

Reparo: PUBLICLY VERIFIABLE LAYER TO REPAIR BLOCKCHAINS

Sri Aravinda Krishnan Thyagarajan* Adithya Bhat† Bernardo Magri‡
Daniel Tschudi§ Aniket Kate¶

January 3, 2020

Abstract

Although blockchains aim for immutability as their core feature, several instances have exposed the harms with perfect immutability. The permanence of illicit content inserted in Bitcoin poses a challenge to law enforcement agencies like Interpol, and millions of dollars were lost in buggy smart contracts in Ethereum. A line of research then spawned on Redactable blockchains with the aim of solving the problem of redacting illicit contents from both permissioned and permissionless blockchains. However, all the existing proposals follow the build-new-chain approach for redactions, and *cannot* be integrated with existing running blockchains, such as Bitcoin and Ethereum.

We present **Reparo**,¹ a generic protocol that acts as a publicly verifiable *layer* on top of any blockchain to perform *repairs*, ranging from fixing buggy contracts to removing illicit contents from the chain. **Reparo** follows the layer design, that facilitates additional functionalities for blockchains while maintaining the same provably security guarantee; thus, **Reparo** can be integrated with existing blockchains and start performing repairs in pre-existent data on the chain. Any system user may propose a repair and a deliberation process ensues resulting in a decision that complies with the repair policy of the chain and is publicly verifiable. Our **Reparo** layer can be easily tailored to different consensus requirements, does not require heavy cryptographic machinery and can, therefore, be efficiently instantiated in any permissioned or permissionless setting. We demonstrate it by giving efficient instantiations of **Reparo** on top of Ethereum (with PoS and PoW), Bitcoin, and Cardano. Moreover, we evaluate **Reparo** with Ethereum mainnet and show that the cost of fixing several prominent smart contract bugs is almost negligible. For instance, the cost of repairing the prominent *Parity Multisig wallet* bug with **Reparo** is as low as 0.000000018% of the Ethers that can be retrieved after the fix.

*Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

†Purdue University, USA

‡Concordium Blockchain Research Center, Aarhus University, Denmark

§Concordium, Zurich, Switzerland

¶Purdue University, USA

¹In the Harry Potter universe, 'Reparo' is the repairing charm that can be used to seamlessly repair broken objects.

Contents

1	Introduction	1
1.1	Existing Solutions And Their Limitations	1
1.2	Our Contributions	2
1.3	Key Ideas	3
1.4	Discussion	4
2	Blockchain Formalism	5
2.1	Blockchain as a Ledger	5
2.2	Blockchain Protocol	6
3	Reparo Protocol	6
3.1	Reparo Description	6
3.2	Consensus-Specific Repair Policies	8
4	Instantiation in Ethereum with Proof of Stake	9
4.1	A Primer on Ethereum	9
4.1.1	Ethereum with PoS	10
4.2	Reparo on Ethereum Protocol	11
4.3	On Security and Optimizations	15
5	Experiments in Ethereum	17
5.1	Special Transactions: <code>repairTx</code> , <code>voteTx</code>	18
5.2	Performing Repairs	18
6	Conclusion and Future Work	21
A	Detailed Related Work	24
B	Security Definitions	25
C	Security Analysis	26
D	Reparo in Ethereum	27
D.1	Reparo in Ethereum with PoW	27
D.2	Repairs in Ethereum	30
D.3	Prominent Bugs	31
E	Other Instantiations	31
E.1	Integrating into Bitcoin	31
E.2	Integrating into Cardano (PoS)	34

1 Introduction

Blockchain as the underlying technology of cryptocurrencies, such as Bitcoin [36] and Ethereum [41] is an append-only, decentralized ledger equipped with public verifiability and immutability. While immutability in blockchains was always considered attractive, it does come with several issues. Immutability in monetary aspects is quite unforgiving; e.g., the infamous DAO attack [11] exploited a re-entrance bug in a smart contract resulting in the loss of 3.6 million ETH. In Ethereum alone, other than the DAO bug² more than 750K ETH worth more than \$150 million [2] (at the time of writing) have been either locked, lost or stolen by malicious attackers or bugs in smart contracts [8, 7, 18]. In a cryptocurrency with a fixed supply of tokens, stolen or locked tokens pose a huge problem of deflation [24], and even worse, could adversely affect the consensus process on systems based on Proof of Stake (PoS), which Ethereum 2.0 plans to adopt [20, 19]. Moreover, writing bug-free software, and therefore smart contracts, seems to be a long-standing hard problem and the situation only worsens when many such buggy contracts are uploaded onto the chain resulting in the loss of hundreds of millions of dollars.

Even much-restricted systems such as Bitcoin suffers from the problem of arbitrary data being inserted in the chain through special transactions,³ where all miners are required to store and broadcast the data for validation purposes. Several academic and law enforcement groups have studied the problem of illicit content insertion in Bitcoin [12, 40, 34]. A malicious user can pay a small fee to post illegal and/or harmful content onto the blockchain via these special transactions. Interpol [12] reported the existence of such arbitrary content in the form of illicit materials like child pornography, copyrighted material, sensitive information, etc. on the Bitcoin blockchain. While screening the contents of a transaction before adding it to the blockchain seems to be a straightforward solution, Matzutt et al. [34] showed the infeasibility of this approach while giving a quantitative analysis of already existing contents in the Bitcoin blockchain. Law enforcement agencies [40] are finding it challenging to deal with this problem,

The new General Data Protection Regulation (GDPR) in the European states has thrown the spotlight on the immutability of *personal information* like addresses, transaction values, and timestamps [30]. These issues could adversely affect the adaptability of existing blockchain-based applications, especially for cryptocurrencies if they want to be a credible alternative for fiat currencies.

1.1 Existing Solutions And Their Limitations

Redactable Blockchains. The seminal work of Ateniese et al. [15] was the first to consider the mutability of blockchains. Their redactable Blockchain protocol aims to redact illicit contents from a blockchain using chameleon hash links [32]. However, their protocol requires the miners to run a Multi-Party Computation (MPC) protocol which can be quite prohibitive in large permissionless systems like Bitcoin. Moreover, their protocol requires modifications to the block structure, making it not useful to remove already existing illicit content in the chain of Bitcoin or release frozen ethers in Ethereum. We refer to this property as *Repairability of Existing Contents (REC)*. Puddu, Dmitrienko and Capkun [39]’s proposal suffers from the same problems, and also, presents the control to modify a transaction by the transaction creator, which is not useful if the creator does not allow the desired modifications. Derler et al. [23] solve the above problem by using attribute-based encryption where the transaction creator lets anyone with the right policy to modify the transaction. While they do not require any large-scale MPC among the

² The DAO bug was fixed in July 16’ by introducing an ad-hoc fix that runs DAO transactions differently; resulting in a hard fork, that gave birth to Ethereum Classic.

³Arbitrary information is permitted in Bitcoin through `OP_RETURN` code, that can store up to 80 bytes of arbitrary data on the blockchain.

Table 1: Comparison of our work with that of the existing redaction solutions. A cross for *Repairability of Existing Contents (REC)* means the proposal is not useful to redact or modify already inserted contents in blockchain.

Proposals	Stateful repairs	System-scale MPC	REC	Public verifiability
Ateniese et al. [15]	×	Required	×	×
Puddu et al. [39]	×	Required	×	×
Derler et al. [23]	×	Not required	×	×
Deuber et al. [25]	×	Not required	×	✓
Tezos [29]	✓	Not required	×	✓
This work - Reparo	✓	Not required	✓	✓

miners, their protocol lacks public verifiability and requires modifications on how the Merkle roots are computed in the blocks, hence does not guarantee REC for Bitcoin or Ethereum. The recent work of Deuber, Magri and Thyagarajan [25] leverages on-chain voting techniques to reach an agreement on the redaction of contents, thereby adding public verifiability to the redactions. However, their protocol also requires modifications to the block header and therefore does not guarantee REC for current systems. Tezos [29] proposed a PoS protocol that can instantiate any blockchain but does not guarantee REC. While lacking formal security guarantees, it also lacks efficiency for multiple updates. Given that all the aforementioned proposals are *build-new-chain* solutions (no REC) and suffer from other issues as discussed, *none* of them are integrable into existing mainstream permissionless blockchain systems guaranteeing REC⁴ [35]. Table 1 summarizes the above discussed limitations. For an extended technical discussion and comparison, we refer the readers to Appendix A.

Hard Forks. Performing a repair by forking away from a faulty point in the blockchain can lead to a loss of blocks. A hard fork requires miners to update their client software and corresponding mining hardware. Every hard fork brings with it an additional consensus rule in order to validate the whole chain. These additional rules demand additional storage and computational capabilities from clients. Hard forks are ad-hoc: in Ethereum, DAO was deemed to be a big enough bug to fork the chain, whereas Parity Multisig Wallet was not. [8]

Pruning. For repair operations such as redactions or removing old content, there are pruning solutions that locally redact contents [10]. However, the primary purpose of this method is space optimization and there is no consensus on what can be removed or redacted. Therefore, a newly joining full node is still expected to receive all the information on the chain for thorough validation.

1.2 Our Contributions

We present Reparo (Section 3), which is the first protocol that acts as a *layer* (in the style of the finality *layer* for blockchains from [33]) on top of *any* existing secure blockchain and allows repair operations on its contents. Our protocol aims to provide a solution that can be easily integrated into virtually any existing blockchain system, and departs from the *build-new-chain* approach in the literature. Although Reparo is bound to the underlying consensus requirements (e.g., PoW, PoS, as discussed in Section 3.2), it can easily be adapted to any flavor of consensus (include permissioned systems) without any overhead. We formally prove that such an integration of Reparo with a secure blockchain satisfies the standard security properties of *chain quality*, *chain*

⁴In case of permissioned setting, Ateniese et al [15]’s proposal has been commercially adopted by a large consultancy company [13, 14].

growth and *editable common prefix* (which were introduced in [25]). We argue that our **Reparo** protocol potentially could improve the parameters of the chain quality of the integrated system.

The main advantage of **Reparo** as a repair layer in contrast to a repairable blockchain is that, after integration into systems like Bitcoin and Ethereum, the contents that already exists in these chains become repairable and thus guaranteeing the REC property: once **Reparo** is incorporated into the clients of Bitcoin or Ethereum, any previously existing contents can be repaired. In this direction, we give a detailed instantiation of our **Reparo** integration into Ethereum when the underlying consensus is PoS (Section 4) and PoW (Appendix D.1), Bitcoin (Appendix E.1) and Cardano (a PoS based system Appendix E.2). We demonstrate how to perform repair operations in Ethereum, which can fix smart contract bugs. For Bitcoin, we show how to perform redaction of arbitrary data entries (illicit data) that are non-monetary information without changing the Bitcoin block structure or using any heavy cryptographic machinery. With respect to [25], our instantiation with **Reparo** has comparable efficiency in terms of time and is significantly better in terms of space efficiency: unlike [25], **Reparo** does not require an additional hash value to be stored in every block header (as detailed in Appendix E.1). Also, **Reparo** is better than a hard fork. For instance, consider a situation where a user accidentally creates a contract with no code in it, it is safe for the user to create a repair transaction with **Reparo** that adds code to this contract without forking. Users of the system can skip the expensive, cumbersome and often times arbitrary⁵ procedures involving a hard fork.

Finally, we offer a proof-of-concept implementation of our **Reparo** protocol integrated into Ethereum. As we show in Section 5, when importing the latest 2 million block sub-chain (from block number 6M to 8M) from the Ethereum main network, our baseline implementation has an overhead of just 12.52% when compared to its vanilla counterpart. The choice of Ethereum is motivated by the wide-spread adaption and generality of the Ethereum’s functionalities.

Practical Implications. Apart from illicit data redaction in Bitcoin, for systems such as Ethereum, a repair involves re-running all the transactions that are affected thus demanding computation from the network. Therefore, a repair proposal must pay (in gas) an amount proportional to the computation spent by the network performing the repair. We measure the repair costs of various existing bugs affecting Ethereum today.

For concreteness, we demonstrate that the Parity Multi Sig Wallet Bug, which locked over 513K ETH, can be repaired today by paying a little over 0.00094 ETH in gas. **Reparo** also gives a mechanism to resolve an issue where users submit a contract creation transaction with no code [9] (due to user errors or buggy wallet code), releasing over 6.53K ETH. Ethereum uses an ad-hoc fix for DAO as it hard-codes a different logic for DAO. **Reparo** can be used to remove this ad-hoc fix by first repairing DAO code (while the fix is still active) and later removing the fix. One could also handle zero-day vulnerabilities and thereby restrict losses.

Future adopters of blockchains such as governments and corporations can use **Reparo** as a provably secure protocol for regulations and maintenance.

1.3 Key Ideas

Reparo acts as a layer on top of an existing blockchain system. The system users agree beforehand what the policy is going to be for performing repair operations on the chain. This policy specifies the constraints and requirements a proposal needs to satisfy for getting approved. As shown in Figure 1, **Reparo** layer constitutes 6 sequential steps for the case of a redaction proposal. Any user can propose a repair proposal rp_1 for block B_1 that redacts some contents in B_1 . The proposal also determines the updated state of the chain unlike previous proposals.

⁵ For example, in Ethereum, DAO was deemed to be an important enough attack requiring a hard fork, whereas the Parity Multisig wallet was not.

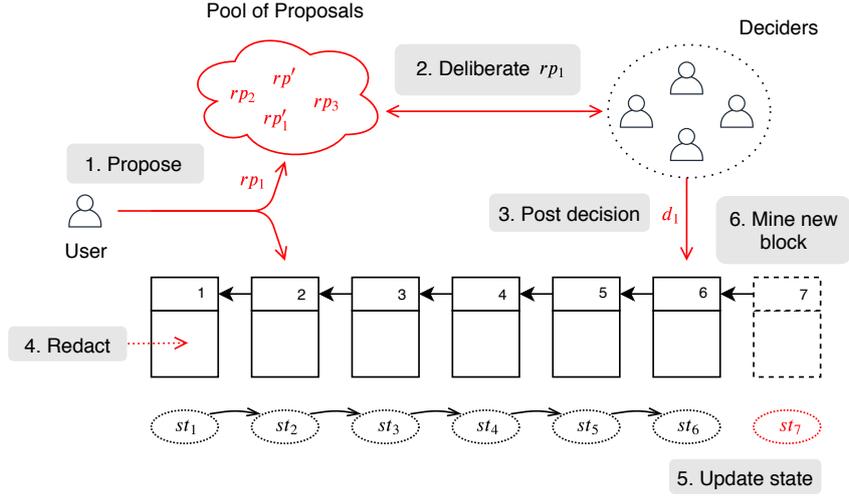


Figure 1: Step by step description of Reparo in case of performing a redaction. The steps are in gray boxes and highlighted with red color. Here a block is partitioned into a block header and block body and the block header stores a pointer to the previous block’s header. State st_i denotes the state of the chain after block i has been mined.

