attenuation function $f : \mathbb{R}_+ \mapsto [0, 1]$ is applied.

[1] The consensus protocol backed by the voting tokens issued, can be any generic Proof-of-Stake protocol. The specifics of which are out of the scope of this paper as it is concerned with identity-augmentation only.

Algorithm 1 The identity-augmented Proof-of-Stake protocol [9] determines the compound market value of voting tokens per participant

```
Require: See table 1
  for $tx \in B$ do
      if $tx$ is IssuanceTransaction then
          $H\{tx.\text{recipient}\}\{tx.\text{sender}\} \leftarrow tx.\text{value}$
      else if $tx$ is Endorsement then
          $R\{tx.\text{recipient}\}\{tx.\text{sender}\} \leftarrow$ GetNominalHoldings($tx$.sender)
      else if $tx$ is Discouragement then
          $R\{tx.\text{recipient}\}\{tx.\text{sender}\} \leftarrow -$ GetNominalHoldings($tx$.sender)
      end if
  end for
  $V \leftarrow \varnothing$
  for $p \in H$.keys do
      $vt \leftarrow \varnothing$
      for $a \in H\{p\}$.keys do
          $tr \leftarrow$ GetTrustworthiness($a$)
          $vt \leftarrow vt \cup H\{p\}\{a\} \cdot tr$
      end for
      $V\{p\} \leftarrow \overline{vt}$                    ▷ $\overline{x}$ denotes the mean of values in $x$
  end for
  return $V$
  procedure GetNominalHoldings(p)
      $nh \leftarrow \varnothing$
      for $h \in H\{p\}$.keys do
          $nh \leftarrow vt \cup H\{p\}\{h\}$
      end for
      return $\overline{nh}$                    ▷ $\overline{x}$ denotes the mean of values in $x$
  end procedure
  procedure GetTrustworthiness(a)
      $tr \leftarrow 0$
      for $r \in R\{a\}$.keys do
          $tr \leftarrow tr + R\{a\}\{r\}$
      end for
      if $tr < 0$ then
          return 0
      else
          return Attenuate($tr$)          ▷ An attenuation function, e.g. $\frac{2}{1+e^{-0.01tr}} - 1$
      end if
  end procedure
```

Calculating Compound Market Values. In order to calculate the compound market value of voting tokens a participant holds, first the nominal values of all tokens issued to them have to be obtained. Subsequently, these nominal values are multiplied by the trustworthiness score of the issuing authority. The compound value is the mean of these values.

Proof-of-Stake Miner Selection. The voting tokens value applicable to Proof-of-Stake miner selection is the compound market value of the token which is based on the overall reputation of the issuing authorities. As nominal values are irrelevant for miner selection, the reputation of an authority has a direct effect on which participants are influential in the system, thereby tying their influence directly to the reputation of the authority they are affiliated with.

3.2. Sybil Attacks on Identity-Augmented Proof-of-Stake

While Proof-of-Stake brought with it a variety of previously unknown attacks (cf. section 2), Sybil attacks are not an attack vector commonly considered applicable to them. Commonly, the total wealth of cryptocurrency available within a Proof-of-Stake network is subject to a systemic limit. This limit can be static or increase continuously through inflation incurred by mining rewards [46]. Since validator selection is contingent on the share of wealth a participant holds, the wealth limit acts as an inbuilt mechanism to prevent Sybil attacks. In an identity-augmented Proof-of-Stake system, where issuers can bring new tokens into circulation at will, such an upper bound is, however, not conceivable. Smaller authorities might only issue voting tokens infrequently to small numbers of participants while larger ones might issue them in high frequency. A reasonable default could be to introduce no upper bound at all, allowing authorities to issue arbitrary amounts of voting tokens to the participants they trust. This would allow for the explosive growth of wealth, potentially leading to a concentration of it originating from a single authority. While this is unproblematic in a scenario where all authorities are trustworthy, it introduces the risk of Sybil attacks: Malicious authorities can issue large numbers of voting tokens to accomplice accounts, allowing them to inflate their influence in the identity-augmented Proof-of-Stake protocol, thereby creating enough voting tokens to conduct a successful majority attack. In the absence of protocol-level limitations on the issuance of voting tokens, a malicious authority would only be constrained by technical parameters (e.g. maximum block sizes that limit how many transactions can be included per period). This limitation would provide no substantial protection from Sybil attacks, making the system vulnerable to a takeover within a short time.

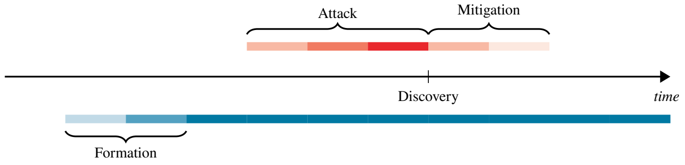

Figure 3: The anatomy of a thwarted Sybil attack on an identity-augmented Proof-of-Stake system. The bottom bar symbolises a trustworthy constituency, the top bar a malicious one.

Despite these blatant risks of a takeover, it can be assumed that an attack spans at least several periods (blocks). Therefore, the following picture of an attack emerges: A Sybil attack on an identity-augmented Proof-of-Stake system follows three stages (cf. figure 3). A formation stage, in which the attacker has not started the attack, but one or more trustworthy constituencies are established and develop. An attack stage, starting with the undermining of an identity authority leading to the creation of malicious authorities that begin issuing frivolous voting tokens. And finally, should the attack be discovered by the wider network, a mitigation phase in which the wider network takes action to remove the malicious authority. Some mitigation strategies for these attacks, all revolving around limiting the growth of wealth of voting tokens originating from a single malicious issuer, have been proposed [8]. These are further discussed in this section. Their effectiveness will subsequently be evaluated in section 5.

3.2.1. Temporal Normalisation of Reputational Signals

When an untrustworthy authority joins the network, an attack conducted by them will be characterised by a sudden influx of frivolous voting tokens. While an attenuation function addresses this as part of the original protocol (cf. section 3.1), this affects only the market value of voting tokens, and therefore the mechanics of Proof-of-Stake miner selection, but not the reputational signals. This can lead to voting tokens of trustworthy constituents rapidly losing market value. Temporal normalisation is a strategy that seeks to mitigate these sudden changes. This strategy mandates that participant's reputational signals mature in value. Similar to market value normalisation, this strategy can apply any function $f : \mathbb{N} \mapsto [0, 1)$, such as a sigmoid function, to the age of the participant. The age of the participant is defined as the number of steps that have passed since they have first received voting tokens. The result of this normalisation is then multiplied with the nominal value of the voting tokens issued to provide a normalised signal strength $[0, 1]$.

