RePoPoW: A Gas-Efficient Superlight Bitcoin Client in Solidity

Anonymous Author(s)

ABSTRACT

Superlight clients enable the verification of proof-of-work-based blockchains by checking only a small representative number of block headers instead of all the block headers as done in SPV. Such clients can be embedded within other blockchains by implementing them as smart contracts, allowing for cross-chain verification. One such interesting instance is the consumption of Bitcoin data within Ethereum by implementing a Bitcoin superlight client in Solidity. While such theoretical constructions have demonstrated security and efficiency, no practical implementation exists. In this work, we put forth the first practical Solidity implementation of a superlight client which implements the NIPoPoW superblocks protocol. Contrary to previous work, our Solidity smart contract achieves sufficient gas-efficiency to allow a proof and counter-proof to fit within the gas limit of a block, making it practical. We provide extensive experimental measurements for gas. The optimizations that enable gas-efficiency heavily leverage a novel technique which we term hash-and-resubmit, which almost completely eliminates persistent storage requirements, the most expensive operation of smart contracts in terms of gas. Instead, the contract asks contesters to resubmit data and checks their veracity by hashing it. Other optimizations include off-chain manipulation of proofs in order to remove expensive look-up structures, and the usage of an optimistic schema. We show that such techniques can be used to bring down gas costs significantly and may have applications to other contracts. Lastly, our implementation allows us to calculate concrete cryptoeconomic parameters for the superblocks NIPoPoWs protocol and in particular to make recommendations about the monetary value of the collateral parameters. We provide such parameter recommendations over a variety of liveness and adversarial bound settings.

KEYWORDS

Blockchain; Superlight clients; NIPoPoWs, Solidity

1 INTRODUCTION

Blockchain interoperability [37] is the ability of distinct blockchains to communicate. This *crosschain* [20, 22, 23, 26, 36] communication enables useful features across blockchains such as the transfer of asset from one chain to another (one-way peg) and back (one-way peg) [26]. To date, there is no commonly accepted decentralized protocol that enables cross-chain transactions. Currently, crosschain operations are only available to the users via third-party applications, such as multi-currency wallets. However, this treatment is opposing to the nature of decentralized currencies.

In general, crosschain-enabled blockchains A, B support the following operations:

- Crosschain trading: A user with deposits in blockchain A, makes a payment to a user in blockchain B.
- Crosschain fund transfer: A user transfers her funds from blockchain A to blockchain B. After the transfer, these funds no longer exist in blockchain A. The user can later decide to transfer any portion of the original amount to the blockchain of origin.

In order to perform crosschain operations, there must be a mechanism to allow for users of blockchain A to discover events that occur in chain B, such that a transaction occurred. A trivial manner to perform such an audit is to participate as a full node in multiple chains. This approach, however, is impractical because a sizeable amount of storage is needed to host entire chains as they grow with time. As of July 2020, Bitcoin [30] chain spans roughly 245 GB, and Ethereum [5, 34] has exceeded 350 GB[1]. Naturally, not all users are able to accommodate this size of data, especially if portable devices are used, such as mobile phones.

One early solution to compress the extensive size of blockchain is addressed by Nakamoto [30] with the Simplified Payment Verification (SPV) protocol. In SPV, only the headers of blocks are stored, saving a considerable amount of storage. However, even with this protocol, the process of downloading and validating all block headers leads to unpleasant user experience. In Ethereum, for instance, headers sum up to approximately 5.1 GB[2] of data. A mobile client needs several minutes, even hours, to fetch all information needed in order to function as an SPV client.

Towards the goal of delivering more practical solutions for blockchain transaction verification, a new generation of *superlight* clients [4, 21, 24, 25] emerged. In these protocols, cryptographic proofs are generated, that prove the occurrence of events inside a blockchain. Better performance is achieved due to the considerably smaller size of proofs compared to the amount of data needed in SPV. By utilizing superlight client protocols, a compressed proof for an event in chain A is constructed and dispatched to chain B. Under the condition that chain B supports smart contracts, the proof is then verified automatically and transparently *on-chain*. This communication is realized without the intervention of third-party applications. An interesting application of such a protocol is the communication between Bitcoin and Ethereum.

[1]The size of the Bitcoin chain was derived from https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/, and the size of the Ethereum chain by https://etherscan.io/chartsync/chaindefaults

[2]Calculated as the number of blocks (10,050,219) times the size of header (508 bytes). Statistics by https://etherscan.io/

Related Work. We use Non-Interactive Proofs of Proof of Work (NIPoPoWs) [25, 26] as the fundamental building block of our solution. This cryptographic primitive is *provably secure* and provides *succinct proofs* regarding the existence of an arbitrary event in a chain. Contrary to the linear growth rate of the underlying blockchain, NIPoPoWs span polylogarithmic size of blocks.

Christoglou [10] provided a Solidity smart contract which is the first implementation of crosschain events verification based on NIPoPoWs, where proofs are submitted and checked for their validity. This solution, however, is impractical due to extensive usage of resources, widely exceeding the Ethereum block gas limit.

Other attempts have been done to address the verification of Bitcoin transactions to the Ethereum blockchain, most notably BTC Relay [9].

Our contribution. Notably, no practical implementation for superlight clients exists to date. In this paper, we focus on constructing a practical client for NIPoPoWs. For the implementation of our client, we refine the NIPoPoW protocol based on a series of keen observations. These refinements allow us to leverage useful techniques that construct a practical solution for proof verification. We believe that this achievement is a decisive step towards the establishment of NIPoPoWs to the application end, therefore it is a significant progress in order to provide a widely accepted protocol that enables crosschain transactions. A summary of our contributions in this paper is as follows:

(1) We developed the first decentralized client that securely verifies crosschain events and is practical. Our client establishes a trustless and efficient solution to the interoperability problem. We implement our client in Solidity, and we verify Bitcoin events to the Ethereum blockchain. The security assumptions we make are no others than SPV [19, 35].
(2) We present a novel pattern which we term *hash-and-resubmit*. Our pattern significantly improves performance of Ethereum smart contracts [5, 34] in terms of gas consumption by utilizing *calldata* space of Ethereum blockchain to eliminate high-cost storage operations.
(3) We create an *optimistic* schema which we incorporate into the design of our client. This design achieves significant performance improvement by replacing linear complexity verification of proofs, with constant complexity verification.
(4) We demonstrate via application that the NIPoPoW protocol is practical, making the cryptographic primitive the first provably secure construction of succinct proofs that is efficient to implement.
(5) We present an econometric analysis on NIPoPoWs. This analysis allows us to provide concrete values of collateral parameters in fiat, as well as the cost of submit and contest actions under different gas price configurations.

Our implementation meets the following requirements:

(1) Security: The client implements a provably secure protocol.
(2) Decentralization: The client is not dependent on third-party applications and operates in a transparent, decentralized manner.
(3) Efficiency: The client comply with all environmental constraints, i.e. block gas limit and calldata size limit of Ethereum blockchain.

We selected Bitcoin as source blockchain as it the most used cryptocurrency, and enabling crosschain transactions in Bitcoin is beneficial to the vast majority of blockchain community. We selected Ethereum as the target blockchain because, besides its popularity, it supports smart contracts, which is a requirement in order to perform on-chain verification.

Structure. In Section 2 we describe the blockchain technologies that are relevant to our work. In Section 3 we put forth the *hash-and-resubmit* pattern. We demonstrate the improved performance of smart contracts using the pattern, and how it is incorporated into our client. In Section 4, we present an alteration of the NIPoPoW protocol that enables the elimination of look-up structures that allows for efficient interactions due to the considerably smaller size of dispatched proofs. In Section 5, we put forth an optimistic schema that significantly lowers the complexity of proofs' structural verification from linear to constant, by introducing a new interaction which we term *dispute phase*. Furthermore, we present a technique that leverages the dispatch of constant number of blocks in contest phase. Finally, in Section 6, we present our cryptoeconomic analysis on RePoPoW. We establish the monetary value of collateral parameters, as well as the cost of each interaction with the client for different gas prices.

2 PRELIMINARIES

Model. We consider a setting where the blockchain network consists of two different types of nodes: The first kind, *full nodes*, are responsible for the maintenance of the chain including verifying it and mining new blocks. The second kind, *verifiers*, connect to full nodes and wish to learn facts about the blockchain without downloading it, for example whether a particular transaction is confirmed. The full nodes therefore also function as *provers* for the verifiers. Each verifier connects to multiple provers, at least one of which is assumed to be honest.

We model full nodes according to the Backbone model [16]. There are n full nodes, of which t are adversarial and $n-t$ are honest. All t adversarial parties are controlled by one colluding adversary $\mathcal{A}$. The parties have access to a hash function H which is modelled as a common Random Oracle [1]. To each novel query, the random oracle outputs κ bits of fresh randomness. Time is split into distinct *rounds* numbered by the integers $1, 2, \cdots$. Our treatment is in the *synchronous model*, so we assume messages *diffused* (broadcast) by an honest party at the end of a round are received by all honest parties at the beginning of the next round. This is equivalent to a network connectivity assumption in which the round

duration is taken to be the known time needed for a message to cross the diameter of the network. The adversary can inject messages, reorder them, sybil attack by creating multiple messages, but not suppress messages.