A group of users called *deciders* deliberate on the proposal (step-2) and after the deliberation process, they post their decision d_1 on the chain (step-3). After the decision is posted for rp_1 , it is removed from the pool of proposals. Miners check if decision d_1 for the proposal rp_1 is positive and if rp_1 adheres to the repair policy guidelines. If so, the miners redact the said contents in block B_1 as per the proposal in step 4. Miners update the state of the chain as per rp_1 (step-5) and propose a new block B_7 with this updated state resulting in an extension of a repaired chain (step-6).

The key novelty in Reparo is in the consensus-based and cost-efficient (compared to recreating an alternate chain) updating of the state st_7 that accounts for the states and contents of previous blocks, new incoming contents and the redaction operation that was performed in step 4. Reparo also allows repair operations other than redactions also, and in that case, the miners skip step 4 and only update (or repair) the state of the chain according to the proposal. This state update takes into account the necessary repairs to the chain that the user wishes to perform.

1.4 Discussion

Notice that Reparo possesses public verifiability of proposals, deliberation and repair operations: Reparo has accountability during and after a deliberation process is over for any repair proposal, referred to as *voting phase accountability* and *victim and new user accountability* [25]. In this section we argue about some crucial features of Reparo that makes it stand apart from the rest of the proposals.

What if users decide to retain redacted data? Similar to previous proposals, Reparo does not enforce complete removal of redacted data from a user’s local storage. Users can still locally keep redacted data, however, once repaired by Reparo the users are not required by the blockchain protocol to store the redacted information. For instance, in the case of illicit content, this means that the miners who locally keep and broadcast illicit (redacted) data can be prosecuted individually if necessary and the system as a whole is not liable.

Can a bad set of deciders retroactively censor transactions? Similar to censorship of transaction inclusion by miners, it is also possible to “censor” transactions retroactively through repair operations. However, this can be easily mitigated by requiring multiple decider sets across different deliberation phases to approve a repair operation. Thus, a single bad set of deciders at a given time interval cannot censor. Moreover, contrary to the censorship on transaction inclusion, attempts to censor through repair operations are publicly verifiable as the transaction is already on chain and the network is aware of the deliberation process.

How is Reparo different from the DAO fix in Ethereum? The hard fork in Ethereum to fix the DAO bug was an ad hoc software patch in the Ethereum client. On the other hand, Reparo is a layer on top of the underlying blockchain system that can handle virtually any kind of repair operations subject to restrictions of the policy.

Using Reparo to perform monetary changes in the state can cause inconsistencies? Although Reparo here is described in a generic way, in Bitcoin for example the repair policy could restrict repair operations to be only redaction of auxiliary data that does not affect user’s balances. For Ethereum, the policy could allow contract bug fixes that indeed affects monetary balances of user accounts.

2 Blockchain Formalism

2.1 Blockchain as a Ledger

A blockchain is simply a chain (or sequence) of blocks that we call \mathcal{C} . The i -th block in the chain \mathcal{C} is denoted by $B_i := \langle \text{header}_i, x_i \rangle$, where $\text{header}_i = (pt, G(x_i), hd)$ and x_i denotes the messages contained in the block. Here, $H : \{0, 1\}^* \rightarrow \{0, 1\}^\kappa$ and $G : \{0, 1\}^* \rightarrow \{0, 1\}^\kappa$ are cryptographic hash functions, pt is the hash of the previous block header $H(\text{header}_{i-1})$ and hd contains some special header data, such as the consensus proof π_i for the block (e.g., a nonce for PoW or a stakeholder signature for PoS). For a block B_i to be considered valid, it needs to satisfy a public set of requirements established by the protocol; the requirements can vary depending on the application of the blockchain, but at the very least the consensus proof π_i , and the set of transactions contained in x_i needs to be valid according to some pre-determined rules.

The rightmost block is called the head of the chain, denoted by $\text{Head}(\mathcal{C})$. The sequence of blocks till the $\text{Head}(\mathcal{C})$ defines the state st of the chain \mathcal{C} . Any chain \mathcal{C} with a head $\text{Head}(\mathcal{C}) := \langle \text{header}, x \rangle$ can be extended to a new longer chain $\mathcal{C}' := \mathcal{C} || B'$ by attaching a (valid) block $B' := \langle \text{header}', x' \rangle$ where $pt = H(\text{header})$ and the state of \mathcal{C}' is updated to st' ; the head of the new chain \mathcal{C}' is $\text{Head}(\mathcal{C}') := B'$. A chain \mathcal{C} can also be empty, and in such a case we let $\mathcal{C} := \varepsilon$. For a chain \mathcal{C} of length n and any $q \geq 0$, we denote by $\mathcal{C}^{\lceil q}$ the chain resulting from removing the q rightmost blocks of \mathcal{C} , $\mathcal{C}^{\lfloor q}$ to denote the chain with the first q blocks of \mathcal{C} and we denote by ${}^q\mathcal{C}$ the chain resulting in removing the q leftmost blocks of \mathcal{C} ; note that if $q \geq n$ (where $\text{len}(\mathcal{C}) = n$) then $\mathcal{C}^{\lceil q} := \varepsilon$ and ${}^q\mathcal{C} := \varepsilon$. If \mathcal{C} is a prefix of \mathcal{C}' we write $\mathcal{C} \prec \mathcal{C}'$. For a more detailed and precise definition of blockchain and its functionalities and assumptions we refer the reader to [17].

A secure blockchain satisfies the properties of common prefix, chain growth and chain quality [27, 37, 31]. It is shown that these properties of the blockchain satisfy *persistence* and *liveness* with which we can classify it as a “healthy” blockchain. Intuitively, in a healthy blockchain after some time period, all honest users of the system will have a consistent view of the chain,

Table 2: Interfaces provided by the blockchain protocol Γ . As an example of the stability interface, in Bitcoin the stable part is the chain pruned by the most recent k blocks (e.g. $k = 6$). The idea is that the stable part of a chain will (with overwhelming probability) remain immutable.

Interface	Description
$\Gamma.\text{validateChain}(\mathcal{C})$	returns 1 iff all the blocks in the chain are valid according to a public set of rules and the links between blocks are well-formed
$\Gamma.\text{validateBlock}(B)$	for a block $B := \langle \text{header}, x \rangle$ returns 1 iff the block is valid; specifically the hash of x is equal to the data-pointer $G(x)$ in header.
$\Gamma.\text{broadcast}(x)$	broadcasts data x to all the parties.
$\Gamma.\text{stable}(\mathcal{C})$	returns the stable part of the chain \mathcal{C} .

and transactions posted by honest users will eventually be included. For formal definition of the above properties, we refer the reader to Appendix B.

2.2 Blockchain Protocol

The basis of our Reparo protocol is a healthy immutable blockchain protocol (e.g., [27, 22]), denoted by Γ . In Γ , parties are categorized into different roles (not mutually exclusive), namely *users* and *miners*. Users can send inputs in the form of messages while miners try to extend the blockchain by creating and appending new blocks containing the users’ messages. The existing blockchain protocols [27, 37, 31] achieve the security properties stated previously, by assuming that the majority of miners are *honest*: honest miners behave according to the protocol. We make the same assumption. Therefore, in protocol Γ we also assume that the majority of miners are honest (if one instantiates Γ with the protocols from [27, 37, 31]).

We refer as *node* to any party in the system, be it a user or a miner. Each node locally stores its current chain \mathcal{C} . We assume that nodes automatically update their local chain whenever there is a *better*⁶ valid chain available. We assume that a node has access to the interfaces as described in Table 2.

3 Reparo Protocol

In this section we show how to extend a given immutable blockchain protocol Γ (as described in Section 2.2) into a *repairable* blockchain protocol Γ' that permits repair operations ranging from *redactions* to *state updates* on the chain.

3.1 Reparo Description

We add an additional category of parties (apart from users and miners) to the set of parties involved in our protocol, namely *deciders*. Note that the categories are *not* mutually exclusive, e.g., a user can also be a decider and/or a miner.

The Reparo protocol, formally described in Figure 2, allows repair operations on the underlying blockchain Γ : redaction of data from Γ and/or special changes in the current state of the chain. It communicates with Γ through the interfaces described in Table 2. Similar to [25], Reparo is parametrized by a repair policy \mathbb{P} . The integration of Reparo with the blockchain protocol Γ is denoted by $\Gamma'_{\mathbb{P}}$. We describe the two flavors of repairs in more details next.

Redactions: Without loss of generality, we consider a redaction to be the removal of the entire content (i.e., all transactions) of a block. The redactions can be made much more fine-grained

⁶According to the “best chain rule” of the underlying blockchain system.

Initialisation (miner/decider). Initialize the *proposal pool*, $\text{propPool} \leftarrow \emptyset$.

Proposal (user). To propose a repair:

1. create repair proposal, $rp = \langle (pt, x'), sp \rangle$.
2. broadcast it to the network, $\Gamma.\text{broadcast}(rp)$.

Update proposal pool (miner/decider). In periodic intervals:

1. collect all valid proposals rp and set $\text{propPool} \leftarrow \text{propPool} \cup \{rp\}$.
2. if proposal rp has a corresponding repair witness w in $\text{stable}(\mathcal{C})$ set $\text{propPool} \leftarrow \text{propPool} / \{rp\}$ to remove rp from the pool.

Deliberation process (decider/miners). For each new proposal $rp \in \text{propPool}$:

1. deciders deliberate and come to a decision, denoted as $w \leftarrow \text{decision}(rp)$.
2. miners add $hd \leftarrow hd \cup \{w\}$, where hd is part of the next new block.

Repairing the chain (miners). For each $w := \langle pt, H(rp), G(x'), sp, b, pf \rangle \in \text{stable}(\mathcal{C})$:

1. if $b = 1$ and $\text{chkApproval}(\mathbb{P}, w) = 1$,
 - (a) replace the data x in the block pointed by pt by the new data x' . If $x' = x$ no action is needed.
 - (b) update state st of chain \mathcal{C} to st' using sp .
2. else if $b = 0$, ignore rp as deciders have rejected.

Chain validation (miners). Update $\Gamma.\text{validateChain}$ to handle repair operations:

1. start validating block B from genesis.
2. for a block $B := \langle \text{header}, x \rangle$, where $\text{header} := (pt, g, hd)$, update $\Gamma.\text{validateBlock}$ to do the following checks:
 - (a) if $G(x) = g$, then no repair has happened, go to step 3.
 - (b) else, retrieve all repair witnesses of the form $w := \langle pt, h', g', sp, 1, pf \rangle \in \mathcal{C}$ where pt points to B .
 - (c) for each of these repair witnesses w in the same order of their retrieval, do the following steps (exactly as it was performed by the miners originally during repairing):
 - (i) check if $\text{chkApproval}(\mathbb{P}, w) = 1$,
 - (ii) check if the corresponding repair proposal $rp := \langle (pt, x'), sp \rangle$ (where $G(x') = g'$ and $H(rp) = h'$) was performed correctly according to the witness w .
 - (iii) check if state updates of \mathcal{C} was correctly performed according to sp .
 - (d) for the final repair witness $w := \langle pt, h', g', sp, 1, pf \rangle$ in the order that was retrieved:
 - (i) check if it holds that $G(x) = g'$ to see if the current data in B (that is pointed to by pt). This check also works for redactions.
 - (ii) check if the final state obtained after applying all the state updates sp from all witnesses is consistent with the state of the chain at B under validation.
3. finally, ensure that all repair operations that have an approved witness w on the chain have been performed. This check can be performed on the fly as we validate blocks from the genesis.

Figure 2: Repairable blockchain protocol $\Gamma_{\mathbb{P}}'$ resulted from adding *Reparo* on top of Γ and with policy \mathbb{P} .

(e.g., on the transaction level) by adding cumbersome details, what we defer to Section 4, when we present an instantiation of **Reparo** on top of Ethereum.

Other repair operations: Or the repair operation acts as any other type of message that changes the current state of the chain from st to st' . For instance, in this case, in contrast to normal messages such as “Alice send some coins to Bob” with Alice authenticating this transfer, these repair messages can alter the current balances of Alice and Bob arbitrarily without requiring authentication from either Alice or Bob.

Repair Proposal. Any user can propose any type of repair request. A repair proposal $rp = \langle (pt, x'), sp \rangle$ consists of a block-pointer pt and the new data x' (or \perp in case of redaction) and sp is a (set of) message(s) that describes the desired state change (state st of \mathcal{C} is changed to st'). If the repair is a redaction, then the original data x stored in the block pointed to by pt is removed, and the state of the chain is updated using sp . If its not a redaction, then only the state is updated and no data needs to be modified physically (as $x' = x$ where x is stored in the block pointed to by pt). This means that the state change sp always describes the transition of state of the chain from st to st' irrespective of whether the repair is a redaction or not.

Deliberation. These proposals are collected by miners and deciders. The deciders then use a *publicly verifiable* decision protocol $\text{decision}(rp)$ to deliberate whether a proposal rp should be accepted or not. The protocol outputs their final decision in the form of a repair witness w . Miners then add the witness into the header data of the next created block. For concreteness, if we let the decision process follow as in [25], where deciders (i.e., the miners themselves) add their votes to the header data hd of their newly created blocks before broadcasting it to the network; the witness can then be easily “extracted” from the header data of all the blocks during the deliberation period, by simply counting how many votes supported the proposal. The repair witness $w := \langle pt, H(rp), G(x'), sp, b, pf \rangle$ consists of a block-pointer pt , a pointer to the corresponding repair proposal $H(rp)$ and the pointer to the new data x' , the proposed state-change sp , the decision bit b , and a proof pf which allows to validate the decision (e.g., verifiable vote count). Note that the proposed state change sp could be empty, which is the case when the repair operation is stateless modification. For security, we require the witness proof pf to be sound, i.e., it should be infeasible for an adversary to produce a valid proof pf' for rp if rp was *not* accepted by the protocol.