3.2.2. Wealth Ceilings

Assuming an attacker conducts attacks that take the time-normalisation of voting token market values into account, this safeguard alone cannot be effective. When a large number of Sybil identities is created by one authority, this authority can, as a result, issue a large number of voting tokens in very short time. Even when using a defensive temporal normalisation function, their cumulative market value will be significant. Similarly to the unmitigated protocol, this renders the efficacy of temporal normalisation contingent solely on the pace at which accomplice accounts can be funded by the attacking authorities. To preempt such attacks, a constituency wealth ceiling that limits the total funds issuable by one authority, can be enforced. This could be a static maximum number of voting tokens issuable by a single authority or a dynamic function that limits the permissible growth of voting tokens depending on the age of the issuing authority. Both can be implemented through ceiling functions $f : \mathbb{N} \mapsto \mathbb{N}$, such as $f(a) = 100 \cdot s$ in the case of a dynamic ceiling or $f(a) = 10,000$ in the case of a static ceiling, where a denotes the steps passed since an authority was first mentioned.

3.2.3. Authority Pruning

Once authorities on the network receive increased rates of discouragement messages, they are likely to pose a danger to the system's stability. For this reason, a declining identity authority should inevitably lose all influence over the network, even when the voting tokens distributed by them still hold value in arithmetic terms. To enforce this, an upper bound for the maximum number of discouragement messages can be introduced. Once an authority reaches the threshold, the protocol would fully devaluate the voting tokens provided by them.

4. Method

The research question is addressed through an agent-based approach. While approaching it through mathematical proof might provide a more reliable answer, agent-based modelling allows to frame Sybil attacks on *IdAPoS* as an emergent phenomenon, the effects of which are not derivable from the inspection of the model parameters alone. This allows for the relaxing of assumptions that may be difficult to accurately specify in the context of mathematical analysis [47]. For this, a system model (cf. section 4.1) is designed and implemented in the 'Kotlin' programming language. The system model is executed stepwise with steps representing blocks. At the end of a step, a number of proposed transactions are randomly sampled from the memory pool and applied to the ledger atomically, similar to how transactions would be included in a block through mining in an actual Blockchain system.

Environmental Values. Table 2 shows the environmental parameters of the system simulation along with the values assumed. Apart from the value for the block size c, that roughly aligns with the average number of transactions per block on the Bitcoin Blockchain, the environmental parameters are curated to provide an *illustrative* environment for a simulation, meaning, they should provide underlying conditions in which mitigation strategies are observable well. In extremis, i.e. where a feeble attacker (characterised by low g_m, me, mn, md) is met by a potent group of existing vigilant constituencies (characterised by high g_h, lf, p_{vig}, p_{he}, p_{ho}, p_{hd}), or vice versa, the simulation cannot produce meaningful outputs. This stems from the very short attack phases scenarios with such parameters suffer from that make it impossible to meaningfully analyse the interplay between opposing actors.

In addition to the environmental parameters which, in a system operating in the wild, would be dictated by the environment it operates in, several aspects of an *IdAPoS* system are explicitly configurable as part of the protocol. Table 3 shows the parameters with which the simulation can be configured, as well as the variations of candidate values that are tested during simulation. Section 5 describes in more detail how combinations of the configurable parameters influence the system's resistance to Sybil attacks.

The outputs of the simulation, the ledger, are written to a log file[2]. The model avoids randomness where possible. Where unavoidable, it uses seeded pseudorandom number generators (PRNG) to make the results of the simulation deterministic and reproducible.

[2] The log files for the different mitigation strategies are published as supplementary dataset [48].

Variable	Description	Value
lf	The number of steps the formation phase in which honest constituencies build up is comprised of	200
li	The number of steps the attack remains undiscovered is comprised of	5
la	The number of steps the mitigation phase in which the malicious activity is discovered is comprised of	100
c	The number of transactions a block holds	3000
p_{vig}	The probability with which an honest participant is created as a 'vigilant' participant, i.e. they discourage malicious authorities	2%
vig	The number of malicious authorities a 'vigilant' participant tries to discourage in a step	5
p_{he}	The probability with which an honest participant endorses a new honest authority in a step	0.01%
p_{ho}	The probability with which an honest participant endorses their own authority in a step	2%
p_{hd}	The probability with which an honest participant discourages their own authority in a step	0.1%
g_h	The growth rate of an honest constituency (endorsement messages emitted/period)	2
g_m	The growth rate of a malicious constituency (endorsement messages emitted/period)	10
me	The number of endorsement messages a malicious participant emits in support of existing malicious authorities	5
mn	The number of endorsement messages a malicious participant emits in support of new malicious authorities	2
md	The number of discouragement messages a malicious participant emits towards existing honest authorities	3

Table 2: Environmental parameters of the system simulation

Value	Description	Domain
at(r)	The attenuation function that is applied to the sum of reputational values (r)	$f : \mathbb{R}_+ \mapsto [0, 1]$
tn(a)	The temporal normalisation function that dampens the value of reputational signals depending on the age of the emiter (a)	$f : \mathbb{N} \mapsto [0, 1)$
wc(a)	The wealth ceiling that is applied to the total issuable amount of voting token as a function of the issuer's age (a)	$f : \mathbb{N} \mapsto \mathbb{N}$
pr	Whether authority pruning is applied	$\mathbb{B}$

Table 3: Configurable aspects of *IdAPoS* and their domains

PRNG are used where the model demands probabilistic selection, i.e. p_{vig}, p_{he}, p_{ho} and p_{hd} as described in table 2, as well as random selection from the memory pool.

4.1. System Model

The simulation (cf. algorithm 2) is initialised with an empty ledger, an empty memory pool and an empty set of participants. It follows the sequence of instructing the agents to act (ACT), which results in those agents emitting transaction proposals to the memory pool, and subsequently selecting a number of transaction proposals from said pool randomly. The selected proposals are then committed to the ledger (MINE). Any effects of the selected proposals on the constituency of the system, i.e. where new participants or authorities are introduced, will be registered by the simulation. New agents will be created accordingly and references to them will be stored within the set of participants. Starting with the following iteration, these newly created agents will participate in the simulation and will be called during the ACT phase of a step.