Blockchain. Each honest full node locally maintains a *chain* C, a sequence of blocks. In understanding that we are developing an improvement on top of SPV, we use the term *block* to mean what is typically referred to as a *block header*. Each block contains the Merkle Tree root [29] of transaction data $\overline{x}$, the hash s of the previous block in the chain known as the *previd*, as well as a nonce value ctr. As discussed in the Introduction, the compression of application data $\overline{x}$ is orthogonal to our goals in this paper and has been explored in independent work [8] which can be composed with ours. Each block $b = s \parallel \overline{x} \parallel ctr$ must satisfy the proof-of-work [12] equation $H(b) \leq T$ where T is a constant *target*, a small value signifying the difficulty of the proof-of-work problem. Our treatment is in the *static difficulty* case, so we assume that T is constant throughout the execution[3]. $H(B)$ is known as the *block id*.

Blockchains are finite block sequences obeying the *blockchain property*: that in every block in the chain there exists a pointer to its previous block. A chain is *anchored* if its first block is *genesis*, denoted $\mathcal{G}$, a special block known to all parties. This is the only node the verifier knows about when it boots up. For chain addressing we use Python brackets $\mathsf{C}[\cdot]$. A zero-based positive number in a bracket indicates the indexed block in the chain. A negative index indicates a block from the end, e.g., $\mathsf{C}[-1]$ is the tip of the blockchain. A range $\mathsf{C}[i{:}j]$ is a subarray starting from i (inclusive) to j (exclusive). Given chains $\mathsf{C}_1, \mathsf{C}_2$ and blocks A, Z we concatenate them as $\mathsf{C}_1\mathsf{C}_2$ or $\mathsf{C}_1 A$ (if clarity mandates it, we also use the symbol $\parallel$ for concatenation). Here, $\mathsf{C}_2[0]$ must point to $\mathsf{C}_1[-1]$ and A must point to $\mathsf{C}_1[-1]$. We denote $\mathsf{C}\{A{:}Z\}$ the subarray of the chain from block A (inclusive) to block Z (exclusive). We can omit blocks or indices from either side of the range to take the chain to the beginning or end respectively. As long as the blockchain property is maintained, we freely use the set operators $\cup$, $\cap$ and $\subseteq$ to denote operations between chains, implying that the appropriate blocks are selected and then placed in chronological order.

During every round, every party attempts to *mine* a new block on top of its currently adopted chain. Each party is given q queries to the random oracle which it uses in attempting to mine a new block. Therefore the adversary has tq queries per round while the honest parties have $(n-t)q$ queries per round. When an honest party discovers a new block, they extend their chain with it and broadcast the new chain. Upon receiving a new chain C' from the network, an honest party compares its length $|\mathsf{C}'|$ against its currently adopted chain C and adopts the newly received chain if it is longer. It is assumed that the honest parties control the majority of the computational power of the network. This *honest majority assumption* states that there is some δ such that $t < (1-\delta)(n-t)$. If so, the protocol ensures consensus among the honest parties: There is a constant k, the *Common Prefix* parameter, such that, at any round, all the chains belonging to honest parties share a common prefix of blocks; the chains can deviate only up to k blocks at the end of each chain [16]. Concretely, if at some round r two honest parties have C_1 and C_2 respectively, then either $\mathsf{C}_1[:-k]$ is a prefix of C_2 or vice versa.

Superblocks. Some valid blocks satisfy the proof-of-work equation better than required. If a block b satisfies $H(b) \leq 2^{-\mu}T$ for some natural number $\mu \in \mathbb{N}$ we say that b is a *μ-superblock* or a block *of level* μ. The probability of a new valid block achieving level μ is $2^{-\mu}$. The number of levels in the chain will be $\log|\mathsf{C}|$ with high probability [24]. Given a chain C, we denote $\mathsf{C}{\uparrow^{\mu}}$ the subset of μ-superblocks of C.

Non-Interactive Proofs of Proof-of-Work (NIPoPoW) protocols allow verifiers to learn the most recent k blocks of the blockchain adopted by an honest full node without downloading the whole chain. The challenge lies in building a verifier who can find the suffix of the longest chain between claims of both honest and adversarial provers, while not downloading all block headers. Towards that goal, the *superblock* approach uses superblocks as samples of proof-of-work. The prover sends superblocks to the verifier to convince them that proof-of-work has taken place without actually presenting all this proof-of-work. The protocol is parametrized by a constant security parameter m. The parameter determines how many superblocks will be sent by the prover to the verifier and security is proven with overwhelming probability in m.

Prover. The prover selects various levels μ and for each such level sends a carefully chosen portion of its μ-level *superchain* $\mathsf{C}{\uparrow^{\mu}}$ to the verifier. In standard blockchain protocols such as Bitcoin and Ethereum, each block $\mathsf{C}[i+1]$ in C points to its previous block $\mathsf{C}[i]$, but each μ-superblock $\mathsf{C}{\uparrow^{\mu}}[i+1]$ does not point to its previous μ-superblock $\mathsf{C}{\uparrow^{\mu}}[i]$. It is imperative that an adversarial prover does not reorder the blocks within a superchain, but the verifier cannot verify this unless each μ-superblock points to its most recently preceding μ-superblock. The proposal is therefore to *interlink* the chain by having each μ-superblock include an extra pointer to its most recently preceding μ-superblock. To ensure integrity, this pointer must be included in the block header and verified by proof-of-work. However, the miner does not know which level a candidate block will attain prior to mining it. For this purpose, each block is proposed to include a pointer to the most recently preceding μ-superblock, for every μ. As these levels are only $\log|\mathsf{C}|$, this only adds $\log|\mathsf{C}|$ extra pointers to each block header.

The exact NIPoPoW protocol works like this: The prover holds a full chain C. When the verifier requests a proof, the prover sends the last k blocks of their chain, the suffix $\chi = \mathsf{C}[-k:]$, in full. From the larger prefix $\mathsf{C}[:-k]$, the prover constructs a proof π by selecting certain superblocks as representative samples of the proof-of-work that took place. The blocks are picked as follows. The prover selects the *highest* level μ^* that has at least m blocks in it and

[3] A treatment of variable difficulty NIPoPoWs has been explored in the soft fork case [37], but we leave the treatment of velvet fork NIPoPoWs in the variable difficulty model for future work.

Algorithm 1 The Prove algorithm for the NIPoPoW protocol in a soft fork

```
function Prove_{m,k}(C)
    B ← C[0]                                          ▷ Genesis
    for μ = |C[−k − 1].interlink| down to 0 do
        α ← C[: −k]{B :}↑^μ
        π ← π ∪ α
        if m < |α| then
            B ← α[−m]
        end if
    end for
    χ ← C[−k :]
    return πχ
end function
```

includes all these blocks in their proof (if no such level exists, the chain is small and can be sent in full). The prover then iterates from level $\mu = \mu^* - 1$ down to 0. For every level μ, it includes sufficient μ-superblocks to cover the last m blocks of level $\mu + 1$, as illustrated in Algorithm 1. Because the density of blocks doubles as levels are descended, the proof will contain in expectation $2m$ blocks for each level below μ^*. As such, the total proof size $\pi\chi$ will be $\Theta(m \log |\mathsf{C}| + k)$. Such proofs that are polylogarithmic in the chain size constitute an exponential improvement over traditional SPV clients and are called *succinct*.

Verifier. Upon receiving two proofs $\pi_1\chi_1, \pi_2\chi_2$ of this form, the NIPoPoW verifier first checks that $|\chi_1| = |\chi_2| = k$ and that $\pi_1\chi_1$ and $\pi_2\chi_2$ form valid chains. To check that they are valid chains, the verifier ensures every block in the proof contains a pointer to its previous block inside the proof through either the *previd* pointer in the block header, or in the interlink vector. If any of these checks fail, the proof is rejected. It then compares π_1 against π_2 using the $\leq_m$ operator, which works as follows. It finds the lowest common ancestor block $b = (\pi_1 \cap \pi_2)[-1]$; that is, b is the most recent block shared among the two proofs. Subsequently, it chooses the level μ_1 for π_1 such that $|\pi_1\{b:\}\uparrow^{\mu_1}| \geq m$ (i.e., π_1 has at least m superblocks of level μ_1 following block b) and the value $2^{\mu_1}|\pi_1\{b:\}\uparrow^{\mu_1}|$ is maximized. It chooses a level μ_2 for π_2 in the same fashion. The two proofs are compared by checking whether $2^{\mu_1}|\pi_1\{b:\}\uparrow^{\mu_1}| \geq 2^{\mu_2}|\pi_2\{b:\}\uparrow^{\mu_2}|$ and the proof with the largest score is deemed the winner. The comparison is illustrated in Algorithm 2.