Repair Policy. Repair policy \mathbb{P} dictates the constraints of different repair operations, e.g., what is the duration of the deliberation period, what type of data can be redacted, what changes in the state are allowed, just as in [25]. For our case, as minimum requirements from a *valid* policy \mathbb{P} , we have (i) a detailed description of what contents can be redacted and what kind of state changes are allowed, (ii) a well defined period (in rounds) for the deliberation process for each proposal, (iii) the header data hd of blocks can not be edited. This implies that repair witnesses of other proposals cannot be edited, and (iv) system parameters that determine block creation are not modified. For instance, one cannot modify the mining difficulty (in case of PoW) that was used in some block in Bitcoin.

Policy Approval. We assume that there is a predicate $\text{chkApproval}(\mathbb{P}, w)$ which determines if a proposal rp is approved. It takes as input the policy \mathbb{P} , a repair witness $w := \langle pt, H(rp), G(x'), sp, b, pf \rangle$ for a repair proposal rp . The predicate outputs 1 if the proposal rp is accepted by the deciders ($b = 1$) with a valid proof pf and complies by the policy \mathbb{P} , and outputs 0 otherwise. For the formal analysis of security we refer the reader to Appendix C.

3.2 Consensus-Specific Repair Policies

Here we discuss how the **Reparo** repair policy \mathbb{P} deals with different consensus specific challenges.

Proof of Work (PoW). The set of deciders are chosen in a sybil-resistant manner. When

the underlying chain is PoW based, one could select the deciders via PoW itself, where the deciders are required to show proof of work. Necessary bounds on the fraction of adversarial deciders are discussed in Appendix C. The repair policy of `Reparo` in this setting need not have any restrictions on the kind of repair operations that can be performed: data can be redacted or modified. However, though `Reparo` does not impose any restrictions, some applications may prefer to have policies that allow only restricted repair operations. For instance, if one is interested to redact arbitrary illicit non-payment data from Bitcoin transactions, the `Reparo` repair policy \mathbb{P} can be set accordingly. On the other hand, if one is interested in fixing buggy contracts that have cost a lot of money and effort (in case of Ethereum), the policy could be set to allow specific repair operations on the state of the system.

Proof of Stake (PoS). In case of PoS based consensus, the repair policy should ensure that repair operations do not invalidate consensus. More specifically, the repair policy \mathbb{P} should disallow redactions of state. This is because PoS inherently relies on the state of the system for consensus, and removing some state information permanently makes the existing consensus proofs unverifiable. Of course, redactions that do not affect the state of the chain can still be performed with `Reparo`. The repair policy should also ensure that during the deliberation process the set of deciders do not change. In other words, the deliberation process should happen in a phase where the set of deciders are fixed. This ensures that any repair operation does not affect any other ongoing deliberation process or the decider set.

4 Instantiation in Ethereum with Proof of Stake

We discuss PoS in Ethereum and then continue to describe the working of ethereum today. We then proceed to detail how one can instantiate our `Reparo` layer protocol of Section 3.1 on top of Ethereum to support repair operations: redaction of transaction contents and/or state updates in the form of smart-contracts “patches” and account balance update (e.g., restitution of stolen coins).

4.1 A Primer on Ethereum

Ethereum [41] is a decentralized virtual machine (Ethereum Virtual Machine or EVM), which runs user programs - *smart contracts* - upon user’s request. Roughly, a contract is a collection of functions and variables, where each function is defined by a sequence of bytecode instructions that operate on the function input and the variables associated with the contract. The contract has an address for users in the network to interact with, and this address depends on the contract creator. A user may interact with a contract through transactions that calls functions in the contract.

Transactions and Block Structure. An Ethereum transaction `tx` can serve two purposes: message calls or special calls. The `tx.from` field is derived from signature values `tx.r`, `tx.s`⁷. The `tx.to` field contains the 160-bit address of the recipient. The `tx.value` field contains the amount of ether (in Wei) transfer from the sender to the recipient, and in case of contract creation it initializes the contract with the amount. The `tx.data` field optionally contains EVM bytecode for contract creation or an encoding of a function call of a contract. There are special reserved recipient addresses like `0x00..0-8` for special calls. These addresses contain native contracts. Native contracts contain instructions that are not executed by the EVM. Similar to Bitcoin, Ethereum has a block header and block content associated with a block. The relevant⁸ contents of the block header are shown and described in Table 3.

⁷These are ECDSA signatures that help derive the public key and thus, the sender of the transaction.

⁸There are other fields like `ommers_hash`, `receipts_root`, `extra_data` in the Ethereum block header.

Table 3: Structure of the Ethereum block header. Relating to the abstract protocol from Figure 2, G is of the form $G = (G_{tx}, G_{st})$, and $hd = (d, t, ctr)$.

Value	Description
<code>parent_hash</code> (pt)	hash ⁹ of the previous block header
<code>state_root</code> (G_{st})	hash of the root node of the state tree, after all transactions are executed and finalizations applied
<code>tx_root</code> (G_{tx})	hash of the root node of the tree structure populated with all the transactions in the block
<code>difficulty</code> (d)	the difficulty of the proof-of-work
<code>timestamp</code> (t)	the timestamp of the block
<code>nonce</code> (ctr)	value used in proof-of-work

Accounts and State. State in Ethereum is denoted by ACC which consists of account objects. There are two types of accounts in Ethereum: the external account and the contract account. Both types of accounts (Acc) contain balance ($Acc.bal$), storage root ($Acc.sr$), nonce ($Acc.nonce$) and code hash ($Acc.h$). $Acc.sr$ is the hash digest of the trie encoding of the state of the contract while code hash $Acc.h$ is the hash of the contract bytecode. An external account has empty $Acc.h$ and $Acc.sr$. The effect of the transactions included in a block have on the accounts is the state of the accounts at the time; reflected in the $Acc.bal$ and $Acc.sr$ fields of accounts at the time of mining the block and consequently in the `state_root` stored in the block header.

The ACC is updated every block by using a global state transition function $\delta : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^*$. This function takes as input the state of the accounts in the previous block and new transactions included in the current block and returns the current state of the accounts in the chain. For block B_i we have $ACC_i \leftarrow \delta(ACC_{i-1}, TX_i)$ where ACC_{i-1} is the state of accounts in B_{i-1} . The output of this function can be thought of as the changing of account states to ACC_i from their previous state ACC_{i-1} after applying the new incoming transactions TX_i . These transactions are validated (signature, balance and nonce checks) according to Ethereum rules before letting them affect the state transition. Note that ACC is analogous to UTXO in Bitcoin and is derived from the block but does not exist as a part of the chain.

4.1.1 Ethereum with PoS

We briefly describe the Ethereum protocol when its PoW consensus is replaced with a PoS consensus like Algorand [28] or Ouroboros Praos (OP) [22]. Algorand is a Byzantine fault tolerant (BFT) consensus, where a set of nodes (known as *committee*) is selected through a sortition procedure based on the weight of the stake they own. Every round, the committee engage in a Byzantine agreement protocol to produce a new block to be appended in the Algorand chain. OP, on the other hand is a slot based consensus protocol where time is divided into slots and blocks are created relative to a slot. Parties with stake can participate in a slot lottery, and winning the lottery (referred to as *slotleader*) allows a stakeholder to create a block in a particular slot. The probability of winning the lottery for a stakeholder is directly proportional to his stake. Assuming for simplicity that users in Ethereum have one account each, then OP dictates that the probability of winning the lottery is: $\phi_f(Acc) := 1 - (1 - f)^{b_{Acc}/S}$, where Acc is the account of the stakeholder, b_{Acc} is the balance in this account, S is the total stake in the system and f is some difficulty parameter. We make block-box use of the sortition procedure or the lottery function for our work.

The main difference in contrast to PoW is that the difficulty d and nonce ctr are not in the

⁹Ethereum uses the 256-bit variant of Keccak/SHA3.

block header. Instead, generating a validity proof for a block is done by the algorithm `prf_pos` that outputs a proof σ on the block with respect to some account (address) referred by `Acc`. Verifying the proof of stake is done by the algorithm `vfy_pos`.

4.2 Reparo on Ethereum Protocol

In this section we describe Reparo on Ethereum when the PoS consensus is instantiated with Algorand or OP. The reasoning is that, in both these proposals the stakeholder also happen to be slotleaders if they are chosen.

As in Section 3, we formally denote a block in Ethereum as $B := \langle \text{header}, TX \rangle (ACC)$, where TX denotes the set of transactions (individual transaction is denoted by Tx), $\text{header} := (pt, G_{tx}(TX), G_{st}(ACC), hd)$ as in Table 3 and ACC denotes the state of the accounts in Ethereum. Here $G_{tx}(TX)$ is the Merkle root of the transactions, $G_{st}(ACC)$ is the Merkle root of the account state.

Regarding the roles of parties when PoS is instantiated with Algorand or OP, we consider *miners* or *slotleaders* in the PoS setting also to be *deciders* of repair proposals. This means that the deliberation process happens on the chain with *slotleaders* voting on proposals by adding special voting transactions in the header data hd of their blocks. Recall that in Algorand *slotleaders* are referred to as the *committee members* who are chosen to propose a block at that round. Due to space constraints, we give a formal description of the protocol in Appendix D (Figure 5).

On a very high level, while performing repair operations, we repair the block contents using new Reparo data structures. This ensures that the block header always remains unchanged while only the block contents are repaired. This enables efficient multiple repairs on a block: multiple repairs on the same block’s state (direct) or the block’s state gets updated multiple times due to a cascading effect (indirect). Physically repairing block contents while making use of the Reparo data structures also improves efficiency in terms of consensus for PoS. This is because the blocks always contain the most recent state of accounts and balances which makes the retrieval of updated stakeholder distribution in case of PoS easier.

To see how repair operations could affect the state of accounts denoted by A , changing contents can affect the state of the concerned accounts, which could subsequently lead to cascading changes to other accounts. This is pictorially described in Figure 3 where a user proposes to fix a buggy contract C in step 1. During the voting period of $\ell (= 5)$ blocks that coincides with a PoS epoch, slotleaders vote for the proposal by adding a vote inside the block that they propose in step 2. The reason for the voting period and the epoch to coincide was explained in Section 3.2. If enough votes are obtained and the proposal satisfies some set of policy guidelines, in step 3 slotleaders fix the contract C according to the proposal. The states of all accounts in the subsequent blocks are updated amounting to this fix in step 4 and a new state A'_g is obtained in step 5. In step 6, this updated state is reflected on the chain by proposing the next block with respect to this state (by including $G_{st}(A'_g)$ in the header).

Proposing Repairs. Any user in the system can request a repair of the chain. The user first broadcasts the candidate transaction Tx^* to the network. Then, the user sends a proposal transaction tx . The `tx.to` address field contains the special address `REQ_ADDR`. `REQ_ADDR` is a native contract for Reparo. The `tx.data` field contains $(H(\text{Tx}), H(\text{Tx}^*))$, the hash of the old version (Tx) and the new version (Tx^*) of the transaction. For smart contract bug fixes, Tx was the buggy-contract creation transaction, while Tx^* is a similar transaction with the bug fixed. The `repairTx` offers processing fees to the slotleader who includes the transaction into the block and could also offer a approval fee to the slotleader who performs the repair after the policy approval. The user also adds the new version of the transaction Tx^* to the candidate

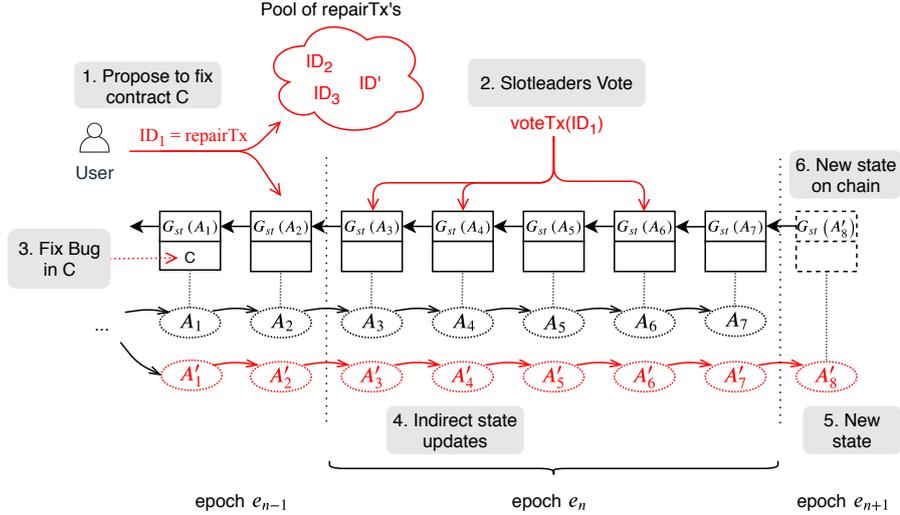


Figure 3: An overview of Reparo in Ethereum with PoS to fix a buggy contract C in block 1. The Reparo layer steps are numbered inside gray boxes and highlighted in red. The voting period starts at the start of the next epoch and lasts for $\ell = 5$ blocks. Proposal ID_1 is approved at block 7. The voting period coincides with the start and end of an epoch.

Algorithm 1: proposeRepair

input : Chain $\mathcal{C} = (B_1, \dots, B_n)$ of length n , an index $j \in [n]$, and new set of transactions TX_j^* .
output: A repair proposal rp_j^* .

- 1: Parse $B_{j-1} := \langle \text{header}_{j-1}, TX_{j-1} \rangle (ACC_{j-1})$;
 - 2: Build the repair proposal $rp_j^* \leftarrow \langle TX_j^* \parallel \delta(ACC_{j-1}, TX_j^*) \rangle$;
 - 3: **return** rp_j^* ;
-

database for the users to validate and consider voting for the proposal¹⁰. A formal description of proposing a candidate block is given in Algorithm 1. The repair proposal rp_j^* can be seen as including the candidate transaction and other unchanged transactions of that block (allowing fine-grained transaction level changes). It also includes the new state of accounts: as the new set of transactions (candidate transaction plus other old transactions) are applied on the state of accounts of the previous block. It is interesting to note here that a repair proposal can be accompanied by a bounty (in tx.value) incentivizing the network to approve the repair. We provide more details in Appendix D.2.

Validating Requests. Nodes validate a repair proposal by checking if the proposed new Tx^* is a well-formed transaction as per rules of Ethereum (correct format, correct signatures, etc) and Tx is in the chain. They also check if the proposal rp_j^* from Algorithm 1 includes the correct state of accounts after applying Tx^* and other unchanged transactions of block B_j on the state of accounts in B_{j-1} . Proposals are rejected as redundant if they are already in the voting phase.