Within the simulation programme, assumptions are made regarding both the environmental parameters and the strategies that the agents pursue. Strategies that agents adopt are based on intuition and no work has been done to optimise strategies of either type of agent. Despite this, the simulation–even where reasonable but suboptimal strategies are employed–should allow for the effectiveness of countermeasures to be compared. The system model is free of any strategic considerations and merely acts as a layer to orchestrate agent actions. Furthermore, it provides data structures to hold all transactions and, in the interest of optimisation, additional data structures that provide convenient access to voting tokens holdings and reputational signals. These can be thought of as caches.

Memory Pool. A Sybil attack on an identity-augmented Proof-of-Stake system is accompanied by an explosion in the number of transactions proposed. Common Blockchain protocols, however, impose significant limits on the number of transactions included in a block. Optimal block sizes in Proof-of-Work systems are severely limited [49]. This limitation is imitated in the simulation by introducing the concept of a naïve memory

Algorithm 2 The algorithm simulating an identity-augmented Proof-of-Stake system

```
Require: See table 2
  $i \leftarrow 0$
  $L \leftarrow \emptyset$                      ▷ The ledger containing committed transactions
  $M \leftarrow \emptyset$                ▷ The memory pool containing uncommitted transactions
  $A \leftarrow \emptyset$                           ▷ The agent instances in the simulation
  $M \leftarrow GenesisTx$
  MINE
  repeat                         ▷ Formation phase, attack phase, mitigation phase
      ACT
      MINE
      $i \leftarrow i + 1$
  until $i \geq lf + li + la$
  procedure MINE
      $txs \leftarrow$ RANDOMSAMPLE(c, M)          ▷ Picks $c$ elements from $M$ randomly
      for $tx \in txs$ do
          $L \leftarrow L \cup tx$
          if $tx$ is GenesisTransaction then
              $A \leftarrow A \cup$ HonestAuthority()
          else if $tx$ is IssuanceTransaction then
              if $tx.sender$ = honest then
                  $A \leftarrow A \cup$ HonestParticipant()
              else
                  $A \leftarrow A \cup$ MaliciousParticipant()
              end if
          else if $tx$ is ReputationalTransaction then    ▷ Endorsement/discouragement
              if $tx.sender$ = honest then
                  $A \leftarrow A \cup$ HonestAuthority()
              else
                  $A \leftarrow A \cup$ MaliciousAuthority()
              end if
          end if
      end for
      $M \leftarrow \emptyset$
  end procedure
  procedure ACT
      for $a \in A$ do
          $M \leftarrow M \cup$ A.ACT
      end for
  end procedure
```

pool that stores proposed transactions locally until they are included in a valid block. Since considerations of the economic properties of transaction mining are beyond the scope of this paper, a simplified strategy for selecting transactions from this pool is applied: In each step of the protocol, transactions, up to the ‘block size’ defined, are chosen randomly from the set of uncommitted transactions.

4.2. Agents

To inform their actions, all agents have visibility of the current ledger (L). Through this, they have knowledge of all participants, along with their voting tokens holdings, and all authorities, along with the reputational signals received by them. They also understand whether their co-actors are honest or malicious. Based on this information, they execute a predefined strategy, encoded in their respective Act methods, which leads to a number of transaction proposals being emitted to the memory pool.

Honest Authorities. Honest authorities facilitate a steadily growing constituency by following a strategy in which they issue a steady stream of voting tokens to new honest participants up to the maximum number of messages configured (cf. algorithm 3).

Algorithm 3 The programme an agent representing a *honest* authority executes

```
Require: The current ledger $L$, the current memory pool $M$, and all configuration
  parameters described in table 2
  procedure Act
      $i \leftarrow 0$
      repeat
          $M \leftarrow M \cup$ Issuance(HonestParticipant())
          $i \leftarrow i + 1$
      until $i \geq g_h$
  end procedure
```

Malicious Authorities. Malicious authorities follow a strategy (cf. algorithm 4) in which they issue voting tokens to new malicious participants up to the maximum number of messages configured.

Algorithm 4 The programme an agent representing a *malicious* authority executes

```
Require: The current ledger $L$, the current memory pool $M$, and all configuration
  parameters described in table 2
  procedure Act
      $i \leftarrow 0$
      repeat
          $M \leftarrow M \cup$ Issuance(MaliciousParticipant())
          $i \leftarrow i + 1$
      until $i \geq g_m$
  end procedure
```

A more complex strategy might provide benefits in a scenario where the malicious nature of an actor is not easily discoverable and actors have to analyse the behaviour of their co-actors to it. In such a scenario, malicious authorities might disguise as honest ones and issue an artificially low number of tokens to avoid being discovered.

Honest Participants. Honest participants are mostly passive. Their strategy is defined based on the assumption that they make use of an identity-augmented Proof-of-Stake system for their individual purposes but are less concerned with participating in its governance. Algorithm 5 shows how they infrequently endorse authorities, and, if they are instantiated as 'vigilant' actors, how they discourage malicious authorities.

Algorithm 5 The programme an agent representing a *honest* participant executes

```
Require: The current ledger $L$, the current memory pool $M$, and all configuration
  parameters described in table 2
  procedure ACT
    shouldEndorseNew ← GETBOOLEAN($P_{eh}$)          ▷ Returns True with $P = P_{eh}$
    shouldEndorseOwn ← GETBOOLEAN($P_{eo}$)
    shouldDiscourageOwn ← GETBOOLEAN($P_{ed}$)
    isVigilant ← GETBOOLEAN($P_{vig}$)
    if shouldEndorseNew then
      $M \leftarrow M \cup$ Endorsement(HonestAuthority())
    end if
    if shouldEndorseOwn then
      $M \leftarrow M \cup$ Endorsement(ownAuthority)
    end if
    if shouldDiscourageOwn then
      $M \leftarrow M \cup$ Discouragement(ownAuthority)
    end if
    if isVigilant then
      MA ← MALICIOUSAUTHORITIES(L)
      targets ← RANDOMSAMPLE(vig, MA)
      for $target \in targets$ do
        $M \leftarrow M \cup$ Discouragement(target)
      end for
    end if
  end procedure
```

It is not specified how honest participants become aware of malicious activity on the network. Conceivably, in an actual implementation of a protocol, they would become aware through off-ledger communication. It is also not modelled here how an honest participant would form an understanding of whether another actor on the network is trustworthy or malicious. In an actual implementation of the protocol, where this information would have to be derived from the ledger alone, identification of malicious actors would be imprecise. Imitating this behaviour in the context of a simulation requires deep game-theoretical consideration that goes beyond the scope of this paper.