In the case of the infix proofs there are some additional things that need to be considered. An adversary prover could skip the blocks of interest and present an honest and longer chain that, if only the suffix verifier where to be used, is considered a better proof. For that reason, the last step of the algorithm in the suffix verifier is changed to not only store the best proof but also combine the two proofs by including all of the ancestor blocks of the losing proof. This is guaranteed to include the blocks of interest. The resulting best proof is stored as a DAG(Directed Acyclic Graph), as in Algorithm 3

Algorithm 2 The Verify algorithm for the NIPoPoW protocol

```
function best-arg_m(π, b)
    M ← {μ : |π↑^μ {b :}| ≥ m} ∪ {0}          ▷ Valid levels
    return max_{μ∈M}{2^μ · |π↑^μ {b :}|}      ▷ Score for level
end function
operator π_A ≥_m π_B
    b ← (π_A ∩ π_B)[−1]                                ▷ LCA
    return best-arg_m(π_A, b) ≥ best-arg_m(π_B, b)
end operator
function Verify^Q_{m,k}(P)
    π̃ ← (Gen)                          ▷ Trivial anchored blockchain
    for (π, χ) ∈ P do                  ▷ Examine each proof in P
        if validChain(πχ) ∧ |χ| = k ∧ π ≥_m π̃ then
            π̃ ← π
            χ̃ ← χ                              ▷ Update current best
        end if
    end for
    return Q̃(χ̃)
end function
```

Algorithm 3 The verify algorithm for the NIPoPoW infix protocol

```
function ancestors(B, blockById)
    if B = Gen then
        return {B}
    end if
    C ← ∅
    for id ∈ B.interlink do
        if id ∈ blockById then
            B′ ← blockById[id]
            ▷ Collect into DAG
            C ← C ∪ ancestors(B′, blockById)
        end if
    end for
    return C ∪ {B}
end function
function verify-infx^D_{ℓ,m,k}(P)
    blockById ← ∅
    for (π, χ) ∈ P do
        for B ∈ π do
            blockById[id(B)] ← B
        end for
    end for
    π̃ ← best π ∈ P according to suffix verifier
    return D(ancestors(π̃[−1], blockById))
end function
```

3 THE HASH-AND-RESUBMIT PATTERN

We now introduce a novel design pattern for Solidity smart contracts that results into significant gas optimization due to the elimination of expensive storage operations. We first

introduce the pattern, and display how smart contracts benefit by using it. Then, we proceed into integrating the pattern into the NIPoPoW client, and we analyze the performance in comparison with the previous work [10].

Motivation. It is essential for smart contracts to store data in the blockchain. However, interacting with the storage of a contract is among the most expensive operations of the EVM [5, 34]. Therefore, only necessary data should be stored and redundancy should be avoided when possible. This is contrary to conventional software architecture, where storage is considered cheap. Usually, the performance of data access in traditional systems is related with time. In Ethereum, however, performance is related to gas consumption. Access to persistent data costs a substantial amount of gas, which has a direct monetary value. One way to mitigate gas cost of reading variables from the blockchain is to declare them public. This leads to the creation of a *getter* function in the background, allowing free access to the value of the variable. But this treatment does not prevent the initial population of storage data, which is significantly expensive for large size of data.

By using the *hash-and-resubmit* pattern, large structures are omitted from storage entirety, and are contained in memory. When a function call is performed, the signature and arguments of the function is included in the transactions field of the body of a block. The contents of blocks are public to the network, therefore this information is locally available to full nodes. By simply observing blocks, a node retrieves data sent to the contract by other users. To interact publicly with this data without the utilization storage, the node *resends* the observed data to the blockchain. The concept of resending data is redundant in conventional systems. However, this technique is very efficient to use in Solidity due to the significantly lower gas cost of memory operations in relation with storage operations.

Related patterns. Towards the goal of implementing gas-efficient smart contracts [6, 7, 15, 17], several patterns have been proposed. Towards eliminating storage operations using data signatures, the utilization of IPFS [2] is proposed by [31] and [18]. However, these solutions do not address availability, which is a main requirement in our application. [11] uses logs to replace storage in a similar manner sparing a great amount of gas. However, this approach does not address consistency, which is consists also a critical one of our critical target. Lastly, [33] proposes an efficient manner to replace read storage operations, but does not address write operations.

Applicability. We now list the cases in which the *hash-and-resubmit* pattern is efficient to use:

The situations in which the *hash-and-resubmit* pattern can be applied are the following:

(1) The application is a Solidity smart contract.
(2) Read/write operations are performed in large arrays that exist in storage. Rehashing variables of small size may result to negligible gain or even performance loss.
(3) The entity that operates on the structures is a full node and observes function calls to the smart contract.

Participants and collaborators. The first participant is the smart contract S that accepts function calls. Another participant is the invoker E_1, who dispatches a large array d_0 to S via a function $\mathsf{func}_1(\mathsf{d}_0)$. Note that d_0 is potentially processed in func_1, resulting to d. The last participant is the observer E_2, who is a full node that observes transactions towards S in the blockchain. This is possible because nodes maintain the blockchain locally. After observation, E_2 retrieves data d. Since this is an off-chain operation, a malicious E_2 potentially alters d before interacting with S. We denote the potentially modified d as d^*. Finally, E_2 acts as an invoker by making a new call to S, $\mathsf{func}_2(\mathsf{d}^*)$. The verification that $\mathsf{d} = \mathsf{d}^*$, which is a prerequisite for the secure functionality of the underlying contract consists a part of the pattern and is performed in func_2.

Implementation. The implementation of this pattern is divided in two parts. The first part covers how d^* is retrieved by E_2, whereas in the second part the verification of $\mathsf{d} = \mathsf{d}^*$ is realized. The challenge here is twofold:

(1) Availability: E_2 must be able to retrieve d without the need of accessing on-chain data.
(2) Consistency: E_2 must be prevented from dispatching d^* that differs from the originally submitted d.

Hash-and-resubmit technique is performed in two stages to face these challenges: (a) the *hash* phase, which addresses *consistency*, and (b) the *resubmit* phase which addresses *availability* and *consistency*.

Addressing availability: During *hash* phase, E_1 makes the function call $\mathsf{func}_1(\mathsf{d}_0)$. This transaction, which includes a function signature (func_1) and the corresponding data (d_0), is added in a block by a miner. Due to blockchain's transparency, the observer of func_1, E_2, retrieves a copy of d_0, without the need of accessing contract data. In turn, E_2 performs *locally* the same set of on-chain instructions operated on d_0 generating d. Thus, availability is addressed through observability.

Addressing consistency We prevent an adversary E_2 from altering d^* by storing the *signature* of d in contract's state during the execution of $\mathsf{func}_1(\mathsf{d})$ by E_1. In the context of Solidity, a signature of a structure is the digest of the structure's *hash*. The pre-compiled `sha256` is convenient to use in Solidity, however we can make use of any cryptographic hash function $\mathsf{H}()$:

$$\mathsf{hash} \leftarrow \mathsf{H}(\mathsf{d})$$

Then, in *rehash* phase, the verification is performed by comparing the stored digest of d with the digest of d^*.

$$\mathsf{require}(\mathsf{hash} = \mathtt{H}(\mathsf{d}^*))$$

In Solidity, the size of digests is 32 bytes. To persist such a small value in contract's memory only adds a constant, negligible cost overhead.

We illustrate the application of the *hash-and-resubmit* pattern in Figure 1.

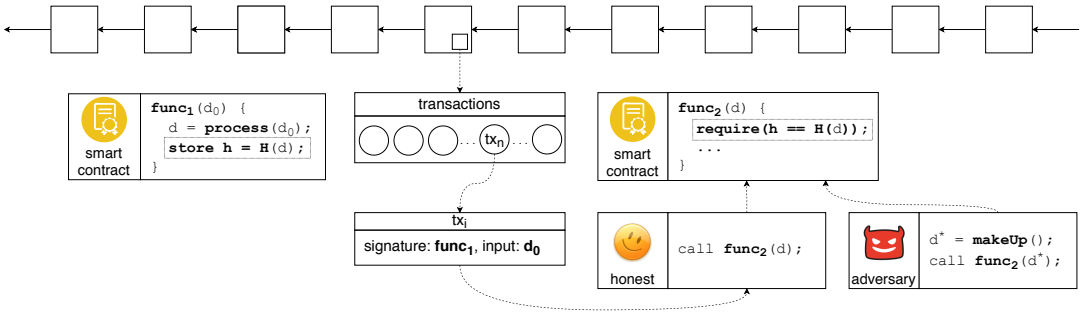


Figure 1: The *hash-and-resubmit* pattern. First, an invoker calls $\mathsf{func}_1(\mathsf{d}_0)$. d_0 is processed *on-chain* and d is generated. The signature of d is stored in the blockchain as the digest of a hash function $\mathsf{H}(\mathsf{d})$. Then, a full node that observes invocations of func_1 retrieves d_0, and generates d by performing the analogous processing on d_0 *off-chain*. An adversarial observer dispatches d^*, where $\mathsf{d}^* \neq \mathsf{d}$. Finally, the node invokes $\mathsf{func}_2(.)$. In func_2, the validation of input data is performed, reverting the function call if the signatures of the input does not match with the signature of the originally processed data. By applying a *hash-and-resubmit pattern*, only the fixed-size signature of d is stored to the contract's state, replacing arbitrarily large structures.

Sample. We now demonstrate the usage of the hash-and-resubmit pattern with a simplistic example. We create a smart contract that orchestrates a game between two players, P_1 and P_2. The winner is the player with the most valuable array. The interaction between players through the smart contract is realized in two phases: (a) the submit phase and (b) the contest phase.

Submit phase: P_1 submits an N-sized array, a_1 and becomes the holder of the contract.

Contest phase: P_2 submits a_2. If compare(a_2, a_1) is true, then the holder of the contract changes to P_2. We provide a simple implementation for compare(), but we can consider any notion of comparison, since the pattern is abstracted from such implementation details.

We make use of the *hash-and-resubmit* pattern by prompting P_2 to provide *two* arrays to the contract during contest phase: (a) a_1^*, which is the originally submitted array by P_1, possibly modified by P_2, and (b) a_2, which is the contesting array.

We provide two implementations of the above described game. In Algorithm 4 we display the storage implementation, while in Algorithm 5 we show the implementation embedding the *hash-and-resubmit* pattern.