Reparo Layer. Reparo has new data structures that help store the block contents (transactions and state of accounts) that enable efficient multiple repairs and chain validation. We have two such data structures: *repair layer Rdb* database and *approved repairs Adb* database. Every block is associated with its own repair layer Rdb database entry that comes into play when the block contents are repaired (directly or indirectly). For repairs that are not redactions,

¹⁰If a candidate transaction does not have a corresponding **repairTx** in the blockchain then the transaction is not included in the candidate pool, and it is treated as spam instead.

Rdb of the block contains the old version `Tx` of the transaction that was repaired. In case of redactions, as discussed in Section 3.2, the policy \mathbb{P} only permits non-payment data to be redacted. These are data whose changes do not affect the state of accounts. In this case the hash of the old version of the transaction $H(\text{Tx})$ is stored in the `Rdb` of the block. In the formal description Figure 5 the entire set of old transactions and state of accounts is stored in the repair layer. We emphasize that this covers the possibility of many transactions in a block being repaired multiple times. We note that as a practical optimization, you only need the old version of the specific transaction to be stored. The *approved repairs* `Adb` database, stores the repair proposal that was approved by the policy at that time. This data structure plays a crucial role in chain validation. The information stored in these data structures do not need any special authentication, as they can be validated using simple hash equality checks: in case of information stored in `Rdb`, the `header` of a block stores the corresponding transaction root and state root, and in case of `Adb`, the chain stores the hash of the repair proposal (in the form of votes as explained later).

Repair Policy. We briefly discuss the repair policy \mathbb{P} for Ethereum with PoS that determines if a repair proposal has been approved or not. Although our voting based deliberation process is similar to the protocol in [25], the deliberation and corresponding policy in `Reparo` is much more complicated. Therefore, $\Gamma'.\text{chkApproval}$ in Figure 5 for the policy \mathbb{P} and a repair proposal returns `approve`, `reject` or `voting`: `approve` and `reject` means that the repair proposal has been approved and rejected, respectively, and `voting` means that the repair proposal is still in deliberation phase.

Policy \mathbb{P} takes in the information from the real world like user discussions, forum discussions, expert opinions, etc. to see if a particular repair proposal is good for the chain or not. In any case, we wish to give a minimum policy requirement for redactions and other repair operations which can later be updated depending on the application. The objective of this minimum requirement is to enable miners to detect malicious repair proposals that aim at double spending or stealing coins. We emphasize that this by no means is a complete set of requirements to detect such behavior, and in fact it may be of independent research interest to frame such policies for various applications. In other words, enforcement of the policy is not done on chain. The minimum requirements from the policy \mathbb{P} are: (1) The deliberation period for the request began at the start of the epoch and ended with the end of the same epoch. (2) The proposal does not propose to modify the address fields or the value field of a transaction. (3) The proposal does not redact or modify votes in the chain. (4) The proposal has received more than ρ fraction of votes (50% of votes) in the epoch (ℓ consecutive blocks which is voting period, and can be decided by the system) after the corresponding `repairTx` is included in the chain. And finally, (5) the proposal is (unambiguously) not a double spend attack attempt (which needs information from the real world for confirmation).

Deliberation by Voting. In the deliberation process, the slotleaders vote for a repair proposal by generating a voting transaction `voteTx` and including that in the block that they propose. If the votes received is approved according to `chkApproval` with policy \mathbb{P} , slotleaders consider these votes as the witness w (as in Figure 2). The `Tx.to` address field of the `voteTx` is a special address `VOTE_ADDR` and the `Tx.data` field contains the hash of the old transaction and the hash of the candidate transaction. Formally, we define the interface:

$H(G_{tx}(TX_j), G_{tx}(TX_j^*)) \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, rp_j^*)$: takes as input a repair proposal rp_j^* and outputs the hash value of the `Tx.data` field of the corresponding repair proposal as a vote.

Note that we use G_{tx} instead of H for TX_j and TX_j^* , as we deal with set of old and new transactions. The reason for using $H(G_{tx}(TX_j), G_{tx}(TX_j^*))$ stems from the need to allow redactions and other repairs to be performed on TX_j or TX_j^* , while later a new user can verify this without providing the original TX_j or TX_j^* itself.

Performing Repairs. Upon approval with respect to the policy \mathbb{P} :

- *Redactions:* These operations are restricted by the policy \mathbb{P} to not affect the state of accounts in Ethereum chain. Withstanding this restriction, the original transaction is replaced by the candidate transaction and the repair layer stores the hash of the old version of the transaction. We use `retainAndRedact(TX_j)` function, which returns the hash of the redacted transaction and all other unedited transactions in TX_j in the original form Algorithm 2. If a version of the transaction is stored in either `Rdb` or `Adb`, it is redacted too ensuring that a redacted transaction is not stored even in the `Reparo` layer.

- *Other repairs:* For other repairs, the old version of the transaction is stored in the repair layer `Rdb` associated with the block. The candidate transaction then replaces the old version in the block. The state of accounts (for this block and the following blocks) is updated according to the repaired transactions. When updating the state for each block, we ensure that the original state of accounts is stored in the corresponding repair layer `Rdb`. Once the state updates reach the head of the chain, the slotleader proposes a new block with this updated state of accounts. Algorithm 2 gives a formal description, where entire set of original transactions and state are stored in the repair layer (thus covering the possibility of multiple repair operations on a block). For improving space efficiency one could store only the old version of the transaction.

Note that, since a repaired block always contains the most recent state, performing multiple indirect state updates is efficient as we only apply the transition function over the block's latest contents during each of the state updates.

Algorithm 2: repairChain

```

input : Chain  $\mathcal{C} = (B_1, \dots, B_n)$  of length  $n$ , repair layer  $\text{Rdb} = (\text{Rdb}_1, \dots, \text{Rdb}_n)$ , and a repair
         proposal  $rp_j^*$ .
output: Chain  $\mathcal{C}'$ , repair layer  $\text{Rdb}'$ 
1: Parse  $B_j := \langle \text{header}_j, TX_j \rangle (ACC_j)$ ,  $rp_j^* := \langle TX_j^* \rangle$ ;
2: Set  $B_j^* \leftarrow \langle \text{header}_j, TX_j^* \rangle$ ;
    $\triangleright$  If block is never repaired then store original state.
3: if  $\text{Rdb}_j = \emptyset$  then if  $rp_j^*$  is a redaction proposal, set  $\text{Rdb}_j \leftarrow \langle \text{retainAndRedact}(TX_j) || ACC_j \rangle$ ,
   otherwise set  $\text{Rdb}_j \leftarrow \langle TX_j \rangle$ ;
4: else Parse  $\text{Rdb}_j := TX_j'$ , and if  $rp_j^*$  is a redaction proposal, set
    $\text{Rdb}_j \leftarrow \langle \text{retainAndRedact}(TX_j') || ACC_j' \rangle$ ;
5: Initialize  $\mathcal{C}' \leftarrow \mathcal{C}^{j-1} || B_j^*$ , and  $\text{Rdb}' \leftarrow \text{Rdb}^{j-1} || \text{Rdb}_j$ ;
    $\triangleright$  Update repair layer of blocks in between
6: for  $i = j + 1$  to  $n$  do
7:   Initialize  $\text{Rdb}_i^* = \emptyset$ ;
8:   Parse  $B_i := \langle \text{header}_i, TX_i \rangle (ACC_i)$ ;
9:   if  $i - 1 = j$  then Set  $B_{i-1} = B_j^*$ ;
10:  Parse  $B_{i-1} := \langle \text{header}_{i-1}, TX_{i-1} \rangle (ACC_{i-1})$ ;
11:  Set  $TX_i^* \leftarrow TX_i$  and  $ACC_i^* \leftarrow \delta(ACC_{i-1}, TX_i)$ ;
12:  Set the block  $B_i^* \leftarrow \langle \text{header}_i, TX_i^* \rangle$ ;
13:  if  $\text{Rdb}_i = \emptyset$  then  $\text{Rdb}_i^* \leftarrow (TX_i)$ ;
14:  Set  $\text{Rdb}' \leftarrow \text{Rdb}' || \text{Rdb}_i^*$ , and  $\mathcal{C}' \leftarrow \mathcal{C}' || B_i^*$ ;
15: return  $\mathcal{C}'$ ,  $\text{Rdb}'$ ;

```

Block Validation. A formal description of the procedure can be found in Algorithm 3 which is invoked during the chain validation. The procedure checks if the transactions included in the block are valid as done currently in Ethereum. It then checks if the hash link is rightly formed. In case no repair proposal has been approved in this block, the only remaining checks are to see if the state of accounts in the block are correct and if the slotleader has produced a valid

Algorithm 3: validateBlock

input : Chain $\mathcal{C} = (B_1, \dots, B_n)$, repair layer $\text{Rdb} = (\text{Rdb}_1, \dots, \text{Rdb}_n)$, block B_{n+1} , repair approved Adb_{n+1} .
output: $\{\perp, (\mathcal{C}', \text{Rdb}')\}$

- 1: Parse $B_{n+1} := \langle \text{header}_{n+1}, TX_{n+1} \rangle (ACC_{n+1})$, where $\text{header}_{n+1} = (pt_{n+1}, G(TX_{n+1}), hd_{n+1})$;
- 2: Parse $B_n := \langle \text{header}_n, TX_n \rangle (ACC_n)$, where $\text{header}_n = (pt_n, G(TX_n), hd_n)$;
- 3: Validate transactions TX_{n+1} , if *invalid* **return** \perp ;
- 4: **if** $pt_{n+1} \neq H(\text{header}_n)$ **then return** \perp ;
- 5: **if** $\text{Adb}_{n+1} = \emptyset \wedge ACC_{n+1} = \delta(ACC_n, TX_{n+1}) \wedge \text{vfy_pos}(\mathcal{C}, hd_{n+1}, (pt_{n+1}, G(TX_{n+1}))) = 1$ **then**
Set $\mathcal{C}' \leftarrow \mathcal{C} || B_{n+1}$, and $\text{Rdb}' \leftarrow \text{Rdb} || \emptyset$, and **return** $(\mathcal{C}', \text{Rdb}')$;
▷ Validate Blocks where repairs were approved
- 6: Initialize $\mathcal{C}' \leftarrow \mathcal{C}$, $\text{Rdb}' \leftarrow \text{Rdb}$;
- 7: **for all** $TX_j^* \in \text{Adb}_{n+1}$ **do**
▷ Perform all the repair operations
- 8: Parse $B_j := \langle \text{header}_j, TX_j \rangle (ACC_j)$, where $\text{header}_j = (pt_j, G(TX_j'), hd_j)$;
- 9: **if** $\text{chkApproval}(\mathbb{P}, H(G_{tx}(TX_j), G_{tx}(TX_j^*))) \neq \text{approve}$ **then return** \perp ;
- 10: Set $ACC_j^* := \delta(ACC_{j-1}, TX_j^*)$;
▷ Perform the repairs as originally performed
- 11: $\mathcal{C}', \text{Rdb}' \leftarrow \text{repairChain}(\mathcal{C}', \text{Rdb}', rp_j^*)$;
- 12: Parse $\mathcal{C}' := (B'_1, \dots, B'_n)$ and $B'_n := \langle \text{header}'_n, TX'_n \rangle$;
▷ Check the state transition after repair
- 13: **if** $ACC_{n+1} = \delta(ACC'_n, TX_{n+1}) \wedge \text{vfy_pos}(\mathcal{C}, hd_{n+1}, (pt_{n+1}, G(TX_{n+1}))) = 1$ **then** Set
 $\mathcal{C}' \leftarrow \mathcal{C} || B_{n+1}$, and $\text{Rdb}' \leftarrow \text{Rdb} || \emptyset$ and **return** $(\mathcal{C}', \text{Rdb}')$;
- 14: **return** \perp ;

proof of stake. If any repair proposals were approved by the policy at this block, the procedure performs these repairs on the chain while performing the required state updates for the blocks. After all the approved repairs until this block have been applied, the procedure checks if the state of accounts in the block under contention are consistent with the updated states of the previous blocks. Finally, it checks if the header is correctly formed with the correct stakeholder as the slotleader.

Chain Validation. On receiving a new chain, the chain validation procedure formally described in Algorithm 4 starts validating the blocks from the genesis of the chain. It first switches the block contents with the corresponding transactions and states stored in the repair layer. This results in the chain (\mathcal{C}_{org}) in its originally mined state. The procedure then validates each block as discussed above using Γ' .validateBlock. This results in performing the repairs (both redactions and other repair operations) as they were performed in sequence. We then obtain a chain in its updated current state and is checked if it is the one that was received.

4.3 On Security and Optimizations

We discuss briefly Reparo's security, and other optimizations possible for our Ethereum instantiation.