Malicious Participants. Malicious participants follow a strategy in which they emit messages endorsing existing and new malicious authorities as well as discouraging honest authorities (cf. algorithm 6). The number of messages emitted is defined in environmental parameters.

Algorithm 6 The programme an agent representing a *malicious* participant executes

```
Require: The current ledger $L$, the current memory pool $M$, and all configuration
  parameters described in table 2
  procedure ACT
      $M \leftarrow$ GETMALICIOUSAUTHORITIES($L$)
      $mas \leftarrow$ RANDOMSAMPLE($mn$, $M$)          ▷ Picks $c$ elements from $M$ randomly
      for $ma \in mas$ do
          $M \leftarrow M \cup$ Endorsement(ma)
      end for
      $H \leftarrow$ GETHONESTAUTHORITIES($L$)
      $has \leftarrow$ RANDOMSAMPLE($md$, $H$)
      for $ha \in has$ do
          $M \leftarrow M \cup$ Discouragement(ha)
      end for
      $i \leftarrow 0$
      repeat
          $M \leftarrow M \cup$ Endorsement(MaliciousAuthority())
          $i \leftarrow i + 1$
      until $i \geq mn$
  end procedure
```

This strategy is assumes that malicious agents do not communicate directly, an assumption that could be relaxed in future work to produce a more aggressive attack strategy. Conceivable improvements would entail selecting endorsing authorities with high trustworthiness instead of picking the authorities to be endorsed randomly.

5. Results

A metric to evaluate the effectiveness of mitigation strategies is the sum of market values of voting tokens (mv) that participants hold. This allows to compare the influence that both honest and malicious participants have on the mining process. A system can be considered overwhelmed at the latest once the sum of market values of voting tokens held by malicious participants exceeds that of voting tokens held by honest participants (i.e., the total value held by malicious participants exceeds $^1/_2$). While this limit can be considered to be optimistic under the assumption that common proof-of-stake protocols require $^2/_3$ of the network to be in agreement in order for it to to remain deadlock-free [34], $^1/_2$ can serve as a firm upper bound. The values are evaluated per protocol step (s), i.e. after selecting and, subsequently, applying transaction proposals selected from the memory pool.

Strategy	at(r)	tn(a)	wc(a)	pr	p_s	p_o	Δ
i	$\frac{2}{1+e^{-0.1r}}-1$	-	-	no	200	203	3
	$\frac{2}{1+e^{-0.01r}}-1$	-	-	no	200	204	4
	$\frac{2}{1+e^{-0.001r}}-1$	-	-	no	200	204	4
	$\frac{2}{1+e^{-0.0001r}}-1$	-	-	no	200	204	4
	$\frac{2}{1+e^{-0.00001r}}-1$	-	-	no	200	204	4
	$\frac{2arctan(0.1r)}{\pi}$	-	-	no	200	204	4
	$\frac{2arctan(0.01r)}{\pi}$	-	-	no	200	204	4
	$\frac{2arctan(0.001r)}{\pi}$	-	-	no	200	204	4
	$\frac{2arctan(0.0001r)}{\pi}$	-	-	no	200	204	4
	$\frac{2arctan(0.00001r)}{\pi}$	-	-	no	200	204	4
ii	$\frac{2}{1+e^{-0.01r}}-1$	$\frac{2}{1+e^{-0.1a}}-1$	-	no	200	210	10
	$\frac{2}{1+e^{-0.01r}}-1$	$\frac{2}{1+e^{-0.01a}}-1$	-	no	200	217	17
	$\frac{2}{1+e^{-0.01r}}-1$	$\frac{2}{1+e^{-0.001a}}-1$	-	no	200	218	18
	$\frac{2}{1+e^{-0.01r}}-1$	$\frac{2}{1+e^{-0.0001a}}-1$	-	no	200	221	21
	$\frac{2}{1+e^{-0.01r}}-1$	$\frac{2}{1+e^{-0.00001a}}-1$	-	no	200	217	17
iii	$\frac{2}{1+e^{-0.01r}}-1$	-	10	no	200	202	2
	$\frac{2}{1+e^{-0.01r}}-1$	-	100	no	200	203	3
	$\frac{2}{1+e^{-0.01r}}-1$	-	1000	no	200	204	4
	$\frac{2}{1+e^{-0.01r}}-1$	-	10000	no	200	204	4
	$\frac{2}{1+e^{-0.01r}}-1$	-	100000	no	200	204	4
iv	$\frac{2}{1+e^{-0.01r}}-1$	-	a	no	200	207	7
	$\frac{2}{1+e^{-0.01r}}-1$	-	$2a$	no	200	206	6
	$\frac{2}{1+e^{-0.01r}}-1$	-	$4a$	no	200	205	5
	$\frac{2}{1+e^{-0.01r}}-1$	-	$8a$	no	200	204	4

	$\frac{2}{1+e^{-0.01r}} - 1$	-	$16a$	no	200	204	4
v	$\frac{2}{1+e^{-0.01r}} - 1$	-	-	yes	200	204	4
vi	$\frac{2}{1+e^{-0.01r}} - 1$	-	4a	yes	200	205	5
vii	$\frac{2}{1+e^{-0.01r}} - 1$	$\frac{2}{1+e^{-0.001a}} - 1$	$4a$	no	200	218	18
viii	$\frac{2}{1+e^{-0.01r}} - 1$	$\frac{2}{1+e^{-0.001a}} - 1$	-	yes	200	219	19
ix	$\frac{2}{1+e^{-0.01r}} - 1$	$\frac{2}{1+e^{-0.001a}} - 1$	$4a$	yes	200	223	23

Table 4: Comparison of mitigation strategies using temporal normalisation functions (tn(a)), wealth ceiling functions (wc(a)), authority pruning (pr) and various attenuation functions (at(r)). The table shows when the attack started (p_s), when the system was overwhelmed (p_o) and the number of periods in between (Δ).

Table 4 shows how the different mitigation strategies, described in more detail later, influence the takeover time of an *IdAPoS*-based system. An 'unmitigated' simulation run serves as the baseline to compare strategies against. This is the result of *IdAPoS* consensus as formalised in algorithm 1 being executed.

5.1. Unmitigated Protocol

The visualisation of the sum of market values of voting tokens provided in figure 4 shows that, once the attack commences, the market value of honest participants' voting tokens declines steeply while the value of voting tokens held by malicious actors appears to grow linearly.