Gas analysis. The gas consumption of the two above implementations is displayed in Figure 2. By using the *hash-and-resubmit* pattern, the aggregated gas consumption for submit and contest is decreased by 95%. This significantly affects the efficiency and applicability of the contract. Note that, the storage implementation exceeds the Ethereum block gas limit[4] for arrays of size 500 and above, contrary to the optimized version, which consumes approximately only $1/10^{th}$ of the block gas limit for arrays of 1000 elements.

[4] As of July 2020, the Ethereum block gas limit approximates 10,000,000 gas units

Algorithm 4 best array using storage

```
1:  contract best-array
2:      best ← ∅; holder ← ∅
3:      function submit(a)
4:          best ← a                          ▷ array saved in storage
5:          holder ← msg.sender
6:      end function
7:      function contest(a)
8:          require(compare(a))
9:          holder ← msg.sender
10:     end function
11:     function compare(a)
12:         require(|a| ≥ |best|)
13:         for i in |best| do
14:             require(a[i] ≥ best[i])
15:         end for
16:         return true
17:     end function
18: end contract
```

Consequences. The most obvious consequence of applying the *hash-and-resubmit* pattern variations is the circumvention of storage structures, a benefit that saves a substantial amount of gas, especially in the cases where these structures are large. To that extend, smart contracts that exceed the Ethereum block gas limit become practical. Furthermore, the pattern enables off-chain transactions, significantly improving the performance of smart contracts.

Known uses. To our knowledge, we are the first to combine the notion of the transparency of the blockchain with data structures signatures to eliminate storage variables from Solidity smart contracts by resubmitting data in a manner that addresses consistency and availability.

Enabling NIPoPoWs. We now present how the *hash-and-resubmit* pattern is used in the context of the NIPoPoW

Algorithm 5 best array using hash-and-resubmit pattern

```
1:  contract best-array
2:      hash ← ∅; holder ← ∅
3:      function submit(a1)
4:          hash ← H(a1)                          ▷ hash saved in storage
5:          holder ← msg.sender
6:      end function
7:      function contest(a1*, a2)
8:          require(hash256(a1*) = hash)          ▷ validate a1*
9:          require(compare(a1*, a2))
10:         holder ← msg.sender
11:     end function
12:     function compare(a1*, a2)
13:         require(|a1*| ≥ |a2|)
14:         for i in |a1*| do
15:             require(a1*[i] ≥ a2[i])
16:         end for
17:     end function
18:     return true
19: end contract
```

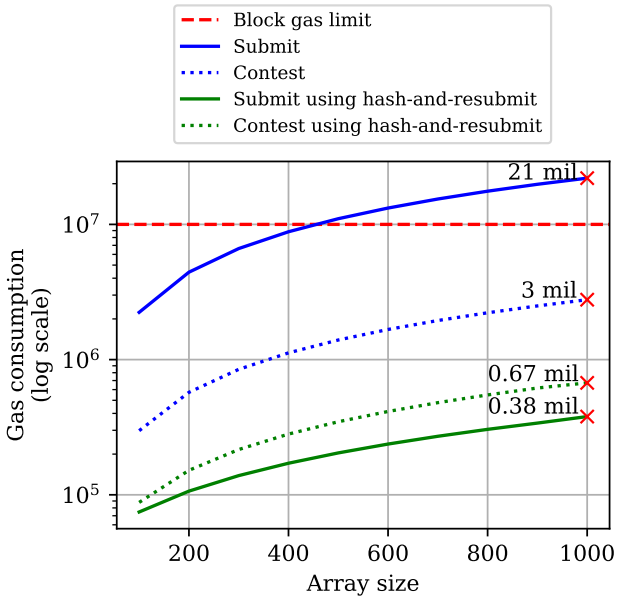


Figure 2: Gas-cost reduction using the *hash-and-resubmit* pattern (lower is better). By avoiding gas-heavy storage operations, the aggregated cost of submit and contest is decreased significantly by 95%.

superlight client. The NIPoPoW verifier adheres to a submit-and-contest-phase schema, and the inputs of the functions are arrays that are processed on-chain. It consists, therefore, a suitable case for our pattern.

In *submit* phase, a *proof* is submitted. In the case of falsity, it is contested by another user in *contest* phase. The user that initiates the contest is a node and monitors the traffic of the verifier. The input of submit function includes the submit proof (π_s) that indicates the occurrence of an *event* (e) in the source chain, and the input of contest function includes the contesting proof (π_c). A successful contest of π_s is realized when π_c has a better score. The score evaluation process is irrelevant to the pattern and remains unchained. The size of proofs is dictated by the value m. We consider $m = 15$ to be sufficiently secure.

In previous work, NIPoPoW proofs are maintained on-chain, resulting to extensive storage operations that limit the applicability of the contract considerably. In our implementation, proofs are not stored on-chain, and π_s is retrieved by the node from calldata. Since we assume a trustless network, π_s potentially alters by the node. We denote the potentially changed π_s as π_s^*. In *contest* phase, π_s^* and π_c are dispatched in order to enable the *hash-and-resubmit* pattern.

For our analysis, we create a chain similar to the Bitcoin chain with the addition of the interlink structure in each block as in [10]. Our chain spans 650,000 blocks, which represent a slightly enhanced Bitcoin chain[5]. From the tip of our chain, we branch two chains that span 100 and 200 additional blocks respectively, as illustrated in Figure 3. Then, we use the smaller chain to create π_s, , and the larger chain to create π_c. By applying the protocol, π_s is submitted, and a contest is initiated with π_c. The contest is successful, since π_s represents a chain consisting of fewer blocks than π_c, therefore encapsulating less proof of work. We select this setting as it provides maximum code coverage, and it descries the most gas-heavy scenario.

In Algorithm 6 we show how *hash-and-resubmit* pattern is embedded into the NIPoPoW client.

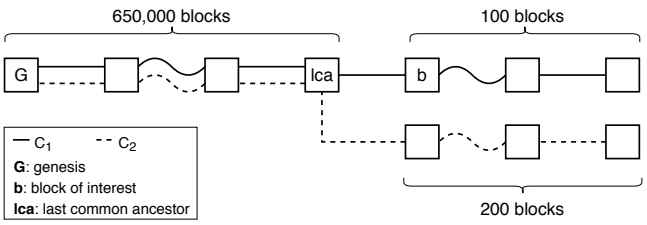


Figure 3: Forked chains for our gas analysis.

In Figure 4, we display how results of the *hash-and-resubmit* implementation differentiate from previous work for the aggregated cost of *submit* and *contest* phases. We observe that, by using the *hash-and-resubmit* pattern, we achieve to increase the performance of the contract 50%. This is a decisive step towards creating a practical superlight client.

Note that, the gas consumption generally follows an ascending trend, however it is not a monotonically increasing function. This is due to the fact that NIPoPoWs are probabilistic structures, the size of which is determined by the distribution of superblocks within the underlying chain. A

[5]Bitcoin spans 631,056 blocks as on May 2020. Metrics by https://www.blockchain.com/

proof that is constructed for a chain of a certain size can be larger than a proof constructed for a slightly smaller chain, resulting to non-monotonic increase of gas consumption between consecutive values of proof sizes.

Algorithm 6 The NIPoPoW client using hash-and-resubmit pattern

```
1:  contract crosschain
2:      events ← ⊥; 𝒢 ← ⊥
3:      function initialize(𝒢_remote)
4:          𝒢 ← 𝒢_remote
5:      end function
6:      function submit(π_s, e)
7:          require(events[e] = ⊥)
8:          require(π_s[0] = 𝒢)
9:          require(valid-interlinks(π))
10:         DAG ← DAG ∪ π_s
11:         ancestors ← find-ancestors(DAG, π_s[-1])
12:         require(evaluate-predicate(ancestors, e))
13:         ancestors = ⊥
14:         events[e].hash ← H(π_s)              ▷ enable pattern
15:     end function
16:     function contest(π*_s, π_c, e)           ▷ provide proofs
17:         require(events[e] ≠ ⊥)
18:         require(events[e].hash = H(π*_s))    ▷ verify π*_s
19:         require(π_c[0] = 𝒢)
20:         require(valid-interlinks(π_cont))
21:         lca = find-lca(π*_s, π_c)
22:         require(π_c ≥_m π*_s)
23:         DAG ← DAG ∪ π_c
24:         ancestors ← find-ancestors(DAG, π*_s[-1])
25:         require(¬evaluate-predicate(ancestors, e))
26:         ancestors = ⊥
27:         events[e] ← ⊥
28:     end function
29: end contract
```

4 REMOVING LOOK-UP STRUCTURES

Now that we can freely retrieve large array structures, we can focus on other types of storage variables. The challenge we face is that the protocol of NIPoPoWs dependents on a Directed Acyclic Graph (DAG) of blocks which is a mutable hashmap. This DAG is needed because interlinks of superblocks can be adversarially defined. By using DAG, the set of ancestor blocks of a block is extracted by performing a simple graph search. For the evaluation of the predicate, the set of *ancestors* of the best blockchain tip is used. Ancestors are included to avoid an adversary who presents an honest chain but skips the blocks of interest.