Security Since the hash function H is modeled as a random oracle (RO), finding a collision on a vote (which is the hash of the ID of the old transaction and the ID of the candidate transaction) is highly improbable. Therefore, when a slotleader votes for a repair proposal in his block, no adversary can claim a different repair proposal for the same vote value. Similarly no adversary can find a different block that hashes to the same hash of an honestly proposed

Algorithm 4: validateChain

input : Chain $\mathcal{C} = (B_1, \dots, B_n)$ of length n , repair layer $\text{Rdb} = (\text{Rdb}_1, \dots, \text{Rdb}_n)$ and approved repairs $\text{Adb} = (\text{Adb}_1, \text{Adb}_2, \dots, \text{Adb}_n)$.
output: $\{0, 1\}$

- 1: Initialize $\mathcal{C}_{org} \leftarrow B_1$;
- 2: **for** $j = 2$ **to** n **do**
- 3: Parse $B_j := \langle \text{header}_j, TX_j \rangle (ACC_j)$, where $\text{header}_j = (pt_j, G(TX'_j), hd_j)$;
- 4: Parse $\text{Rdb}_j := (TX^* || ACC^*)$;
- 5: **if** $TX^* || ACC^* = \emptyset$ **then** $TX_j^* \leftarrow TX_j$;
- 6: **else** $TX_j^* || ACC_j^* \leftarrow TX^* || ACC^*$;
 \triangleright In case of redactions, we have $TX^* = \text{retainAndRedact}(TX_j)$, from which the original transaction merkle root $G_{tx}(TX^*)$ can be computed
- 7: **if** $G_{tx}(TX'_j) \neq G_{tx}(TX_j^*) \vee G_{st}(ACC'_j) \neq G_{st}(ACC_j^*)$ **then return 0** ;
- 8: $\mathcal{C}_{org} \leftarrow \mathcal{C}_{org} || \langle \text{header}_j, TX_j^* \rangle$;
- 9: Initialize $\text{Rdb}^* \leftarrow \emptyset$;
- 10: Parse $\mathcal{C}_{org} := (B_1^{org}, \dots, B_n^{org})$;
 \triangleright Validate each block starting at genesis
- 11: **for** $j = 2$ **to** n **do**
- 12: Set $op \leftarrow \text{validateBlock}(\mathcal{C}_{org}^{j-1}, \text{Rdb}^*, B_j^{org}, \text{Adb}_j)$;
- 13: **if** $op = \perp$ **then return 0** ;
- 14: **else** Parse $op := (\mathcal{C}', \text{Rdb}')$;
- 15: Set $\mathcal{C}_{org} \leftarrow \mathcal{C}' ||^j \mathcal{C}_{org}$, $\text{Rdb}^* \leftarrow \text{Rdb}'$;
- 16: **if** $\mathcal{C}_{org} = \mathcal{C} \wedge \text{Rdb}^* = \text{Rdb}$ **then return 1** ;
- 17: **return 0**

block. Therefore an adversary cannot break the integrity of the chain. Together, they imply the *unforgeability* of votes: if an adversary wishes to vote, he has to possess enough stake to propose a block with his vote himself. Assuming appropriate threshold on adversarial stakes and the honest stakeholders follow the policy \mathbb{P} , Reparo integration satisfies *editable common prefix* and preserves chain quality and chain growth.

Effect on Stake Distribution. Reparo’s repair operations, like fixing smart contract bugs, affects the balances of users and hence the stake distribution is altered. During the deliberation phase, it is the stakeholders who vote for a request fully aware of how the stake distribution change affects them. Assuming rational slotleaders and honest majority in the stakeholders, a slotleader votes for those repair requests that are not obvious double spend attacks and has least negative impact on his stake. The honest behaviour is enforced through public verifiability of a slotleader’s votes.

Optimizations. To lower the costs of repair operations in Ethereum, we propose *Depth-based future approval*: Depending on the depth d of the contract that needs repairs, the system can have a parameter p that integrates the fix into the main chain in block number d/p after approval. For example, at block number 8M if a contract deployed at block number 1M was found to have a vulnerability, then with $d = 7M$, $p = 1000$, the fix will be integrated into the chain 7,000 blocks after the corresponding `repairTx` is approved. This alleviates the computational load on the network by giving them more time to perform repairs that are deep in the chain. Few other optimizations are discussed in Appendix D.

5 Experiments in Ethereum

In this section, we report a proof-of-concept implementation of the `Reparo` protocol on top of Ethereum [4].

We implement two new types of transactions, namely `repairTx` and `voteTx`, and measure their performance with respect to a baseline transaction in Ethereum. We also measure the overhead of implementing these special transactions on the Ethereum main network by measuring the time taken to import the *latest* 2 million blocks. We measure the time taken to import the blockchain because these introduce overheads for syncing (fully/partially) with the network (see Table 4).

In Ethereum, computation is measured in terms of the gas it needs to run the transaction in the Ethereum Virtual Machine (EVM). Hence, we take a look at the gas costs to repair (by fixing) some popular bugs by computing the transaction dependency graph for the contract creation transaction for these bugs. We estimate the gas cost to re-run all the dependent transactions and provide real-world numbers on the cost of such repairs in Table 6.

Setup and System Configurations. We modify the Go client for Ethereum (`geth`) for our experiments. We use the version `1.9.0-unstable-2388e425-20190528` from the official Github repository as the base version. We set the `geth` cache size to 10,000 MB and disable the P2P discovery (using the `--nodiscover` flag). The import was done using an export file consisting of blocks from block number 6,000,000 to 8,038,219 (latest block as of Jun 27, 2019) created by the export command from a fully synced node.

Our experiments employed the following hardware/software configuration: *CPU*: 24 core, 64-bit, Intel® Xeon® Silver 4116 CPU clocked at 2.10 GHz; *RAM*: 128 GB; *OS*: Ubuntu; *Kernel*: 4.15.0-47-generic.

Chain Selection. We choose a subset of the chain from block number 6,000,000 to 8,038,219 from the main network. This subset spans almost 1 year (approx. 342 days). It is representative because the chosen subset accounts for 49.75% of the total gas (and thus computation) in Ethereum. The first 6 million blocks have 13,472,636.72 million gas (Mgas) while the chosen subset has 13,339,193.15 Mgas. In terms of transaction volume, the first 6 million blocks have 273,900,932 transactions while the chosen subset has 209,746,714 transactions. This accounts for 43.37% of the transaction volume in Ethereum. In terms of block volume, the subset accounts for 25.36% of the whole chain. Another reason to consider this subset is the uniformity of the consensus algorithm. Ethereum uses different consensus rules depending on the block number (based on whether the block is a part of Homestead, Frontier or Metropolis release). The chosen subset has a mix of Byzantium (62.80%) and Constantinople (37.20%) forks, both of which are a part of the Metropolis release. The rules are similar apart from the reduced block rewards and EVM instruction optimizations in Constantinople.

System-Level Optimizations. We employ the following system-level optimizations in our implementation.

1. *Database choice for Reparo*: `geth` implements three types of key-value databases: *Memory Databases* which reside in the system memory, *Cached* and *Uncached Databases* which reside on the disk. The repair layer only stores active requests and the votes for these requests. Hence, a memory database is ideal to implement `repairTx` and `voteTx`.

2. *Native Contracts for repairTx and voteTx*: Native contracts are client-side implementations of functionalities that are too complex or expensive (in terms of gas) to be implemented inside the EVM. For example, the Ethereum yellow paper [41] uses native contracts to perform SHA3 and *ecrecover* (a function that returns the address from ECDSA signature values r, s). We use native contracts to support `Reparo`.

3. *Fast sync and light-client friendliness*: Fast sync is a mode used by the Ethereum clients.

In this mode, the clients download the entire chain but only retain the state entries for the recent blocks (pruning). In bandwidth, our implementation only needs to download $|\mathcal{C}| + m$ from full nodes, where m is the number of updates and the final space storage is still $|\mathcal{C}|$ as the nodes can discard the repair layer after syncing.

5.1 Special Transactions: `repairTx` , `voteTx`

`repairTx` and `voteTx` have special `to` addresses `REQ_ADDR = 0x09` and `VOTE_ADDR = 0x0D` respectively.

The transactions are always collected in the transaction pool. We modify the transaction pool logic, specifically `validateTx()`. After ensuring well-formedness of inputs, for `repairTx` we check that the `data` field is exactly 64 bytes long and the first 32 bytes correspond to the transaction hash of an existing transaction in the chain. For `voteTx`, we check that the `data` field contains exactly 32 bytes.

The input for `repairTx` consists of hash of the transaction $H(\text{Tx})$ which is to be repaired and the hash of the proposed new transaction $H(\text{Tx}^*)$. The validation logic ensures that `Tx` exists in the blockchain (repaired blockchain) by adding a new function `isTransactionTrue()`. In the implementation of the native code for this transaction, we add the request to the request memory database, indexed by $ID = H(H(\text{Tx})||H(\text{Tx}^*))$ and initialize it with 0 votes. This database is created on demand. The footprint of the database is small as we will need to process about 16,384 repair requests before occupying 1 MB. In contrast, the default cache memory used by the client ranges from 512 to 4096 MB depending on the client version and is therefore a safe assumption to make.

The input for `voteTx` is the `ID` described previously. The validation logic ensures that the input is well-formed (of correct length). In the implementation of the native code for this transaction, we check if the request exists in the request memory database. If found, it increments the vote by one. Otherwise, it throws an error and aborts the transaction.

To evaluate the performance overheads of the special transactions on the client (and the network), we compare it with a baseline transfer transaction involving a transfer of ETH between two accounts. The transfer function has the lowest gas requirements (21,000). `repairTx` (5.90% overhead) takes 76.09 ms and `voteTx` (0.055% overhead) takes 71.89 ms when compared to a transfer transaction which takes 71.85 ms on an average over 100 iteration. (Refer Table 4.)

5.2 Performing Repairs

In this series of experiments, we analyze the impact of supporting `Reparo` on client software. For every block, supporting `Reparo` adds an overhead of checking for approved repairs. If approvals are found, we repair the block body accordingly. In this section, we analyze the read-write overheads to support the repair, the cost of building new states and applying transaction

Table 4: Comparison of operations between the modified client and the unmodified client

Operation	Type	Client Type	
		Unmodified	Modified
<code>repairTx</code>	Time (ms)	-	76.09
<code>voteTx</code>	Time (ms)	-	71.89
Transfer	Time (ms)	71.85	71.85
Import	Time (Hours)	100.46	113.04
Import	Speed (Mgas/s)	37.42	33.08

Table 5: Contract creation transaction dependency graph for some prominent bugs. We trace the number of transactions that are dependent on the contract creation transaction. We also compute the total (maximum) number of accounts whose state could be affected if the concerned transaction were repaired.

Bug	Accts. Affected	Tx Affected
DAO	55.51 M	474.45 M
QCX	55.31 M	470.63 M
Parity	48.67 M	399.08 M
REXmls	52.44 M	437.21 M
No Code	2.98 K	2.98 K

Table 6: Estimated repair costs (today) using `Reparo`. m , K , M and T stand for milli(10^{-3}), Kilo(10^3), Mega(10^6) and Tera(10^{12}) multipliers respectively.

Bug	ETH Stuck	Costs of repair in <code>Reparo</code>		
		Tx Re-runs	Gas	ETH
DAO	3.60 M	474.45 M	30.17 M	30.17 K
QCX	67.32 K	20.77 K	805.89 M	0.80
Parity	517.34 K	1	9.39 M	94.00 m
REXmls	6.67 K	1.04 K	114.74 M	0.114
No Code	6.53 K	2.98 K	438.85 M	439.00 m

dependencies to repair some real-world bugs (check Table 7 for details about these bugs). We use an unedited (clean) chain for our experiments.

Read-Write Costs. In this experiment, we measure the time to update the data of a block. This experiment helps to estimate the I/O overheads of transaction updates in the blockchain. A repair consists of finding a transaction in the blockchain and replacing it with a new transaction. The transaction repair overhead consists of the time taken by a node to read the transaction metadata (block hash, block number and the transaction index in the block) and write the new transaction data. We point the old hash to the new transaction data so that when the hash of the old transaction is accessed, the repaired transaction is furnished by the blockchain. We measure the read and write times for 10,000 random transactions from random blocks in the chain. Random transactions ensure that internal (database, software or operating system) caches do not skew the measurements. The time taken to read the metadata is $649.81\mu\text{s}$ and the write operation takes 2.32 ms on average over 100 runs for each of the 10,000 transactions.

Import Costs. In this experiment, we evaluate the time it takes to import our chain subset using the modified and unmodified versions of the client. The `geth` client imports blocks in batches. For example, the first batch consists of 3 blocks, starting from block number 6,000,000, the next batch consists of 13 blocks and so on. We log the amount of gas (in million gas) in such batches and the time elapsed for the import (and thus compute the speed). We perform 3 iterations on both the modified and unmodified clients. We plot these speeds for the entire import process for the unmodified and modified clients in Figure 4. The average speed for the two clients are presented in Table 4. In Figure 4, we present the speed as the import progresses. As evident from the graph, for most of the parts the modified client is equal to or slightly slower than the unmodified client. This is reasonable in the real world as the slight import delay per block can be accounted for by reducing the gas limit of the block (and thus the computation performed on each block allowing `Reparo` to utilize the remaining time).

On average, the unmodified client takes 100.46 hours to import 2,038,219 blocks whereas our modified client takes 113.04 hours to import the same blocks. (Refer to Table 4.) This is just 12.52% overhead for a full import of more than 2 million blocks. It does not have any

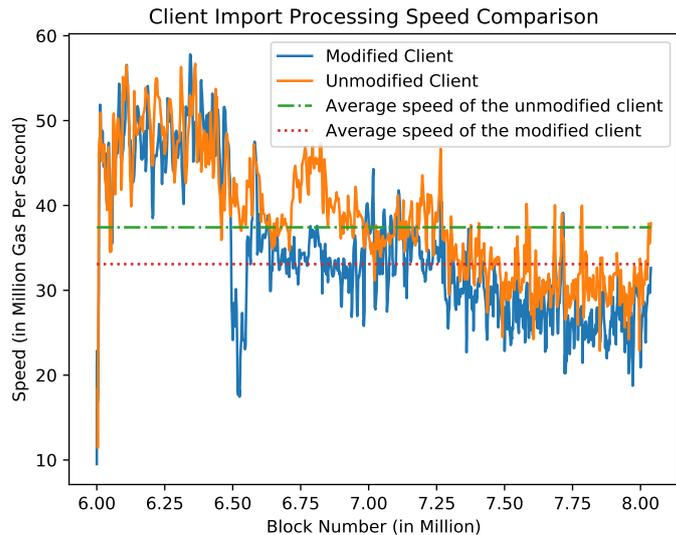


Figure 4: Batched speed comparison of the modified client and the unmodified client. The modified client has a modified `validateTx` rule, a new function `isTransactionTrue`, and modified structures with flags to detect a dirty (edited chain and blocks).

significant effect on the block generation, block validation or block propagation as this can be tweaked by reducing the difficulty and/or gas limit of the blocks.

The average amount of gas processed by the unmodified client is 37.42 million gas per second whereas the modified client processes 33.08 million gas per second (Table 4). This 11.59% overhead is due to the hard coding of rules for special transactions whose conditions are checked for every transaction. This overhead does not cause any problems as the average gas limit for an Ethereum block is 8,000,000 (which is under 33 Mgas/s) [3] and both the nodes perform optimally to sync the latest blocks and propagate. Note that this affects the full sync nodes only. Note that the light clients, such as Parity [6] for example, skip verification of states and are thus unaffected.

Transaction Dependency Graph. To estimate the amount of extra gas required to repair a transaction Tx, we compute transaction dependency graphs for contract creation transactions. We in fact measure the worst possible scenario when a transaction has an impact on all transactions that follow directly and indirectly from it. We perform this by marking the affected accounts and checking every transaction if they have a marked account in the `from` or `to` fields. We choose some of the popular bugs (described in Table 7) *DAO*, *Parity Multisig Wallet* bug, *QCX*, *REXmIs* and a class of bugs *No code*. The chosen contracts occur very deep in the blockchain giving us realistic estimates for repairs. Table 5 shows the cascading effects of changing transactions.