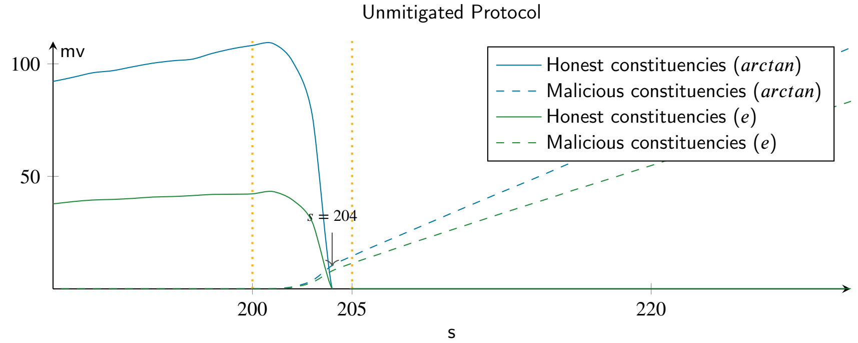

Figure 4: After 204 periods (s) the market value (mv) of tokens held by malicious authorities exceeds that held by honest ones.

When no mitigation strategies are applied, the system can be taken over within a brief period prior to the attack being discovered (cf. table 2). This confirms the intuitive assumption made earlier (cf. section 3.2) that *IdAPoS* is highly vulnerable to Sybil attacks if no mitigation strategies are applied.

5.2. Temporal Normalisation of Reputational Signals

To evaluate the temporal normalisation strategy, the effect of a temporal normalisation function $f(r) = \frac{2}{1+e^{-d \cdot r}} - 1$ with varying growth rates d is simulated. Table 4 shows that $d = 0.0001$ yielded the best results. Here, t represents the age of the holder which is defined as the number of periods between the holder receiving the first endorsement message and the current step. This modifier is subsequently applied to all reputational signals emitted by the holder.

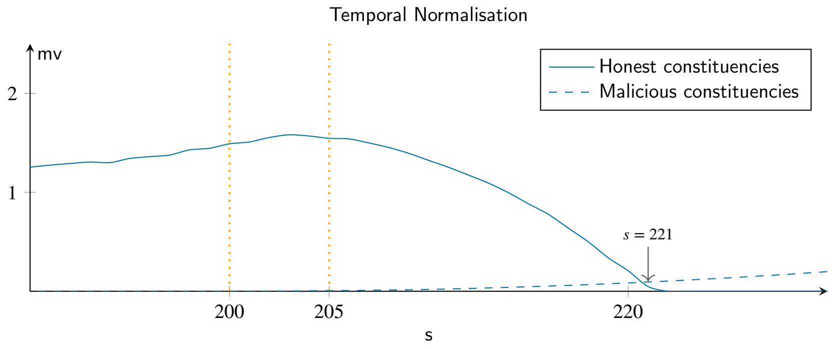

Figure 5: When applying temporal normalisation with the temporal normalisation function $f(r) = \frac{2}{1+e^{-0.0001r}} - 1$ only after 221 periods (s) the market value (mv) of tokens held by malicious authorities exceeds that held by honest ones.

The diminishing effect this has on the sum of market values of voting tokens held by honest participant is evident from the plot shown in figure 5. In the context of temporal normalisation, the attenuation function used is of particular importance. When using a logistic function $f(x) = \frac{1}{1+e^{-k(x-x_0)}}$ as attenuation function, as in the experiment, the logistic growth rate influences the point of overwhelming the system. A steeper curve will lead to a system being overwhelmed more quickly, while a flatter curve will push out the time at which the system is overwhelmed. In the extreme case ($k = 0$) will never be overwhelmed but new constituencies will not be able to join the system. Consequently, where $k = \infty$, temporal normalisation will have no effect.

5.3. Wealth Ceilings

Wealth ceilings define the maximum number of voting tokens an issuer can bring into circulation up to a certain age t. Here, t is defined as the number of steps passed between the authority receiving their first endorsement message and the current step. Ten experiments are executed with different modalities of wealth ceilings: First, static wealth ceilings are applied, second dynamic wealth ceilings. Figure 6 shows how the choice of wealth ceiling parameters influences takeover time. While the most severely limiting constant function tested ($wc(a) = 10$) has a negative effect on takeover time, likely to due the capping of the growth of the trustworthy authority in the formation phase, dynamic wealth ceilings, particularly those that map directly to the constituency size (i.e. $wc(a) = a$), have a small positive effect.

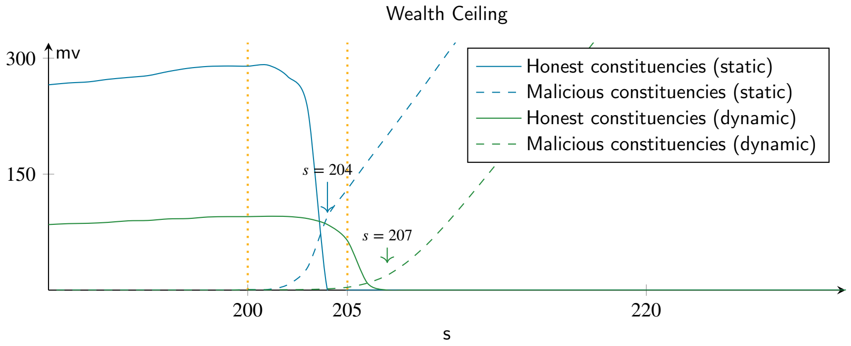

Figure 6: The simulation of static and dynamic wealth ceiling strategies shows that static strategies are less effective, with $wc(a) = 10$ leading to the system being overwhelmed after only 204 periods (s) when the market value (mv) of tokens held by malicious authorities exceeds that held by honest ones. Dynamic strategies (e.g. $wc(a) = a$) lead to minor improvements in system stability.

5.4. Authority Pruning

A threshold is reached where discouragement messages an authority has received exceed the number of endorsement messages an authority has received. During each step, it can be calculated whether this threshold has been reached per authority. Should an authority cross it, it will be removed from the pool of actors. Figure 7 shows how this strategy in its current form has no effect on takeover time. This is likely because trustworthy authorities have already received large numbers of endorsement messages before the start of the attack.