This logic is intuitive and efficient to implement in most traditional programming languages such as C++, JAVA, Python, JavaScript, etc. However, as our analysis demonstrates, such an implementation in Solidity is significantly

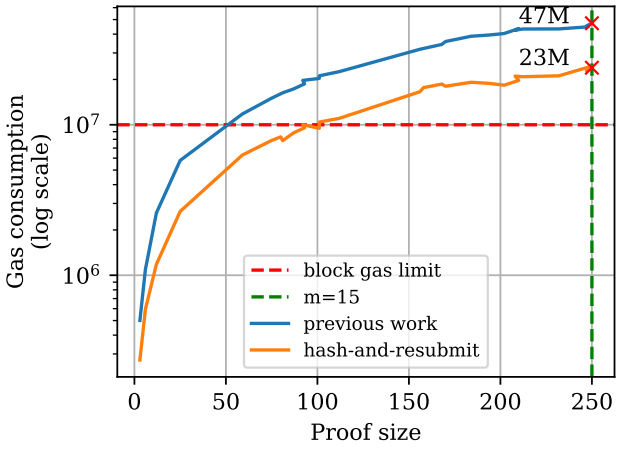


Figure 4: Performance improvement using hash-and-resubmit pattern in NIPoPoWs compared to previous work for a secure value of m (lower is better). The gas consumption is decreased by approximately 50%.

expensive. Albeit Solidity supports constant-time look-up structures, hashmaps are only contained in storage. This affects the performance of the client, especially for large proofs.

We make a keen observation regarding potential positions of the *block of interest* in proofs, which leads us to the construction of an architecture that does not require DAG, ancestors or other complementary structures. To support this claim, we adopt the notation from [25]. Additionally, we denote the initially submitted proof as π_s and the contesting proof as π_c. In this context, we consider the predicate p to be of the fork: "does block B exist inside proof π?", where B denotes the block of interest of proof π. The entity that performs the submission is $\mathsf{E_s}$, and the entity that initiates a contest is $\mathsf{E_c}$.

Position of block of interest. NIPoPoWs are sets of sampled interlinked blocks, which means that they can be perceived as chains. If π_1 differs from π_2, then a fork is created at the index of the last common ancestor (LCA). The block of interest lies at a certain index within π_s and indicates a stable predicate [25, 28] that is true for π_s. A submission in which B is absent from π_s is aimless, because it automatically fails since no element of π_s satisfies p. On the contrary, π_c tries to prove the *falseness* of the underlying predicate. This means that, if the block of interest is included in π_c, then the contest is aimless. We freely use the term aimless to also characterize components that are included in such actions i.e. aimless proof, aimless blocks etc. We use the term meaningful to describe non-aimless actions and components.

In the NIPoPoW protocol, proofs' segments $\pi_s\{:\mathsf{LCA}\}$ and $\pi_c\{:\mathsf{LCA}\}$ are merged to prevent adversaries from skipping or adding blocks, and the predicate is evaluated against $\pi_s\{:\mathsf{LCA}\} \cup \pi_c\{:\mathsf{LCA}\}$. We observe that $\pi_c\{:\mathsf{LCA}\}$ can be

omitted, because no block B exists such that $\{\mathsf{B} : \mathsf{B} \notin \pi_s\{:\mathsf{LCA}\} \wedge \mathsf{B} \in \pi_c\{:\mathsf{LCA}\}\}$ where B results into positive evaluation of the predicate. This is due to the fact that, in a meaningful contest, B is not included in π_c. Consequently, π_c is only meaningful if it forks π_s at a block that is prior to B.

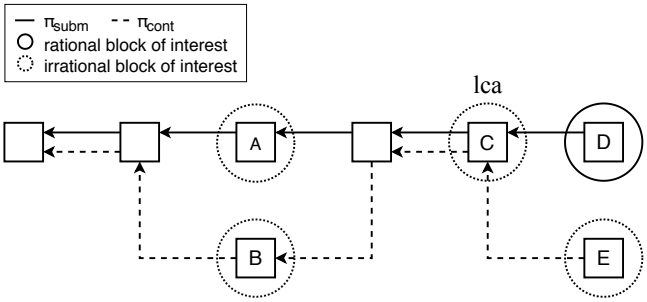


Figure 5: Fork of two chains. Solid lines connect blocks of π_s, and dashed lines connect blocks of π_{cont}. In this configuration, blocks in dashed circles are aimless blocks of interest, and the block in the solid circle is a meaningful block of interest. Blocks B, C and E are aimless because they exist in π_c. Block A is aimless because it belongs to the subchain $\pi_s\{:\mathsf{LCA}\}$

In Figure 5 we display a fork of two proofs. Solid lines connect blocks of π_s and dashed lines connect blocks of π_c. By examining which scenarios are rational depending on different positions of the block of interest, we observe that blocks `B`, `C` and `E` do not qualify, because they are included only in π_c. Block `A` is included in $\pi_s\{:\mathsf{LCA}\}$, which means that π_c is an irrational contest because the LCA comes after B. Therefore A is an irrational block as a component of an irrational contest. Given this configuration, the only rational block of interest is `D` and its predecessors.

Minimal forks. By combining the above observations, we derive that, π_c can be truncated into $\pi_c\{:\mathsf{LCA}\}$ without affecting the correctness of the protocol. We term this truncated proof π_c^f. Security is preserved by requiring π_c^f to be a *minimal fork* of π_s. A minimal fork is a fork chain that shares exactly one common block with the main chain. A proof $\tilde{\pi}$, which is minimal fork of π, has the following attributes:

(1) $\pi\{lca\} = \tilde{\pi}[0]$
(2) $\pi\{lca:\} \cap \tilde{\pi}[1:] = \emptyset$

By requiring π_c^f to be a minimal fork of π_s, we prevent adversaries from dispatching an augmented π_c^f to claim better score against π_s. Such an attempt is displayed in Figure 6.

In Algorithm 7, we show how minimal fork technique is incorporated into our client replacing the DAG and ancestors. In Figure 7 we show how the performance of the client improves. We use the same test case as in *hash-and-resubmit*.

By applying the minimal-fork technique, he achieve to decrease gas consumption by 55%: *submit* phase costs 4,700,000 gas, and *contest* phase costs 4,900,000 million gas. This is a notable result, since each phase is now below the block gas limit.

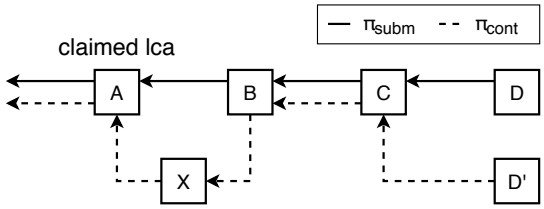


Figure 6: An adversary attests to contest with a malformed proof. Adversary proof consists of blocks {A, X, B, C, D'} that achieves better score against submit proof {A, B, C, D}. This attempt is rejected due to the minimal-fork requirement.

Algorithm 7 The NIPoPoW client using the minimal fork technique

```
1:  contract crosschain
2:      ...
3:      function submit(π_s, e)
4:          require(π_s[0] = G)
5:          require(events[e] = ⊥)
6:          require(valid-interlinks(π_s))
7:          require(evaluate-predicate(π_s, e))
8:          events[e].hash ← H(π_s)
9:      end function
10:     function contest(π_s*, π_c^f, e, f)        ▷ f: Fork index
11:         require(events[e] ≠ ⊥)
12:         require(events[e].hash = H(π_s*))
13:         require(valid-interlinks(π_c^f))
14:         require(minimal-fork(π_s*, π_c^f, f))   ▷ Minimal fork
15:         require(π_c^f ≥_m π_s*)
16:         require(¬evaluate-predicate(π_c^f, e))
17:         events[e] ← ⊥
18:     end function
19:     function minimal-fork(π_1, π_2, f)
20:         if π_1[f] ≠ π_2[0] then                 ▷ Check fork head
21:             return false
22:         end if
23:         for b_1 in π_1[f+1:] do                 ▷ Check disjoint proofs
24:             if b_2 in π_2[1:] then
25:                 return false
26:             end if
27:         end for
28:         return true
29:     end function
30: end contract
```

5 PROCESSING FEWER BLOCKS

Leveraging an optimistic schema. In smart contracts, in order to ensure that users comply with the underlying application's rules, certain actions need to be performed on-chain, e.g. verification of data, balance checks etc. In a different approach, actions that deviate from the protocol are reverted

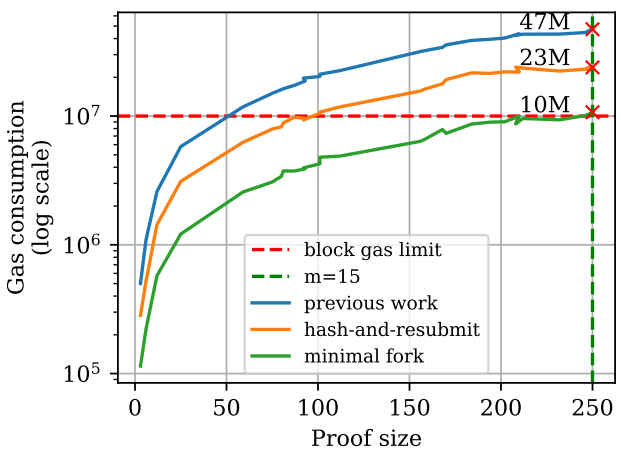


Figure 7: Performance improvement using minimal fork (lower is better). The gas consumption is decreased by approximately 55%.

Process	Gas cost	Total %
verify-interlink	2,200,000	53%
submit	4,700,000	100%

Table 1: Gas usage of function verify-interlink compared to overall gas consumption of submit.