Repairs. We employ a policy which allows editing any contract call in order to repair the chain. We qualify our previous pessimistic analysis by arguing that most of the repairs have small transaction dependency graphs. This is due to the localization of impact to a few accounts. We bound the number of transactions that need to be re-run to transactions that interact with the contract. This coupled with the fact that we are performing a repair ensures a small transaction dependency graph which significantly reduces the repair costs. Table 6 we highlight the impact of such localizations. We sum the gas in all such transactions to estimate the gas cost of repairs and thus the ETH. Note that we always pay the miners (and hence the network) for the extra computation. We use a gas price of 1 GWei/gas (market price at the time of

writing) for our conversions. We refer the interested readers to Appendix D.3 for more details about the bugs and our solutions.

6 Conclusion and Future Work

This work presents **Reparo**, a secure, systematic way to make any blockchain forget the “forgettable”. We present a generic protocol that is adaptable to consensus requirements, and achieves public verifiability and secure chain repairs guaranteeing REC for current mainstream blockchains. We then design and analyze an important application of the protocol in Ethereum to fix contract bugs, and report the implications and feasibility of these repairs for popular contract bugs such as DAO and Parity Multi Sig Wallet. We also provide optimizations that can make the implementation more robust and realizable. We show that, in Ethereum, vulnerabilities, if found, (and existing vulnerabilities) can be immediately isolated to reduce the transaction dependency and repaired efficiently and securely.

In the future, we aim to realize the **Reparo** protocol on permissioned systems such as Hyperledger. We also intend to study the impact of **Reparo** on off-chain protocols and whether it can be used to improve them. Among other repair operations, **Reparo** also offers a means to propose, deliberate and incorporate new features into Bitcoin and Ethereum given the respective communities currently do this in an ad-hoc manner [1, 5].

Acknowledgment

We would like to thank Andrew Miller for his valuable comments and constructive feedback.

The first author was supported by the German research foundation (DFG) through the collaborative research center 1223, and by the state of Bavaria at the Nuremberg Campus of Technology (NCT). NCT is a research cooperation between the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Technische Hochschule Nürnberg Georg Simon Ohm (THN). This work also has been partially supported by the National Science Foundation under grant CNS-1846316.

References

- [1] bitcoin/bips: Bitcoin improvement proposals. <https://github.com/bitcoin/bips>. Accessed: 2019-07-31.
- [2] Coinmarketcap. <https://coinmarketcap.com>.
- [3] Ethereum average gas limit chart. <https://etherscan.io/chart/gaslimit>. Accessed: 2019-07-28.
- [4] Ethereum whitepaper. <https://github.com/ethereum/wiki/wiki/White-Paper>.
- [5] ethereum/eips: The ethereum improvement proposal repository. <https://github.com/ethereum/EIPs>. Accessed: 2019-07-31.
- [6] Getting synced. parity tech documentation. <https://wiki.parity.io/Getting-Synced>. Accessed: 2019-07-28.
- [7] Hundreds of millions of dollars locked at ethereum 0x0 address and smart contracts' addresses - how, why and what to do? <https://medium.com/@maltabba/hundreds-of->

millions-of-dollars-locked-at-ethereum-0x0-address-and-smart-contracts-addresses-how-4144dbe3458a.

- [8] Major issues resulting in lost or stuck funds. <https://github.com/ethereum/wiki/wiki/Major-issues-resulting-in-lost-or-stuck-funds>.
- [9] "reclaiming of ether in common classes of stuck accounts · issue 156 · ethereum/eips". <https://github.com/ethereum/EIPs/issues/156>. Accessed: 2019-07-31.
- [10] State tree pruning. <https://blog.ethereum.org/2015/06/26/state-tree-pruning/>.
- [11] Understanding the dao attack. <https://www.coindesk.com/understanding-dao-hack-journalists>.
- [12] Interpol cyber research identifies malware threat to virtual currencies. *Interpol*, 2015. <https://tinyurl.com/y9wfekr6>.
- [13] Downside of bitcoin: A ledger that can't be corrected. *The New York Times*, 2016. <https://tinyurl.com/ydxjlf9e>.
- [14] Giuseppe Ateniese, Michael T Chiaramonte, David Treat, Bernardo Magri, and Daniele Venturi. Rewritable blockchain, May 8 2018. US Patent 9,967,096.
- [15] Giuseppe Ateniese, Bernardo Magri, Daniele Venturi, and Ewerton Andrade. Redactable blockchain—or—rewriting history in bitcoin and friends. In *Security and Privacy (EuroS&P), 2017 IEEE European Symposium on*, pages 111–126. IEEE, 2017.
- [16] Christian Badertscher, Peter Gazi, Aggelos Kiayias, Alexander Russell, and Vassilis Zikas. Ouroboros genesis: Composable proof-of-stake blockchains with dynamic availability. In *ACM CCS 18: 25th Conference on Computer and Communications Security*, pages 913–930. ACM Press, 2018.
- [17] Christian Badertscher, Ueli Maurer, Daniel Tschudi, and Vassilis Zikas. Bitcoin as a transaction ledger: A composable treatment. *Lecture Notes in Computer Science*, pages 324–356. Springer, Heidelberg, 2017.
- [18] Lorenz Breidenbach, Philip Daian, Florian Tramèr, and Ari Juels. Enter the hydra: Towards principled bug bounties and exploit-resistant smart contracts. pages 1335–1352, 2018.
- [19] Vitalik Buterin. ethereum/wiki - proof of stake faq. <https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ>.
- [20] Vitalik Buterin and Virgil Griffith. Casper the friendly finality gadget. *arXiv preprint arXiv:1710.09437*, 2017.
- [21] Jan Camenisch, David Derler, Stephan Krenn, Henrich C Pöhls, Kai Samelin, and Daniel Slamanig. Chameleon-hashes with ephemeral trapdoors. In *IACR International Workshop on Public Key Cryptography*, pages 152–182. Springer, 2017.
- [22] Bernardo David, Peter Gazi, Aggelos Kiayias, and Alexander Russell. Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. *Lecture Notes in Computer Science*, pages 66–98. Springer, Heidelberg, 2018.

- [23] David Derler, Kai Samelin, Daniel Slamanig, and Christoph Striecks. Fine-grained and controlled rewriting in blockchains: Chameleon-hashing gone attribute-based. In *ISOC Network and Distributed System Security Symposium – NDSS 2019*. The Internet Society, 2019.
- [24] Dominic Deuber, Nico Döttling, Bernardo Magri, Giulio Malavolta, and Sri Aravinda Krishnan Thyagarajan. Minting mechanisms for blockchain – or – moving from cryptoassets to cryptocurrencies. Cryptology ePrint Archive, Report 2018/1110, 2018. <https://eprint.iacr.org/2018/1110>.
- [25] Dominic Deuber, Bernardo Magri, and Sri Aravinda Krishnan Thyagarajan. Redactable blockchain in the permissionless setting. In *2019 IEEE Symposium on Security and Privacy, SP 2019, Proceedings, 20-22 May 2019, San Francisco, California, USA*, 2019.
- [26] Martin Florian, Sophie Beaucamp, Sebastian Henningsen, and Björn Scheuermann. Erasing data from blockchain nodes. *arXiv preprint arXiv:1904.08901*, 2019.
- [27] Juan A. Garay, Aggelos Kiayias, and Nikos Leonardos. The bitcoin backbone protocol: Analysis and applications. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology – EUROCRYPT 2015, Part II*, volume 9057 of *Lecture Notes in Computer Science*, pages 281–310. Springer, Heidelberg, April 2015.
- [28] Yossi Gilad, Rotem Hemo, Silvio Micali, Georgios Vlachos, and Nickolai Zeldovich. Algorand: Scaling byzantine agreements for cryptocurrencies. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 51–68. ACM, 2017.
- [29] LM Goodman. Tezos—a self-amending crypto-ledger. URL: [https://www.tezos.com/static/papers/white paper.pdf](https://www.tezos.com/static/papers/white%20paper.pdf), 2014.
- [30] Luis-Daniel Ibanez, Kieron O’Hara, and Elena Simperl. On blockchains and the general data protection regulation. 2018.
- [31] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. Ouroboros: A provably secure proof-of-stake blockchain protocol. In *Annual International Cryptology Conference*, pages 357–388. Springer, 2017.
- [32] Hugo Krawczyk and Tal Rabin. Chameleon signatures. In *ISOC Network and Distributed System Security Symposium – NDSS 2000*. The Internet Society, February 2000.
- [33] Bernardo Magri, Christian Matt, Jesper Buus Nielsen, and Daniel Tschudi. Afgjort – a semi-synchronous finality layer for blockchains. Cryptology ePrint Archive, Report 2019/504, 2019. <https://eprint.iacr.org/2019/504>.
- [34] Roman Matzutt, Jens Hiller, Martin Henze, Jan Henrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle. A quantitative analysis of the impact of arbitrary blockchain content on bitcoin. In *Proceedings of the 22nd International Conference on Financial Cryptography and Data Security (FC)*. Springer, 2018.
- [35] Brian McCurtis. “redactable blockchains” are not blockchains! <https://medium.com/@brianmccurtis/redactable-blockchains-are-not-blockchains-914b91569e62>.
- [36] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

- [37] Rafael Pass, Lior Seeman, and Abhi Shelat. Analysis of the blockchain protocol in asynchronous networks. In *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part II*, pages 643–673, 2017.
- [38] Eugenia Politou, Fran Casino, Efthimios Alepis, and Constantinos Patsakis. Blockchain mutability: Challenges and proposed solutions. *arXiv preprint arXiv:1907.07099*, 2019.
- [39] Ivan Puddu, Alexandra Dmitrienko, and Srdjan Capkun. μ chain: How to forget without hard forks. *IACR Cryptology ePrint Archive*, 2017:106, 2017.
- [40] Giannis Tziakouris. Cryptocurrencies—a forensic challenge or opportunity for law enforcement? an interpol perspective. *IEEE Security & Privacy*, 16(4):92–94, 2018.
- [41] Gavin Wood et al. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151:1–32, 2014.

A Detailed Related Work

To solve the problem of illicit data stored in blockchains, Ateniese et al. [15] proposed the first redactable blockchain protocol which uses chameleon hash links [21]. Their solution is catered to the permissioned setting where select miners can come together and redact contents from the blockchain using a large scale MPC protocol. Unfortunately, their large scale MPC in a dynamic entry-exit permissionless setting seems to make it infeasible for their proposal to solve the above discussed problems in Bitcoin and Ethereum. Their proposal apart from requiring modifications to the block header structure, does not work with SHA256 and requires chameleon hashes. Therefore the protocol is not backward compatible to any of the existing chains like Bitcoin or Ethereum.

Puddu et al. [39] proposed a protocol where senders encrypt all but one version of their transactions to the miners along with a *mutation* policy. The un-encrypted version remains the valid version on the chain. The miners abiding by the policy can decrypt (via MPC if decryption keys are shared) alternate versions and make those versions valid. However, a malicious sender may not include any alternate version at all or may have a mutation policy where only he can make retrieve the alternate versions. Moreover, similar to [15], this proposal too suffers from scalability issues with large scale MPC in a permissionless setting and is not backward compatible with existing chains.

Derler et al. [23] proposed attribute based modification of chain contents while relying on chameleon hashes. Unlike in [15], here chameleon hashes are not used for hash links but for transaction hashes in computing the merkle root. Any user can tag an object with an access policy before posting it on the blockchain and only the users with attributes satisfying this policy can later decide to modify the object. However, in their setting changing from the old version to the new version of the transaction does not affect the transaction merkle root. Therefore a user cannot decide whether and where something was changed or not. This creates problems when the underlying consensus mechanism is PoS that relies on the state (i.e., accounts and balances) of the chain. A new user can no longer verify the consensus that was generated with the old version of the transaction as it does not exist anymore. In other words, the proposal lacks accountability as the rewritings are indistinguishable and lacks verifiability of (state dependent) consensus with respect to new users. Also, similar to [39] this proposal relies on the user to set an attribute policy which may not be useful if the user sets a policy that only his own or his colluder’s attributes can satisfy.

Deuber, Magri and Thyagarajan [25] proposed the first efficient redactable protocol for a permissionless setting. They rely on achieving voting based consensus to perform redactions. The block structure is modified to have two hash links instead of one and if the previous block is redacted, one of the links breaks while the other holds. This gives accountability and public verifiability of where and what was redacted from the lens of a new user unlike the above mentioned proposals. They do not make use of any heavy cryptographic machinery and therefore achieve better efficiency in the permissionless setting. As their focus is only to redact illicit contents that does not affect payment information, their protocol fails to deal with stateful edits like doing smart contract bug fixes in Ethereum that have a cascading effect. Since their protocol is tailor made for PoW based systems, it is unclear how they do stateless or stateful edits in PoS based systems. And finally, their protocol is not backward compatible with any existing blockchains given their requirement of the block structure modification. For a more detailed comparison with [25], we refer the reader to Appendix E.1.

Florian et al. [26] propose for miners to locally drop harmful data. Although efficient in case of Bitcoin, they do not focus on global consensus on what to be erased. Differing miners end up in different forks which severely limits their functionality, which is aggravated in case of stateful edit operations. Politou et al. [38] present a comprehensive survey of the various solutions that have been proposed to edit a blockchain and also give the relevance of GDPR laws for blockchains.

Tezos [29] proposed a generic and self-amending blockchain. They provide a generic interface for meta-upgrades, i.e changes to the code. The interface is generic and can be instantiated on any blockchain such as Bitcoin [36] and Ethereum [41]. Tezos creates a testnet with the proposed changes/amendments and if there is sufficient confidence in the testnet (via votes from stakeholders) promotes the testnet as the main protocol.

This scheme has several drawbacks. Tezos can instantiate any blockchain protocol by using the appropriate genesis block. However, Tezos cannot be instantiated on an existing blockchain. In other words, Tezos cannot be used to repair existing blockchains.

Another drawback of this scheme is that at any point in time, only one proposal is under consideration (by being in the testnet). The proposal under consideration is always the one with the most approvals. Even if all the nodes in the system agree to the change, it takes a minimum of four quarters and two rounds of voting to integrate the change. This is inefficient when compared to our proposed scheme.