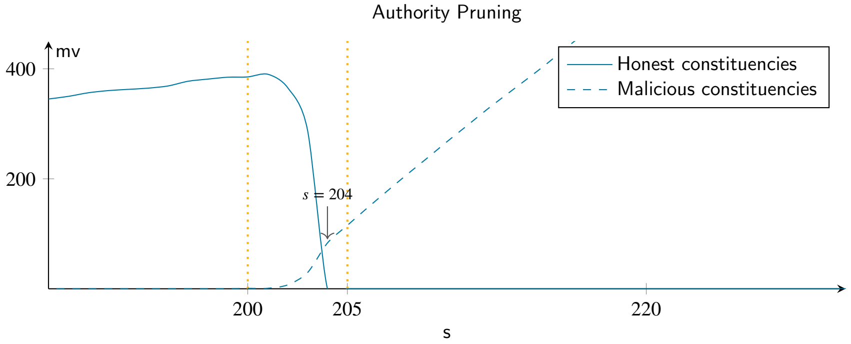

Figure 7: Authority pruning does not provide a measurable improvement over the unmitigated protocol: similar to the this, the market value (mv) of tokens held by malicious authorities exceeds that held by honest ones after only 204 periods (s).

5.5. *Combining Strategies*

The strategies previously presented in isolation are not mutually exclusive. Therefore, ‘super strategies’ can be devised. To create those, representative configurations of strategies previously tested are combined. Figure 8 shows how super strategies improve the performance of the underlying individual strategies.

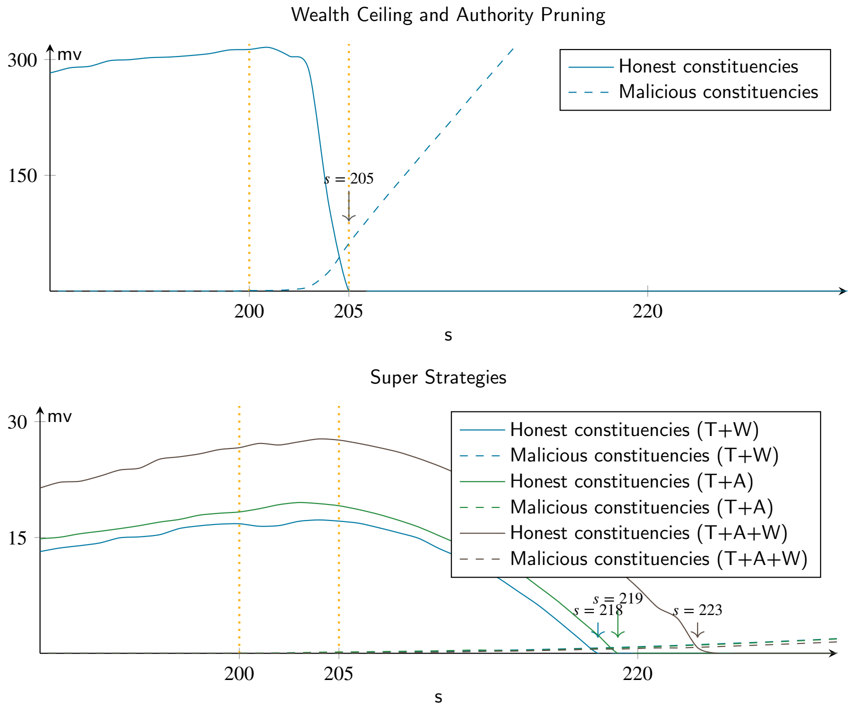

Figure 8: Combining the individual mitigation strategies of temporal normalisation (T), wealth ceilings (W) and authority pruning (A) creates super strategies that can prolong the time (s) it takes until the market value (mv) of tokens held by malicious authorities exceeds that held by honest ones.

The most performant super strategy simulated is the ‘TAW’ (temporal normalisation, authority pruning, wealth ceiling) strategy. Here, a logistic attenuation function ($at(r) = \frac{2}{1+e^{-0.01r}} - 1$) is used in combination with temporal normalisation ($tn(a) = \frac{2}{1+e^{-0.001a}} - 1$), authority pruning, and a dynamic wealth ceiling ($wc(a) = 4a$).

5.6. *Memory Pool Oversaturation*

Simulation reveals that the number of messages submitted to an identity-augmented Proof-of-Stake system under attack is very large. This is due to malicious participants aggressively endorsing accomplices. As discussed before (cf. section 4.1), the simulation imitates common Blockchain protocols by severely limiting the maximum number of transactions included in a block.

As shown in figure 9, this leads to the vast majority of transactions proposed to the memory pool being discarded in the later phases of an attack. The probability of a

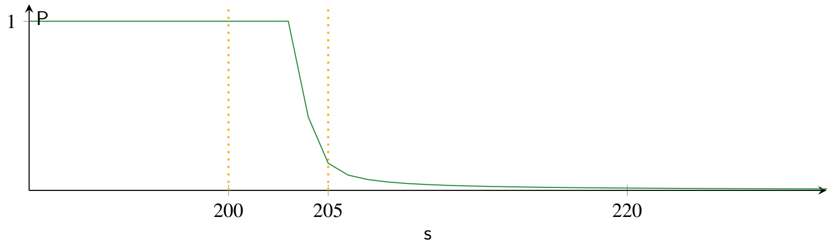

Figure 9: The probability P of a message submitted into the naïve memory pool being mined decreases steeply after an attack commences in step $s = 200$.

proposed transaction being executed in a block is still 100% in step 200 when the attack commences, but drops to under 1% by step 227. While economic strategies for miners that lead to including some transactions while excluding others have been researched in the context of cryptocurrencies [50, 51], the simplistic random selection method applied to the naïve memory pool does not consider these. Transactions that are not selected by the end of the simulation are discarded.

6. Conclusion

Permissionless blockchain technology, by virtue of being trustless and therefore not relying on central authorities, is an appropriate building block to enable self-governance of systems, small and large. Yet, established consensus protocols for such systems do not exist, and common protocols, albeit popular, can be considered inherently undemocratic. *IdAPoS* targets this shortcoming, but is highly susceptible to Sybil attacks. In this paper we apply agent-based simulation to show that mitigation strategies exist that can effectively prolong the time it takes for an attacker to take over an identity-augmented Proof-of-Stake system. Temporal normalisation, a technique through which established participants are given more weight, proved to be the most effective individual mitigation strategy simulated. Individual strategies were outperformed by 'TAW', a super strategy that combines temporal normalisation with authority pruning, a strategy that removes authorities that have received predominantly negative feedback, and wealth ceilings that limit the total number of identities a single authority can control. Using this super strategy, the system was able to withstand an attack for significantly longer (23 periods instead of 4 periods in the unmitigated protocol). The findings of this study may be considered a further validation of the viability of identity-based consensus protocols, a research area highly relevant in an increasingly more digitalised and fragmented world.