	Phase	Gas		Phase	Gas	Phase	Gas
	submit	4.7		submit	2.2	submit	2.2
	contest	4.9		dispute	1.3	contest	4.9
I.	**Total**	**9.6**	**II.**	**Total**	**3.5**	**Total**	**7.1**

Table 2: Performance per phase. Gas units are displayed in millions. I: Gas consumption prior to dispute phase incorporation. II: Gas consumption for two independent sets of interactions (submit & dispute), (submit & contest).

after honest users indicate them, not allowing diverging entities to gain advantages. Such applications that do not check the validity of actions by default, but rather depend on the intervention of other users are characterized "optimistic". In the Ethereum community, several projects(refs) have been emerged that incorporate the notion of *optimistic interactions*. We observe that such a schema can be embedded in the NIPoPoW protocol, resulting to significant performance gain.

We discussed how the verification in the NIPoPoW protocol is realized in two phases. In *submit* phase, the verification of the π_s is performed. This is necessary in order to prevent adversaries from injecting blocks that do not belong to the chain, or changing existing blocks. A proof is valid for submission if it is *structurally correct*. Correctly structured NIPoPoWs has the following requirements:

(1) The first block of the proof is the genesis block of the underlying blockchain.
(2) Every block has a valid interlink.

Asserting the existence of genesis in the first index of a proof is an inexpensive operation of constant complexity. However, confirming the interlink correctness of all blocks is a process of linear complexity to the size of the proof. Albeit the verification is performed in memory, sufficiently large proofs result into costly submissions. This consists the most demanding function of *submit* phase. In Table 1 we display the cost of valid-interlink function which determines the structural correctness of a proof in comparison with the overall gas used in submit.

Dispute phase. We observe that the addition of a phase in our protocol alleviates the burden of verifying all elements of the proof by enabling the indication of an individual incorrect block. This phase, which we term *dispute* phase, leverages selective verification of the submitted proof at a certain index, which, as a constant operation, significantly reduces the gas cost of the verification process.

In the NIPoPoW protocol, when an proof π_s is submitted by E_s, it is retrieved by a node E_c from the calldata and the proof is checked for its validity *off-chain*. The critical observation we make here, is that, in order to prove that π_s is structurally invalid, E_c only needs to indicate the index in which π_s fails the interlink verification. In the protocol that incorporates *dispute* phase, E_c calls dispute(π_s^*, i) for a structurally incorrect proof, where i indicates the disputing index of π_s^*. Therefore, only one block is interpreted *on-chain* rather than the entire span of π_s^*.

Note that, this additional phase does not imply increased rounds of interactions between E_s and E_c. In the case where π_s is invalidated by *dispute* phase, *contest* phase is skipped. Similarly, in the case in which π_s is structurally correct, but represents a dishonest chain, E_c proceeds directly to *contest* phase.

In Table 2 we display the gas consumption for two independent cycles of interactions:

(1) Phases *submit* and *dispute* for the case in which π_s is structurally incorrect.
(2) Phases *submit* and *contest* for the case in which π_s is structurally correct, but represents a dishonest chain.

In Algorithm 8, we show the implementation of *dispute* phase with submit and valid-single-interlink functions.

Isolating the best level. As discussed, *dispute* and *contest* phases are mutually exclusive. Unfortunately, constant-time verification cannot be applied in a contest without increasing the rounds of interactions for the users. However, we derive a major optimization for *contest* phase by observing the process of score evaluation.

In NIPoPoWs, after the last common ancestor is found, each fork of the proofs is evaluated in terms of proof of work score. Each level encapsulates different score of proof of work, and the level with the best score is the representative of the

underlying proof. Since the common blocks of the two proofs naturally gather the same score, only the disjoint portions need to be addressed. Consequently, the position of the LCA determines the span of the proofs that will be included in the score evaluation process. Furthermore, it is impossible to determine the score of a proof in *contest* phase due to the lack of knowledge of the LCA block.

When π_s is known, the score of both proofs is calculated. This means that, after π_s is retrieved from the calldata, the scores of π_s and π_c is known to E_c, as well as the level in which each proof encapsulates the most proof of work. We illustrate the blocks that participate in the formulation of of a proof's score in Figure 8. In the light of this observation, E_c only submits the blocks which consist the *best level* of π_c. The number of these blocks is constant, as it is determined by the security parameter m, which is irrelevant to the size of the underlying blockchain.

The calculation of best level of π_c is an *off-chain* process. Naturally, an adversarial E_c can intentionally dispatch a level of π_c which is different than the best level. However, this is an irrational action, since different levels only undermine the score of π_c. On the contrary, due to the consistency property of *hash-and-resumbit*, π_s cannot be altered. We denote the best level of π_c^f as $\pi_c^{f,\uparrow^b}$.

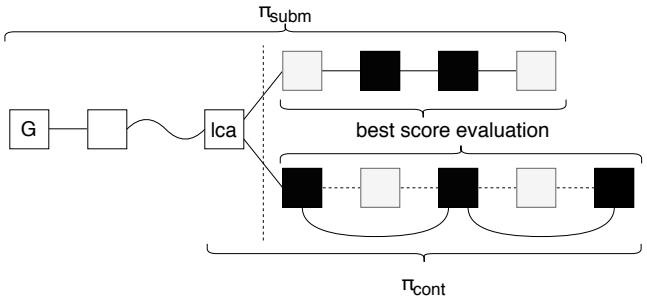


Figure 8: Fork of two proofs. Colored blocks after the lca determine the score of each proof. Black blocks belong to the level that has the best score. Only black blocks are part of the best level of the contesting proof.

The utilization of *best-level* methodology greatly increases the performance of the client, because the complexity of the majority of contest functions is related to the size of π_c. In Table 3, we demonstrate the difference in gas consumption in *contest* phase after using *best-level*. The performance of most functions is increased by approximately 85%. This is due to the fact that the size of π_c is decreased accordingly. For $m = 15$, $\pi_c^{f,\uparrow^b}$ consists of 31 blocks, while π_c^f consists of 200 blocks. Notably, the calculation of score for $\pi_c^{f,\uparrow^b}$ needs 97% less gas. We achieve such a discrepancy because the process of score calculation for multiple levels demands the use of a temporary hashmap which is a storage structure. In contrast, the evaluation of score of a individual level which is performed entirely in memory.

Process		Gas Cost ($\times 10^3$)	Total		Gas Cost ($\times 10^3$)	Total
valid-interlinks		900	18%		120	10%
minimal-fork		1,900	39%		275	18%
args (π_s)		750	16%		750	51%
args (π_c)		950	19%		20	1%
other		400	8%		300	20%
contest	**I**	4,900	100%	**II**	1,465	100%

Table 3: Gas usage in contest. I: Before utilizing best level. II: After utilizing best level.

In Figure 9, we illustrate the performance gain of the client using *dispute* phase and best-level contesting proof. The aggregated gas consumption of *submit* and *contest* phases is greatly reduced to 3,500,000 gas. This is critical threshold regarding applicability of the contract, since a cycle of interactions now fits effortlessly inside a single Ethereum block.

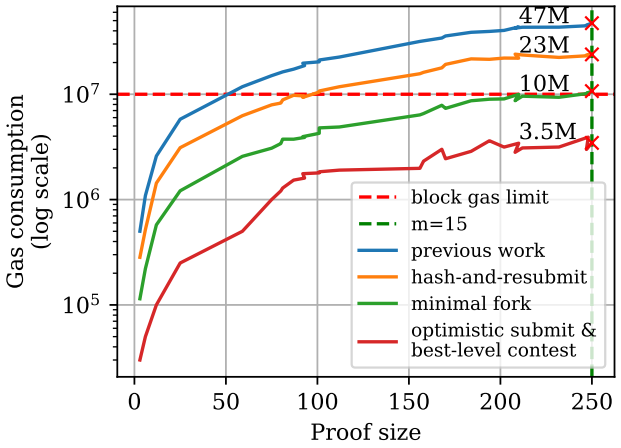


Figure 9: Performance improvement using optimistic schema in submit phase and best level proof in contesting proof (lower is better). The gas consumption is decreased by approximately 65%.

6 CRYPTOECONOMICS

Fairness. We now present our economical analysis on RePoPoW. We have already discussed that the NIPoPoW protocol is performed in distinct phases. In each phase, different entities are prompted to act. As in SPV, the security assumption that is made is that at least one honest node is connected to the verifier contract and serves honest proofs. However, the process of contesting a submitted proof by a honest node does not come without expenses. Such an expense is the computational power a node has to consume in order to fetch a submitted proof from the calldata and construct a contesting proof, but most importantly, the gas

Algorithm 8 The NIPoPoW client enhanced with dispute phase and best-level contesting

```
1: contract crosschain
2:     ...
3:     function submit(π_s, e)
4:         require(π_s[0] = G)
5:         require(events[e] = ⊥)
6:         require(evaluate-predicate(π_s, e))
7:         events[e].hash ← H(π_s)
8:     end function
9:     function dispute(π*_s, e, i)            ▷ i: Dispute index
10:        require(events[e] ≠ ⊥)
11:        require(events[e].hash = H(π*_s))
12:        require(¬valid-single-interlink(π_s, i))
13:        events[e] ← ⊥
14:    end function
15:    function valid-single-interlink(π, i)
16:        l ← π[i].level
17:        if π[i+1].intelink[l] = π[i] then
18:            return true
19:        end if
20:        return false
21:    end function
22:    function contest(π*_s, π_c^{f,↑b}, e, f)
23:        require(events[e] ≠ ⊥)
24:        require(events[e].hash = H(π*_s))
25:        require(valid-interlinks(π_c^{f,↑b}))
26:        require(minimal-fork(π*_s, π_c^{f,↑b}, f))
27:        require(arg-at-level(π_c^{f,↑b}) > best-arg(π*_s[f:]))
28:        require(¬evaluate-predicate(π_c^{f,↑b}, e))
29:        events[e] ← ⊥
30:    end function
31:    function arg-at-level(π)
32:        l ← π[−1].level    ▷ Pick proof level from a block
33:        score ← 0                ▷ Set score counter to 0
34:        for b in π do
35:            if (b.level≠l) then
36:                continue
37:            end if
38:            score ← score+2^l
39:        end for
40:        return score
41:    end function
42: end contract
```

that has to be paid in order to dispatch a the proof to the Ethereum blockchain. Therefore, it is essential to provide motives to nodes, while, on the contrary, adversaries have to be dishearten from submitting invalid proofs. We refer to the principle of promoting honest actions against adversarial actions as "fairness".