B Security Definitions

The *common-prefix* property states that if one take the chains of two honest users at distinct time slots, the shorter chain (minus a few blocks) is a prefix of the longer chain. This property implies the immutability of the underlying blockchain Γ . *Chain growth* property intuitively says that the chain \mathcal{C} will eventually grow in number of blocks as the protocol progresses. The *chain quality* property says that the ratio of blocks produced by malicious users in the chain \mathcal{C} can be upper bounded.

Definition 1 (Common Prefix [27]). *The chains $\mathcal{C}_1, \mathcal{C}_2$ possessed by two honest parties at the onset of the slots $sl_1 < sl_2$ are such that $\mathcal{C}_1^{\lceil k} \preceq \mathcal{C}_2$, where $\mathcal{C}_1^{\lceil k}$ denotes the chain obtained by removing the last k blocks from \mathcal{C}_1 , where $k \in \mathbb{N}$ is the common prefix parameter.*

Definition 2 (Chain Growth [27]). *Consider the chains $\mathcal{C}_1, \mathcal{C}_2$ possessed by two honest parties at the onset of two slots sl_1, sl_2 , with sl_2 at least s slots ahead of sl_1 . Then it holds that $\text{len}(\mathcal{C}_2) - \text{len}(\mathcal{C}_1) \geq \tau \cdot s$, for $s \in \mathbb{N}$ and $0 < \tau \leq 1$, where τ is the speed coefficient.*

Definition 3 (Chain Quality [27]). *Consider a portion of length ℓ -blocks of a chain possessed by an honest party during any given round, for $\ell \in \mathbb{N}$. Then, the ratio of adversarial blocks in this ℓ segment of the chain is at most μ , where $0 < \mu \leq 1$ is the chain quality coefficient.*

C Security Analysis

In this section we formally argue the security properties of the repairable blockchain $\Gamma'_{\mathbb{P}}$ resulting from the composition of an immutable blockchain Γ and our repair layer **Reparo** in the presence of a *valid* policy \mathbb{P} . By validity of \mathbb{P} we mean that the policy satisfies the minimum requirements listed above.

Recall that the underlying blockchain Γ is assumed to satisfy the security properties of *Chain growth*, *Chain quality* and *common prefix*, formally stated in Appendix B. Also, note that the assumptions of the underlying blockchain Γ must still hold (e.g., trusted majority), and in particular this means that in a PoS blockchain the majority of the stake must be in the hands of honest users during the entire lifetime of the system. We show that the protocol Γ' still *preserves* chain growth and chain quality. By preservation of the property we mean that the resulting protocol has *at least* the same guarantees as the original protocol Γ , but potentially stronger.

Chain Growth. Assuming that Γ satisfies chain growth it is not hard to see that the **Reparo** added on top of Γ *does not* influence the chain growth rate of the resulting protocol Γ' , as **Reparo** does not dictate how often new blocks are created and appended to the chain. We give the corollary statement below.

Corollary 1. *If Γ satisfies (τ, s) -chain growth, then $\Gamma'_{\mathbb{P}}$ preserves (τ, s) -chain growth for any valid policy \mathbb{P} .*

Chain Quality. Interestingly, when we assume that the majority of the deciders are honest and a majority endorsement is required for **decision** to output a witness, the **Reparo** protocol can potentially “improve” the chain quality coefficient of the resulting repairable blockchain Γ' . To see this, consider an adversarially produced block $B_i \in \mathcal{C}$. A repair operation rp proposed and accepted for block B_i could be seen as “turning” the block B_i into an *honest* block since the contents of B_i are now agreed by the protocol. This is because for the repair to be performed, it needs to be accepted by **decision**(rp) which needs a majority of the deciders (i.e., miners in the case of a permissionless setting) to endorse. Hence, by the honest majority assumption of the underlying blockchain Γ , any accepted repair operation must be backed by *at least* 1 honest miner (or more, depending on the policy \mathbb{P}), thereby increasing the ratio of honest blocks in \mathcal{C} .

Corollary 2. *For all witnesses $w_i \in \mathcal{C}$, let $pf \in w_i$ be a sound proof. If Γ satisfies (μ, ℓ) -chain quality, then $\Gamma'_{\mathbb{P}}$ preserves (μ, ℓ) -chain quality for any valid policy \mathbb{P} .*

Common Prefix. It can happen that two honest miners will perform the same repair operation at different times, and in the period in between it can happen that they *do not* have a common prefix. Note however, that a repair is only performed once its accepting repair witness is in the stable part of the chain. Hence, we do not have to deal with rollbacks. There are two observations to be made:

- This time period is small (i.e., 1 network delay). If the witness is stable for one miner then it must become stable for the other honest miners as soon as they see all the blocks that the first miner saw. Therefore, the repair operation might briefly disturb common prefix, but not for long.

- If a repair operation is a redaction that does not alter the state of the chain, the common prefix can be momentarily violated, but at no point it is violated when just considering the state of the chain.

Even though Γ' does not satisfy the common-prefix property as stated in Section B, following the lines of [25], we show that the protocol Γ' satisfies the *Editable common prefix* property introduced by [25].

Definition 4 (Editable Common prefix). *The chains $\mathcal{C}_1, \mathcal{C}_2$ of length l_1 and l_2 , respectively, possessed by two honest parties at the onset of the slots $sl_1 \leq sl_2$ satisfy one of the following:*

1. $\mathcal{C}_1^{[k]} \preceq \mathcal{C}_2$, or
2. for each $B_j \in \mathcal{C}_2^{[(l_2-l_1)+k]}$ such that $B_j \notin \mathcal{C}_1^{[k]}$, it must be the case that $\exists w_i \in \mathcal{C}_2^{[(l_2-l_1)+k]}$ such that $\text{chkApproval}(\mathbb{P}, w_i) = \text{approve}$ and $pt_i := H(B_j)$.

Here, $\mathcal{C}_2^{[(l_2-l_1)+k]}$ denotes the chain obtained by pruning the last $(l_2 - l_1) + k$ blocks from \mathcal{C}_2 , \mathbb{P} denotes the chain policy, repair witness w corresponds to a redaction proposal rp , pt_i is the pointer contained in w_i , and $k \in \mathbb{N}$ denotes the common prefix parameter.

Theorem 1. *If Γ satisfies k -common prefix, then $\Gamma'_{\mathbb{P}}$ satisfies k -editable common prefix for a valid policy \mathbb{P} .*

Proof. If no repair operations were performed in the chain \mathcal{C} , then the protocol $\Gamma'_{\mathbb{P}}$ behaves exactly like the protocol Γ . Henceforth the common prefix property follows directly.

However, in case of some repair operations, consider an adversary \mathcal{A} that proposes a repair $rp := \langle (H(\text{header}_i), x'_i), sp \rangle$ to repair contents of B_i in chain \mathcal{C}_2 . The proposal is later accepted and the repair witness $w = \langle H(\text{header}_i), H(rp), G(x'_i), sp, 1, pf \rangle$ is included in the chain which is then executed by an honest party P_2 at slot sl_2 . Observe that by the unforgeability property of the witness proof pf , \mathcal{A} is not able to efficiently produce a valid proof pf' for another repair proposal \hat{rp} that was not accepted. Therefore, since P_2 is honest and incorporated the repair rp in \mathcal{C}_2 , it must be the case that rp was accepted by *at least* the majority of the deciders. Thus making all the honest parties incorporate the repair rp . This concludes the proof. \square

D Reparo in Ethereum

Formal description of the Reparo protocol in Ethereum with a PoS consensus is given in Figure 5.

D.1 Reparo in Ethereum with PoW

We present here how one would instantiate Reparo in Ethereum with PoW based consensus done currently. We present here only the differences from the PoS instance we presented in Section 4.

Repair Policy.

A repair proposal is approved according to $\Gamma'.\text{chkApproval}$ with policy \mathbb{P} , if the following conditions hold:

- The proposal does not propose to modify the address fields or the value field of a transaction.
- The proposal is unambiguously not a double spend attack attempt (needs information from the real world for confirmation).

¹¹Ethereum uses Greedy Heaviest Order Sub Tree (GHOST) protocol to rank chains and this is slightly different from the longest chain rule used in Bitcoin.

The Ethereum protocol PoS consists of a sequence of rounds (slots) r and an epoch e consists of ℓ rounds.

Initialization. We initialize new databases the repair layer $\text{Rdb}_0 \leftarrow \text{genesis}$, and the approved repairs $\text{Adb}_0 \leftarrow \emptyset$, set round $r \leftarrow 1$ and an empty list of repair proposals $\text{propPool} \leftarrow \emptyset$.

For a given epoch e and for each round r , first initialize $\text{Adb}_r \leftarrow \emptyset$, $\text{Rdb}_r \leftarrow \emptyset$ and we describe the following sequence of execution.

Proposal. A node creates a repair proposal $rp_j^* \leftarrow \Gamma'.\text{proposeRepair}(\mathcal{C}, j, TX_j^*)$ (refer Algorithm 1) for block $B_j, j \in [r-1]$ using transactions TX_j^* . It then broadcasts it to the network.

Update Proposal pool. Collect all repair proposals rp_j^* from the network and add rp_j^* to propPool iff rp_j^* is valid; otherwise discard rp_j^* . If r is the beginning of an epoch, then set $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) := \text{voting}$ where v_j is a vote for rp_j^* .

Repairing the chain. For all repair proposals $rp_j^* := \langle TX_j^* \rangle \in \text{propPool}$, we denote a vote $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, rp_j^*)$ and do:

1. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{approve}$, then call algorithm $(\mathcal{C}', \text{Rdb}') \leftarrow \Gamma'.\text{repairChain}(\mathcal{C}, \text{Rdb}, rp_j^*)$ (refer Algorithm 2). Here j -th block in \mathcal{C}' is $\langle \text{header}_j, rp_j^* \rangle$ and subsequent blocks' states are updated accordingly. Then do the following,
 - (a) Add TX_j^* to Adb_r and remove rp_j^* from propPool
 - (b) set local chain $\mathcal{C} = \mathcal{C}'$ and update $\text{Rdb} = \text{Rdb}'$
2. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{reject}$, then remove rp_j^* from propPool
3. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{voting}$, then do nothing

Deliberation process. For all repair proposals $rp_j^* \in \text{propPool}$ satisfying $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{voting}$ (where v_j is vote for rp_j^*), that the node is willing to endorse,

1. Parse the proposal $rp_j^* := \langle TX_j^* \rangle$
2. Generate $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, s_j^*)$. Create a vote transaction voteTx with $\text{voteTx.data} = v_j$
3. Broadcast voteTx .

Proposing a new block. Collect all transactions, denoted by TX from the network for the r -th round and try to build a new block B_r :

1. (*Determine state transition from the head of the chain*). Repair the chain by applying the repair proposals that are approved: $\forall rp_j^* = TX_j^*$ such that $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{approve}$, where $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, rp_j^*)$, set $\text{ACC} = \delta(\text{ACC}_{r-1}, TX)$.
2. (*Proof of Stake*). Extend chain and Reparo data structures as follows,
 - (a) Let Acc^* be the account of the stakeholder, generate $\sigma \leftarrow \text{prf_pos}(\mathcal{C}, \text{Acc}^*, (H(\text{header}_{r-1}), G(TX)))$
 - (b) Set $hd := (\sigma, \text{Acc}^*)$ and complete header by appropriately setting other values
 - (c) Set new block $B_r \leftarrow \langle \text{header}, TX \rangle$
 - (d) Extend local chain $\mathcal{C} \leftarrow \mathcal{C} || B_r$, the repair layer $\text{Rdb} \leftarrow \text{Rdb} || \text{Rdb}_r$ and the approved repairs $\text{Adb} \leftarrow \text{Adb} || \text{Adb}_r$
 - (e) Then broadcast $(\mathcal{C}, \text{Rdb}, \text{Adb})$ to the network

Updating the chain. When a node receives \mathcal{C} , Rdb , and Adb , check if the chain is valid by calling $\Gamma'.\text{validateChain}(\mathcal{C}, \text{Rdb}, \text{Adb}) = 1$. Accept the new chain if the new chain is valid as per PoS's fork resolution rule.

Figure 5: Reparo protocol integration into Ethereum with Algorand or Ouroboros Praos as the underlying consensus and parameterized by policy \mathbb{P} . Meaning prf_pos is instantiated with Algorand or Praos.

The Ethereum protocol consists of a sequence of rounds r .

Initialization. We initialize new databases the repair layer $\text{Rdb}_0 \leftarrow \text{genesis}$, and the approved repairs $\text{Adb}_0 \leftarrow \emptyset$, set round $r \leftarrow 1$ and an empty list of repair proposals $\text{propPool} \leftarrow \emptyset$.

For each round r , first initialize $\text{Adb}_r \leftarrow \emptyset$, $\text{Rdb}_r \leftarrow \emptyset$ and we describe the following sequence of execution.

Proposal. A node creates a repair proposal $rp_j^* \leftarrow \Gamma'.\text{proposeRepair}(\mathcal{C}, j, TX_j^*)$ (refer Algorithm 1) for block $B_j, j \in [r-1]$ using transactions TX_j^* . It then broadcasts it to the network.

Update Proposal pool. Collect all repair proposals rp_j^* from the network and add rp_j^* to propPool iff rp_j^* is valid; otherwise discard rp_j^* .