Future Work. Given the findings around memory pool oversaturation, future research should consider the effects of mining mechanics more carefully, by including them in a more holistic model. Furthermore, future research should further develop and confirm these initial findings by increasing the robustness of the model by optimising the attack strategies employed by malicious actors. Similarly, future work should find a systematic approach to determining optimal parameters for system properties and mitigation strategies, specifically for the attenuation function used in temporal normalisation. Finally, rigorously formalising the problem in the form of a statistical model or through differential equations would provide a stronger basis for an argument in support of *IdAPoS'* viability.

References

[1] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, last accessed 16 December 2020 (2008).
URL `https://bitcoin.org/bitcoin.pdf`

[2] V. Buterin, Ethereum whitepaper, last accessed 25 December 2020 (2013).
URL `https://ethereum.org/en/whitepaper`

[3] W. Dai, b-money, last accessed 17 November 2020 (1998).
URL `http://www.weidai.com/bmoney.txt`

[4] C. Jarvis, Cypherpunk ideology: objectives, profiles, and influences (1992–1998), Internet Histories (2021) 1–27.

[5] M. Platt, R. J. Bandara, A.-E. Drăgnoiu, S. Krishnamoorthy, Information privacy in decentralized applications, in: M. H. U. Rehman, D. Svetinovic, K. Salah, E. Damiani (Eds.), Trust Models for Next-Generation Blockchain Ecosystems, EAI/Springer Innovations in Communication and Computing, Springer, 2021, In press.

[6] J. R. Douceur, The Sybil attack, in: P. Druschel, F. Kaashoek, A. Rowstron (Eds.), Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS 2002), Vol. 2429 of Lecture Notes in Computer Science, Springer, Cambridge, MA, USA, 2002, pp. 251–260.

[7] I. Weber, V. Gramoli, A. Ponomarev, M. Staples, R. Holz, A. B. Tran, P. Rimba, On availability for blockchain-based systems, in: 36th Symposium on Reliable Distributed Systems (SRDS), IEEE, Hong Kong, 2017, pp. 64–73.

[8] M. Platt, P. McBurney, Self-governing public decentralised systems, in: T. Groß, L. Viganò (Eds.), Proceedings of the 10th International Workshop on Socio-Technical Aspects in Security and Trust (STAST), Vol. 12812 of Lecture Notes in Computer Science, Springer, Guildford, UK, 2021, pp. 154–167.

[9] M. Platt, P. McBurney, IdAPoS–identity augmented proof-of-stake, Manuscript submitted for publication (2021).

[10] D. Hyland-Wood, S. Khatchadourian, A future history of international blockchain standards, The Journal of the British Blockchain Association 1 (1) (2018) 1–10.

[11] X. Fu, H. Wang, P. Shi, A survey of Blockchain consensus algorithms: mechanism, design and applications, Science China Information Sciences 64 (2).

[12] B. R. Sutherland, Blockchain's first consensus implementation is unsustainable, Joule 3 (4) (2019) 917–919.

[13] F. Leal, A. E. Chis, H. González–Vélez, Performance evaluation of private Ethereum networks, SN Computer Science 1 (5).

[14] A. Dhillon, G. Kotsialou, P. McBurney, L. Riley, Voting over a distributed ledger: An interdisciplinary perspective (aug 2020). `doi:10.31235/osf.io/34df5`.

[15] S. Park, M. Specter, N. Narula, R. L. Rivest, Going from bad to worse: from internet voting to blockchain voting, Journal of Cybersecurity 7 (1).

[16] Y. Hermstrüwer, Democratic blockchain design, Journal of Institutional and Theoretical Economics 175 (1) (2019) 163–177.

[17] T. Redshaw, Bitcoin beyond ambivalence, Thesis Eleven 138 (1) (2017) 46–64.

[18] C. Tozzi, Decentralizing democracy: approaches to consensus within blockchain communities, Teknokultura. Revista de Cultura Digital y Movimientos Sociales 16 (2) (2019) 181–195.

[19] C. Dwork, M. Naor, Pricing via processing or combatting junk mail, in: E. F. Brickell (Ed.), Proceedings of the 12th Annual International Cryptology Conference (CRYPTO' 92), Vol. 740 of Lecture Notes in Computer Science, Springer, Santa Barbara, CA, USA, 1992, pp. 139–147.

[20] A. Back, A partial hash collision based postage scheme, last accessed 19 December 2020 (Mar. 1997).
URL `http://www.hashcash.org/papers/announce.txt`

[21] P. McCorry, S. F. Shahandashti, F. Hao, A smart contract for boardroom voting with maximum voter privacy, in: A. Kiayias (Ed.), Proceedings of the 21st International Conference on Financial Cryptography and Data Security (FC 2017), Vol. 10322 of Lecture Notes in Computer Science, Springer, Sliema, Malta, 2017, pp. 357–375.

[22] S. Xiao, X. A. Wang, H. Wang, Large-scale electronic voting based on Conflux consensus mechanism, in: L. Barolli, F. Xhafa, O. K. Hussain (Eds.), Proceedings of the 13th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2019), Vol. 994 of Advances in Intelligent Systems and Computing, Springer, Sydney, Australia, 2019, pp. 291–299.

[23] S. King, S. Nadal, PPCoin: Peer-to-peer crypto-currency with proof-of-stake, last accessed 19 December 2020 (Aug. 2012).
URL `https://decred.org/research/king2012.pdf`

[24] QuantumMechanic, Proof of stake instead of proof of work, last accessed 22 May 2020 (Jul. 2011).
URL `https://bitcointalk.org/index.php?topic=27787.0`

[25] G. Fanti, L. Kogan, S. Oh, K. Ruan, P. Viswanath, G. Wang, Compounding of wealth in proof-of-stake cryptocurrencies, in: I. Goldberg, T. Moore (Eds.), Proceedings of the 23rd International Conference on Financial Cryptography and Data Security (FC 2019), Vol. 11598 of Lecture Notes in Computer Science, Springer, Frigate Bay, St. Kitts and Nevis, 2019, pp. 42–61.

[26] W. Wang, D. T. Hoang, P. Hu, Z. Xiong, D. Niyato, P. Wang, Y. Wen, D. I. Kim, A survey on consensus mechanisms and mining strategy management in blockchain networks, IEEE Access 7 (2019) 22328–22370.

[27] M. Borge, E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, B. Ford, Proof-of-personhood: Redemocratizing permissionless cryptocurrencies, in: Proceedings of the 2nd European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, Paris, France, 2017, pp. 23–26.