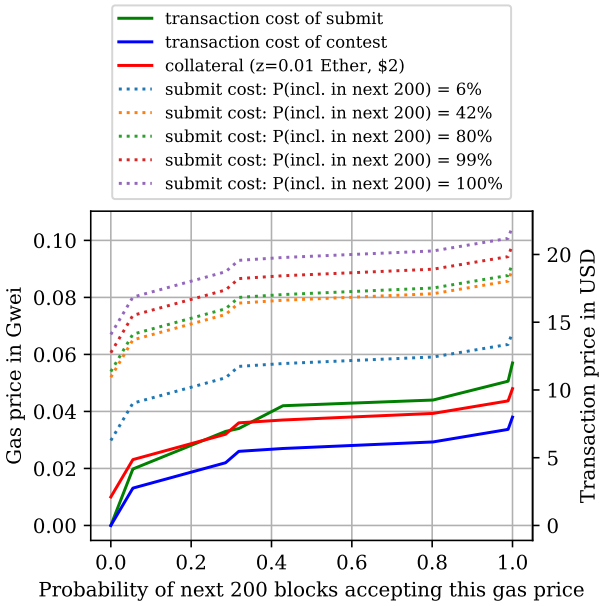
In NIPoPoWs, fairness is addressed by the establishment of a monetary value termed collateral. In *submit* phase, the user pays this collateral in addition to the expenses of the function call, and, if the proof is contested successfully, the collateral is paid to the user that achieves to invalidate the proof. If the proof is not contested, then the collateral is returned to the original issuer. This treatment incentivizes nodes to participate to the protocol, and discourages adversaries from joining. It is critical that the collateral covers all the expenses of the entity issuing the contest.

Collateral. The determination of a fair collateral is not trivial. Thorough analysis has to be made regarding to the gas consumption of all involved phases, as well as the immediacy of required actions. In the Ethereum blockchain, the priority of transactions is determined by the gas price [34] a user assigns to the underlying transaction. This means that the user can chose the estimated time in which transactions are published. Since the duration of contest period in NIPoPoWs is bounded by a finite number of rounds n, the probability of the contesting proof to be included within the following n blocks must be significant. Otherwise, it is possible for an invalid proof to be established due to the lack of challenge. Albeit the user that initiates the submission may demand a direct interaction, and thus selects a high gas price, the value of the collateral is not affected, as the burden of the node is not determined by the priority of the publication of the initial proof.

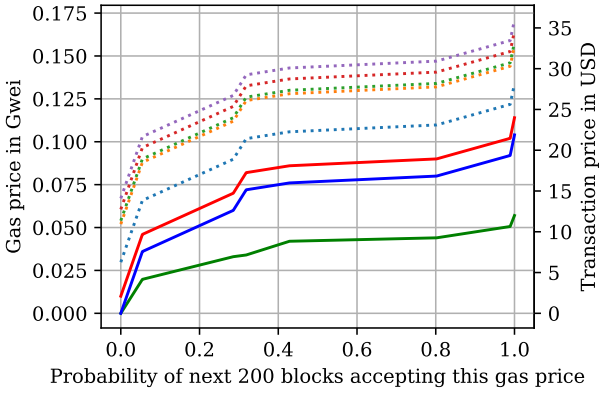
Analysis. We examine various cases in which an invalid submission is followed by a successful contest. Generally, we expect for an adversary to provide a proof of a chain that is a fork of the honest chain at some index relatively close to the tip. This is due to the fact that the ability of an adversary to sustain a fork chain is exponentially weakened as the honest chain progresses. We consider a fork of 100 blocks sufficient to describe an attempt of a power adversary. However, we examine further cases. Our experiments include fraud proofs of chains that fork an honest Bitcoin-like chain 100, 100,000 and 650,000 blocks prior to the tip. The last experiment essentially represents the case of selfish mining [14] from Bitcoin's genesis. We define as Z the profit the node gains in case of a successful contest so that the collateral equals to Z + the expended gas. We consider 0.1 Ether ($2) to be a sufficient amount for Z. Different gas prices formulate different costs for contest.

We use ETH Gas Station [13] to calculate the probabilities of transactions' inclusions with respect to gas price, a tool that is widely used by the Ethereum community. In Figure 10, we illustrate the economical analysis of our client. Green and blue solid lines in each graph display the transaction cost in USD for each phase as a function of the gas price. The red solid line in each graph represents the collateral as a function of the probability of the contest transaction's inclusion in the following 200 blocks. As this probability approaches 1, the gas price needs to be increased. Dashed lines illustrate the total cost of a submission in USD for several selections of collateral, as a product of desired value of gas price. The selection for gas price of the submit phase does not affect collateral, but dictates the immediacy of the transaction by the user.

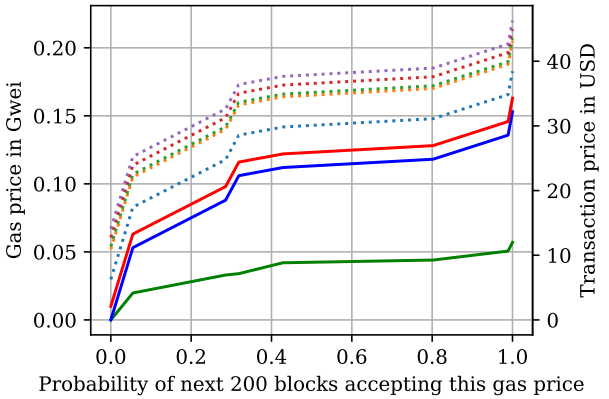
In Figure 10a we observe that the transaction cost of a submission for a chain that equals the length of Bitcoin as a mid priority transaction (80% probability of getting accepted in the next 200 blocks) is \$9.25, while the contest transaction in very high priority (100% probability of being included in the next 200 blocks) is \$7.99. The collateral in this case is \$9.99 (contest cost+Z).



(a) Collateral analysis for fork proof at index 100.



(b) Collateral analysis for fork proof at index 100.000.



(c) Collateral analysis for fork proof at index 650.000.

Figure 10: Economical analysis of RePoPoW.

APPENDIX

A HASH-AND-RESUBMIT VARIATIONS

In order to enable selective dispatch of a segment of interest, different hashing schemas can be adopted, such as Merkle Trees [29] and Merkle Mountain Ranges [27, 32]. In this variation of the pattern, which we term *merkle-hash-and-resubmit*, the signature of an array d is Merkle Tree Root (MTR). In *resubmit* phase, d[m:n] is dispatched, accompanied by the siblings that reconstruct the MTR of d.

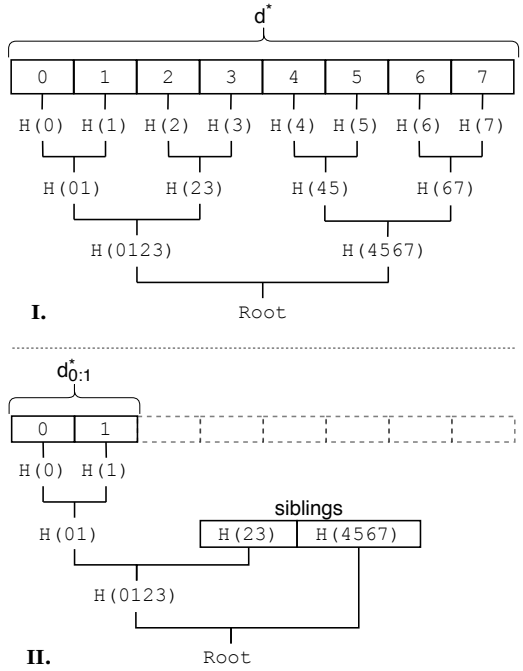


Figure 11: I. The calculation of root in *hash* phase. II. The verification of the root in *resubmit* phase. H(k) denotes the digest of element k. H(kl) denotes the result of H(H(k) | H(l))

This variation of the pattern removes the burden of sending redundant data, however it implies on-chain construction and validation of the Merkle construction. In order to construct a MTR for an array d, |d| hashes are needed for the leafs of the MT, and |d| − 1 hashes are needed for the intermediate nodes. For the verification, the segment of interest d[m:n] and the siblings of the MT are hashed. The size of siblings is approximately $log_2(|\mathsf{d}|)$. The process of constructing and verifying the MTR is displayed in Figure 11.

In Solidity, different hashing operations vary in cost. An invocation of sha256(d), copies *data* in memory, and then the *CALL* instruction is performed by the EVM that calls a pre-compiled contract. In the current state of the EVM, *CALL* costs 700 gas units, and the gas paid for every word when expanding memory is 3 gas units [34]. Consequently, the expression 1 × sha256(d) costs less than |d|×sha256(1) operations. A different cost policy applies for keccak [3] hash function, where hashing costs 30 gas units plus 6 additional gas far each word for input data [34]. The usage of keccak dramatically increases the performance in comparison with sha256, and performs better than plain rehashing if the product of on-chain processing is sufficiently larger that the originally dispatched data. Costs of all related operations are listed in Table 4.