[28] E. Pournaras, Proof of witness presence: Blockchain consensus for augmented democracy in smart cities, Journal of Parallel and Distributed Computing 145 (2020) 160–175.

[29] M. T. de Oliveira, L. H. Reis, D. S. Medeiros, R. C. Carrano, S. D. Olabarriaga, D. M. Mattos, Blockchain reputation-based consensus: A scalable and resilient mechanism for distributed mistrusting applications, Computer Networks 179 (2020) 107367.

[30] P. Gaži, A. Kiayias, A. Russell, Stake-bleeding attacks on proof-of-stake blockchains, in: Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), IEEE, Zug, Switzerland, 2018, pp. 85–92.

[31] P. Daian, R. Pass, E. Shi, Snow White: Robustly reconfigurable consensus and applications to provably secure proof of stake, in: I. Goldberg, T. Moore (Eds.), Proceedings of the 23rd International Conference on Financial Cryptography and Data Security (FC 2019), Vol. 11598 of Lecture Notes in Computer Science, Springer, Frigate Bay, St. Kitts and Nevis, 2019, pp. 23–41.

[32] A. Kiayias, A. Russell, B. David, R. Oliynykov, Ouroboros: A provably secure proof-of-stake blockchain protocol, in: J. Katz, H. Shacham (Eds.), Proceedings of the 37th Annual International Cryptology Conference (CRYPTO 2017), Vol. 10401 of Lecture Notes in Computer Science, Springer, Santa Barbara, CA, USA, 2017, pp. 357–388.

[33] K. Liao, J. Katz, Incentivizing blockchain forks via whale transactions, in: M. Brenner, K. Rohloff, J. Bonneau, A. Miller, P. Y. Ryan, V. Teague, A. Bracciali, M. Sala, F. Pintore, M. Jakobsson (Eds.), Proceedings of the International Conference on Financial Cryptography and Data Security (FC 2017), Vol. 10323 of Lecture Notes in Computer Science, Springer, Sliema, Malta, 2017, pp. 264–279.

[34] W. Y. M. M. Thin, N. Dong, G. Bai, J. S. Dong, Formal analysis of a proof-of-stake blockchain, in: Proceedings of the 23rd International Conference on Engineering of Complex Computer Systems (ICECCS), IEEE, Melbourne, Australia, 2018, pp. 197–200.

[35] V. Buterin, V. Griffith, Casper the friendly finality gadget, arXiv e-prints (Oct. 2017). `arXiv:1710.09437`.

[36] W. Li, S. Andreina, J.-M. Bohli, G. Karame, Securing proof-of-stake blockchain protocols, in: J. Garcia-Alfaro, G. Navarro-Arribas, H. Hartenstein, J. Herrera-Joancomartí (Eds.), Proceedings of the International Workshops Data Privacy Management, Cryptocurrencies and Blockchain Technology (ESORICS 2017), Vol. 10436 of Lecture Notes in Computer Science, Springer, Oslo, Norway, 2017, pp. 297–315.

[37] S. Kanjalkar, J. Kuo, Y. Li, A. Miller, Short paper: I can't believe it's not stake! resource exhaustion attacks on PoS, in: I. Goldberg, T. Moore (Eds.), Proceedings of the 23rd International Conference on Financial Cryptography and Data Security (FC 2019), Vol. 11598 of Lecture Notes in Computer Science, Springer, Frigate Bay, St. Kitts and Nevis, 2019, pp. 62–69.

[38] M. Hesse, T. Teubner, Reputation portability – quo vadis?, Electronic Markets 30 (2) (2019) 331–349.

[39] V. Buterin, Chain interoperability, last accessed 25 December 2020 (Sep. 2016). URL `https://www.r3.com/wp-content/uploads/2017/06/chain_interoperability_r3.pdf`

[40] B. Pillai, K. Biswas, V. Muthukkumarasamy, Cross-chain interoperability among blockchain-based systems using transactions, The Knowledge Engineering Review 35.

[41] C. Decker, R. Wattenhofer, A fast and scalable payment network with Bitcoin duplex micropayment channels, in: A. Pelc, A. A. Schwarzmann (Eds.), Proceedings of the 17th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS), Vol. 9212 of Lecture Notes in Computer Science, Springer, Edmonton, AB, Canada, 2015, pp. 3–18.

[42] A. Zamyatin, D. Harz, J. Lind, P. Panayiotou, A. Gervais, W. Knottenbelt, XCLAIM: Trustless, interoperable, cryptocurrency-backed assets, in: Proceedings of the 2019 Symposium on Security and Privacy (SP), IEEE, San Francisco, CA, USA, 2019, pp. 193–210.

[43] P. Robinson, R. Ramesh, J. Brainard, S. Johnson, Atomic crosschain transactions white paper, arXiv e-prints (Feb. 2020). `arXiv:2003.00903`.

[44] M. Platt, F. Pierangeli, G. Livan, S. Righi, Facilitating the decentralised exchange of cryptocurrencies in an order-driven market, in: Proceedings of the 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS), IEEE, Paris, France, 2020, pp. 30–34.

[45] P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman, Reputation systems, Communications of the ACM 43 (12) (2000) 45–48.

[46] F. Irresberger, Coin concentration of proof-of-stake blockchains, Leeds University Business School Working Paper 19-04, Leeds University (2018). `doi:10.2139/ssrn.3293694`.

[47] S. C. Bankes, Agent-based modeling: A revolution?, Proceedings of the National Academy of Sciences 99 (Supplement 3) (2002) 7199–7200.

[48] M. Platt, Simulations of sybil attacks on identity-augmented proof-of-stake, Mendeley Data, V2 (2021). `doi:10.17632/3g8s2g69b3.2`.

[49] S. Jiang, J. Wu, Bitcoin mining with transaction fees: A game on the block size, in: 2019 International Conference on Blockchain, IEEE, Atlanta, GA, USA, 2019, pp. 107–115.

[50] I. Eyal, E. G. Sirer, Majority is not enough: Bitcoin mining is vulnerable, in: N. Christin, R. Safavi-Naini (Eds.), Proceedings of the 18th International Conference on Financial Cryptography and Data Security (FC 2014), Vol. 8437 of Lecture Notes in Computer Science, Springer, Christ Church, Barbados, 2014, pp. 436–454.

[51] L. Luu, J. Teutsch, R. Kulkarni, P. Saxena, Demystifying incentives in the consensus computer, in: Proceedings of the 22nd Conference on Computer and Communications Security (SIGSAC), ACM, Seoul, South Korea, 2015, pp. 706–719.