Operation	Gas cost
load(d)	$\mathsf{d}_{bytes} \times 68$
sha256(d)	$\mathsf{d}_{words} \times 3 + 700$
keccak(d)	$\mathsf{d}_{words} \times 6 + 30$

Table 4: Gas cost of EVM operations as of June 2020.

The merkle variation can be potentially improved by dividing d in larger chunks than 1 element. We leave this analysis for future work.

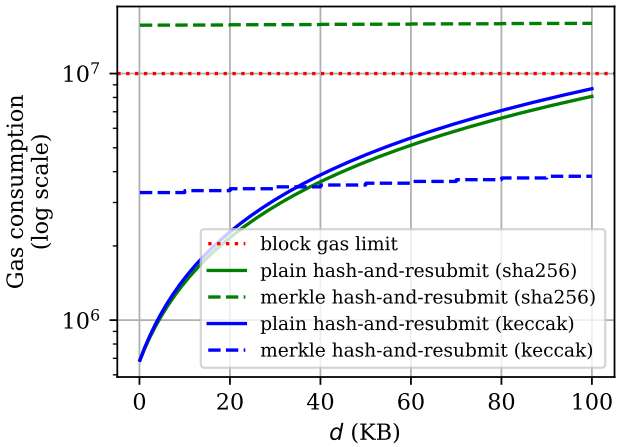


Figure 12: Trade-offs between *hash-and-resubmit* variations. In the vertical axis the gas consumption is displayed. In the horizontal axis the size of d. The size of d_0 is 10KB bytes, and the hash functions we use are the pre-compiled sha256 and keccak.

In Table 5 we display the operations needed for hashing and verifying the underlying data for both variations of the pattern as a function of data size. In Figure 12 we demonstrate the gas consumption for dispatched data of 10KB, and varying size of on-chain process product.

phase per variance	plain hash and resubmit	merkle hash and resubmit
hash	H(d)	H(d_{elem}) × |d| H(digest) × (|d| − 1)
resubmit	load(d) + H(d)	load(d[m:n]) + load($siblings$) + H(d[m:n]) + H($digest$)×|$siblings$|

Table 5: Summary of operations for *hash-and-resubmit* pattern variations. d is the product of on-chain operations and d_{elem} is an element of d. H is a hash function, such as sha256 or keccak, *digest* is the product of H(.) and *siblings* are the siblings of the Merkle Tree constructed for d.

REFERENCES

[1] Mihir Bellare and Phillip Rogaway. 1993. Random oracles are practical: A paradigm for designing efficient protocols. In *Proceedings of the 1st ACM conference on Computer and communications security*. ACM, 62–73.

[2] Juan Benet. 2014. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561* (2014).

[3] Guido Bertoni, Joan Daemen, Michaël Peeters, and GV Assche. 2011. The keccak reference. *Submission to NIST (Round 3)* 13 (2011), 14–15.

[4] Benedikt Bünz, Lucianna Kiffer, Loi Luu, and Mahdi Zamani. 2020. Flyclient: Super-Light Clients for Cryptocurrencies. (2020).

[5] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* (2014).

[6] Ting Chen, Xiaoqi Li, Xiapu Luo, and Xiaosong Zhang. 2017. Under-optimized smart contracts devour your money. In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 442–446.

[7] Ting Chen, Zihao Li, Hao Zhou, Jiachi Chen, Xiapu Luo, Xiaoqi Li, and Xiaosong Zhang. 2018. Towards saving money in using smart contracts. In *2018 IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER)*. IEEE, 81–84.

[8] Alexander Chepurnoy, Charalampos Papamanthou, and Yupeng Zhang. 2018. Edrax: A Cryptocurrency with Stateless Transaction Validation. *IACR Cryptology ePrint Archive* 2018 (2018), 968.

[9] Joseph Chow. 2014. BTC Relay. Available at: https://github.com/ethereum/btcrelay. (Dec 2014). https://github.com/ethereum/btcrelay

[10] Georgios Christoglou. 2018. *Enabling crosschain transactions using NIPoPoWs*. Master's thesis. Imperial College London.

[11] ConsenSys. 2016. A Guide to Events and Logs in Ethereum Smart Contracts. Available at: https://consensys.net/blog/blockchain-development/guide-to-events-and-logs-in-ethereum-smart-contracts/. (June 2016). https://consensys.net/blog/blockchain-development/guide-to-events-and-logs-in-ethereum-smart-contracts/

[12] Cynthia Dwork and Moni Naor. 1992. Pricing via processing or combatting junk mail. In *Annual International Cryptology Conference*. Springer, 139–147.

[13] ethgasstation. 2020. ETH Gas Station. Available at: https://github.com/ethgasstation. (May 2020). https://ethgasstation.info

[14] Ittay Eyal and Emin Gün Sirer. 2014. Majority is not enough: Bitcoin mining is vulnerable. In *International conference on financial cryptography and data security*. Springer, 436–454.

[15] Josselin Feist, Gustavo Grieco, and Alex Groce. 2019. Slither: a static analysis framework for smart contracts. In *2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*. IEEE, 8–15.

[16] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. 2015. The bitcoin backbone protocol: Analysis and applications. *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (2015), 281–310.

[17] Neville Grech, Michael Kong, Anton Jurisevic, Lexi Brent, Bernhard Scholz, and Yannis Smaragdakis. 2018. Madmax: Surviving out-of-gas conditions in ethereum smart contracts. *Proceedings of the ACM on Programming Languages* 2, OOPSLA (2018), 1–27.

[18] Adil H. 2017. Off-Chain Data Storage: Ethereum & IPFS. Available at: https://medium.com/@didil/off-chain-data-storage-ethereum-ipfs-570e030432cf. (October 2017). https://medium.com/@didil/off-chain-data-storage-ethereum-ipfs-570e030432cf

[19] Ethan Heilman, Alison Kendler, Aviv Zohar, and Sharon Goldberg. 2015. Eclipse Attacks on Bitcoin's Peer-to-Peer Network.. In *USENIX Security Symposium*. 129–144.

[20] Kostis Karantias. 2019. *Enabling NIPoPoW Applications on Bitcoin Cash*. Master's thesis. University of Ioannina, Ioannina, Greece.

[21] Kostis Karantias, Aggelos Kiayias, and Dionysis Zindros. 2019. Compact Storage of Superblocks for NIPoPoW Applications. In *The 1st International Conference on Mathematical Research for Blockchain Economy*. Springer Nature.

[22] Kostis Karantias, Aggelos Kiayias, and Dionysis Zindros. 2019. Proof-of-Burn. In *International Conference on Financial Cryptography and Data Security*.

[23] Aggelos Kiayias, Peter Gaži, and Dionysis Zindros. 2019. Proof-of-Stake Sidechains. In *IEEE Symposium on Security and Privacy*. IEEE, IEEE.

[24] Aggelos Kiayias, Nikolaos Lamprou, and Aikaterini-Panagiota Stouka. 2016. Proofs of Proofs of Work with Sublinear Complexity. In *International Conference on Financial Cryptography and Data Security*. Springer, 61–78.

[25] Aggelos Kiayias, Andrew Miller, and Dionysis Zindros. 2020. Non-Interactive Proofs of Proof-of-Work. In *International Conference on Financial Cryptography and Data Security*. Springer.

[26] Aggelos Kiayias and Dionysis Zindros. 2019. Proof-of-Work Sidechains. In *International Conference on Financial Cryptography and Data Security*. Springer, Springer.

[27] Ben Laurie, Adam Langley, and Emilia Kasper. 2013. RFC6962: Certificate Transparency. *Request for Comments. IETF* (2013).

[28] Yuan Lu, Qiang Tang, and Guiling Wang. 2020. Generic Superlight Client for Permissionless Blockchains. *arXiv preprint arXiv:2003.06552* (2020).

[29] Ralph C Merkle. 1987. A digital signature based on a conventional encryption function. In *Conference on the Theory and Application of Cryptographic Techniques*. Springer, 369–378.

[30] Satoshi Nakamoto. 2009. Bitcoin: A peer-to-peer electronic cash system. (2009). http://www.bitcoin.org/bitcoin.pdf

[31] Tak. 2019. Store data by logging to reduce gas cost. Available at: https://github.com/ethereum/EIPs/issues/2307. (October 2019). https://github.com/ethereum/EIPs/issues/2307

[32] Peter Todd. October 2012. Merkle Mountain Ranges. (October 2012). https://github.com/opentimestamps/opentimestamps-server/blob/master/doc/merkle-mountain-range.md.

[33] Franz Volland. 2018. Memory Array Building. Available at: https://github.com/fravoll/solidity-patterns. (April 2018). https://fravoll.github.io/solidity-patterns/memory_array_building.html

[34] Gavin Wood. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper* 151 (2014), 1–32.

[35] Karl Wüst and Arthur Gervais. 2016. *Ethereum eclipse attacks*. Technical Report. ETH Zurich.

[36] Alexei Zamyatin, Mustafa Al-Bassam, Dionysis Zindros, Eleftherios Kokoris-Kogias, Pedro Moreno-Sanchez, Aggelos Kiayias, and William J Knottenbelt. 2019. SoK: Communication across distributed ledgers. (2019).

[37] Dionysis Zindros. 2020. *Decentralized Blockchain Interoperability*. Ph.D. Dissertation.