Repairing the chain. For all repair proposals $rp_j^* := \langle TX_j^* \rangle \in \text{propPool}$, we denote a vote $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, rp_j^*)$ and do:

1. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{approve}$, then call algorithm $(\mathcal{C}', \text{Rdb}') \leftarrow \Gamma'.\text{repairChain}(\mathcal{C}, \text{Rdb}, rp_j^*)$. Here j -th block in \mathcal{C}' is $\langle \text{header}_j, rp_j^* \rangle$ and subsequent blocks' states are updated accordingly. Then do the following,
 - (a) Add TX_j^* to Adb_r and remove rp_j^* from propPool
 - (b) set local chain $\mathcal{C} = \mathcal{C}'$ and update $\text{Rdb} = \text{Rdb}'$
2. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{reject}$, then remove rp_j^* from propPool
3. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{voting}$, then do nothing

Mining a new block. Collect all transactions, denoted by TX from the network for the r -th round and try to build a new block B_r :

1. (*Deliberation process*). For all repair proposals $rp_j^* \in \text{propPool}$ that the node is willing to endorse,
 - (a) Parse the proposal $rp_j^* := \langle TX_j^* \rangle$
 - (b) Generate $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, s_j^*)$. If $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{voting}$ then create a vote transaction voteTx with $\text{voteTx.data} = v_j$
 - (c) Update $TX \leftarrow TX || \text{voteTx}$
2. (*Determine state transition from the head of the chain*). Repair the chain by applying the repair proposals that are approved: $\forall rp_j^* = \langle TX_j^* \rangle$ such that $\Gamma'.\text{chkApproval}(\mathbb{P}, v_j) = \text{approve}$, where $v_j \leftarrow \Gamma'.\text{Vt}(\mathcal{C}, rp_j^*)$, set $\text{ACC} = \delta(\text{ACC}_{r-1}, TX)$.
3. (*Mining*). Extend chain and Reparo data structures as follows,
 - (a) Perform standard Ethereum mining and set new block $B_r \leftarrow \langle \text{header}, TX \rangle$
 - (b) Extend local chain $\mathcal{C} \leftarrow \mathcal{C} || B_r$, the repair layer $\text{Rdb} \leftarrow \text{Rdb} || \text{Rdb}_r$ and the approved repairs $\text{Adb} \leftarrow \text{Adb} || \text{Adb}_r$
 - (c) Then broadcast $(\mathcal{C}, \text{Rdb}, \text{Adb})$ to the network

Updating the chain. When a node receives \mathcal{C}, Rdb , and Adb , check if the chain is valid by calling $\Gamma'.\text{validateChain}(\mathcal{C}, \text{Rdb}, \text{Adb}) = 1$. Accept the new chain if the new chain is valid as per Ethereum's fork resolution rule¹¹.

Figure 6: Reparo protocol integration into Ethereum with PoW based consensus and parameterized by policy \mathbb{P}

- The proposal does not redact or modify votes in the chain.
- The proposal has received more than ρ fraction of votes (50% of votes) in ℓ consecutive blocks (voting period, that can be decided by the system) after the corresponding repairTx is included in the chain.

Performing Repair Operations. Upon approval with respect to the policy \mathbb{P} repair operations are performed as they were performed in the PoS variant. Additionally, in the PoW

system, there are no restrictions on the redaction policy except for the ones discussed above. Unlike the PoS variant, we can redact transactions as a whole since the consensus is independent of the state of accounts.

Block Validation. Block validation is the same as in the PoS variant, except that now instead of checking for PoS consensus we check if the block has the correct nonce for PoW. A formal description of the procedure can be found in Algorithm 5.

Security and public verifiability properties are the same way as in the PoS variant. We describe the security argument here for completeness.

Security. Since the hash function H is modeled as a random oracle (RO), finding a collision on a vote (which is the hash of the ID of the old transaction and the ID of the candidate transaction) is highly improbable. Therefore, when a miner votes for a repair proposal in his newly mined block, no adversary can claim a different repair proposal for the same vote value. Same property of the hash function H also ensures that no adversary can find a different block that hashes to the same hash of an honestly mined block. Therefore an adversary cannot break the integrity of the chain. Together, they imply the *unforgeability* of votes, as if an adversary wishes to vote, he has to mine a block with his vote himself. Assuming majority of the miners are honest in Ethereum, **Reparo** integration with Ethereum satisfies *editable common prefix* and preserves chain quality and chain growth with respect to repair policy \mathbb{P} .

Optimizations. *State Assertion:* Instead of recomputing the state, a `repairTx` can propose a state for the affected contract and accounts. The protocol can then inject this state if allowed by the policy. This method is inexpensive as it is without any cascading computation and is useful for users who accidentally locked their funds [9].

D.2 Repairs in Ethereum

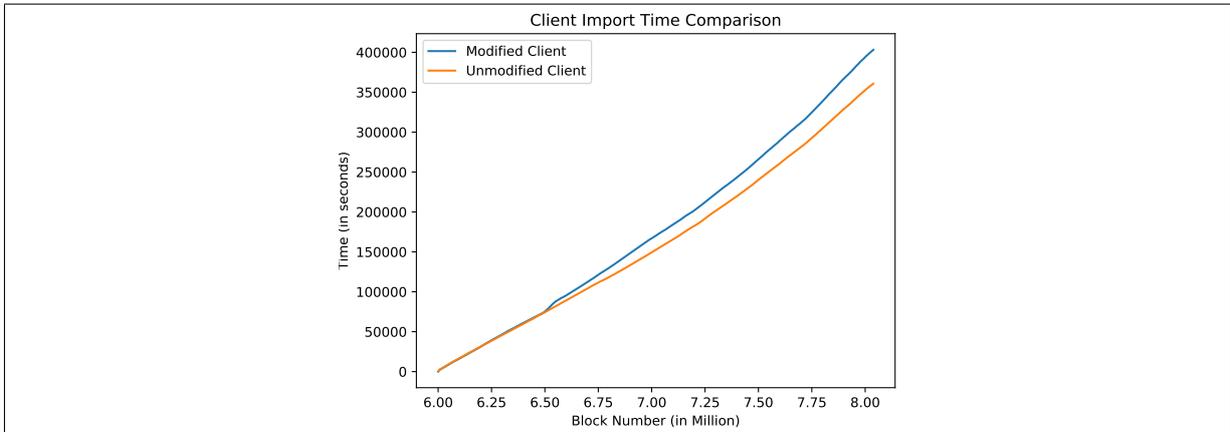


Figure 7: Client Comparison: The cumulative time taken to import 2,038,219 blocks.

Estimating Repair Time. Figure 7 plots the cumulative time taken to process the blocks plotted against the block number. For example, taking a point at block number 6,500,753, the point on the orange line reports the time 75,548.49s. It would take 75,548.49s to recompute the state after changing a transaction at a block 500,753 deep and rebuilding the states. This can be used in conjunction with the data in Figure 4 to determine the time it would take the network to build new states and also measure the impact of a repair on the import for archive or full sync nodes.

D.3 Prominent Bugs

In Table 7 we provide the information of some of the popular bugs.

DAO. This is a re-entrancy bug in the contract that allowed a maliciously crafted call to drain the balance of the contract before it subtracted the balance from the user. We propose to fix this contract by updating all DAO contract creation contracts with the bug fixed code. This is different from the ad-hoc solution employed by Ethereum today. Ethereum hard-coded the address for DAO and executes the contract differently. This ensured that the blockchain should have no transaction dependency because the blockchain already has the state with the contract fixed. This in conjunction with the repair proposal allows an inexpensive repair for DAO even though it has a lot of dependent transactions.

Parity Multisig Wallet Bug. The Parity Mutli Sig Wallet is a library contract that had a bug which had a public constructor that allowed any user to take control of the contract. A user took ownership of the contract and accidentally killed it. We propose to repair this contract by undoing the transaction that killed the contract. The transaction dependency is unaffected as it just resurrects a dead contract. This enables all Parity Multisig Wallet holders to safely recover their funds.

QuadrigaCX (QCX) and REXmls. These contracts have hardcoded wrong addresses in the contract which sent the ICO ETH to an incorrect address (an account that does not exist) thereby permanently locking the coins in those contracts. We propose to repair this bug by proposing a repair transaction with the same code but with the correct address, which can be used to recover and return the lost funds.

No code contract. There are 2,986 such contract creation transactions which have money but no code in the creation call. The idea to solve the no code contract problem, is to allow the user to add code to the contract. We give a template of the code in Figure 8. The contract allows the user who locked the money in a contract to retrieve the money.

E Other Instantiations

In this section we discuss how our protocol can be instantiated into other systems like Bitcoin and Cardano. Note that Cardano is a Proof of Stake [31, 16] based system. Even though Bitcoin and Ethereum are PoW based systems, we discuss Bitcoin instantiation because redaction of illicit data entries is a major problem in Bitcoin and we want to highlight our improvements compared to the work of Deuber et al. [25].

E.1 Integrating into Bitcoin

Deuber et al. [25] instantiated their redactable blockchain protocol with Bitcoin and showed how to redact harmful illicit content from data pockets in transactions. We now describe how to instantiate our **Reparo** protocol from Figure 2 on top of Bitcoin for removal of arbitrary non-payment data bytes (*stateless* repairs) and any other repair operations if such a need arises. We primarily focus on the former repair operations in case Bitcoin as removal of illicit data entries is immediate critical problem to solve.

The main differences between [25] and ours when **Reparo** is integrated into Bitcoin are:

¹²DAO with transaction hash: 0xe9ebfecc2fa10100db51a - 4408d18193b3ac504584b51a4e55bdef1318f0a30f9

¹³Quadriga CX with transaction hash: 0xf4c8423215e8abb2810ff - 0b2eb82a93ce660a0ce651df5f14005e08f6e25318e

¹⁴Parity with transaction hash: 0x348ec4b5a396c95b4a5524ab0 - ff61b5f6e434098cf6e5c1a6887bed2bc35625d

¹⁵REXmls (imbrex) token with transaction hash: 0xcb6b1e83452608 - 65b5e164e09584634960ec1d02d80d3bbb4e1533c77393d216

```

pragma solidity ^0.5.0;

contract simpleWithdraw {
    address private owner;
    uint256 money;
    constructor() public payable {
        owner = msg.sender;
        money = msg.value;
    }

    function withdraw() public {
        if (msg.sender == owner) {
            selfdestruct(msg.sender);
        }
    }
}

```

Figure 8: Contract to retrieve lost money

1. Our protocol does not require the modification of the Bitcoin block header.
2. As a consequence of the above point, we do not require an additional hash value to be stored in every Bitcoin block header like their protocol. This makes our protocol much more space efficient than theirs.
3. Our protocol is immediately integrable into Bitcoin in a backward compatible fashion. This means that already existing illicit data entries in Bitcoin can be redacted once **Reparo** is fit on top of Bitcoin today.
4. Unlike [25], we handle both stateless redactions and stateful repair operations.

We refer to [25] for basic understanding of how Bitcoin transactions work and how arbitrary data can be inserted into transactions.

Regarding user roles, as described in Section 4, *miners* also happen to be *deciders* in the instance of Bitcoin. *Miners* vote for proposal s as they happen to mine blocks similar to the protocol in [25].

Data Structures. Note that unlike [25] we do not require any modification of the Bitcoin block structure. Instead we have a *repair layer* Rdb for each block that is empty when the block is mined.

Proposing Repairs. Similar to [25] we have a special transaction `repairTx` that contains the hash of the old and new version of the transaction in its output script. In other words the transaction id's Tx_{ID} and Tx^*_{ID} are stored. In this case Tx^* is the candidate transaction. The user then broadcasts `repairTx` and Tx^* to the network; `repairTx` requires a transaction fee to be included in the blockchain, while Tx^* is added to a pool of candidate transactions. The candidate transaction Tx^* is validated by checking its contents with respect to Tx , and if it is valid, then it can be considered for voting. For stateless repairs like redacting arbitrary non-payment data entries only, it is checked if the only difference between Tx and Tx^* is that of the missing data entry in Tx^* .

Repair Policy. Our protocol is parameterized by a repair policy parameter \mathbb{P} similar to [25]. For the case of redacting non-payment data entries we follow the same basic policy recom-

mendations as theirs. For completeness, we detail the policy requirements here. A proposed redaction is approved valid if the following conditions hold:

- It is identical to the transaction being replaced, except that it can remove data.
- It can only remove data that can never be spent, e.g., `OP_RETURN` output scripts.
- It does not redact votes for other redactions in the chain.
- It received more than 50% of votes in the 1024 consecutive blocks (voting period) after the corresponding `repairTx` is stable in the chain.

Similar to the discussion on repair policy for Ethereum Section 4, we argue that the policy for handling such stateful repair proposals can be quite event specific and strongly dependent on auxiliary information from the real world. The auxiliary information from the real world helps to see if a particular stateful repair proposal is good for the chain or not. As a minimum requirement from the policy (to help miners detect malicious proposals), a proposed stateful repair operation is approved valid if the following conditions hold:

- It is unambiguously not a double spend attack attempt (needs auxiliary information from the real world for confirmation).
- It cannot propose to repair a transaction thereby making it an invalid spend.
- It does not change the amounts being transacted.
- It does not redact or modify votes in the chain.
- It received more than 50% of votes in the 1024 consecutive blocks (voting period) after the corresponding `repairTx` is stable in the chain.

Deliberation for a proposal is done via voting by miners. Voting for a candidate transaction Tx^* simply means that the miner includes $\text{repairTx}_{\text{ID}} = H(\text{Tx}_{\text{ID}} || \text{Tx}_{\text{ID}}^*)$ in the coinbase (transaction) of the new block he produces. After the voting phase is over as determined by the policy \mathbb{P} , the candidate transaction is removed from the candidate pool.

Performing Repair Operations. Our protocol slightly varies from [25] in this regard. In case of stateless repair operations like redactions, once a candidate transaction has been approved by the redaction policy, the miners in the network replace the old version of the transaction (being repaired) and replaces it with the candidate transaction, while storing the hash of the old version in the corresponding repair layer `Rdb` of the block. In case of stateful repair operations, the entire old version of the transaction is stored in the corresponding repair layer of the block.

Updating UTXO with Stateful Repair Operations. Since stateful repair operations could involve changing payment information which could affect the UTXO database after the mining of a block, special care needs to be taken in updating the UTXO. When a repair operation proposal is approved, the miners perform the repair operation as discussed above. The UTXOs after each of the following blocks upto the most recent block is accordingly updated. This could mean that some transactions in these blocks become invalid spends as their input is no longer in the UTXO database at the time that block was mined (after performing the repair operation). This is the cascading state update that was discussed previously in the case of Ethereum Appendix D.1. After performing the repair operation and reflecting it in the UTXO database at each subsequent block, the miners try to mine a new block by including transactions that are consistent with the updated UTXO database at that time.

Chain Validation. To validate a full chain a miner needs to validate all the blocks within the chain. We discuss the general case where both stateless (redactions) and stateful repair operations could have been performed on the chain. Consequently our chain validation procedure is different from [25] as ours is more generic. For validation, the miner uses the repair layer of the blocks in the chain to *go back in time* to the mined version of the blocks. This ensures that the miner is now having the the original state of the chain. Note that this holds true even in case of stateless redactions as the repair layer **Rdb** would contain the hash of old version of the transaction. The miner now validates the chain from the genesis by “re-mining” the chain, with the catch that instead of solving for PoW, he verifies the existing PoW. Since the miner is validating the blocks as if when they were freshly mined, a valid PoW (in the past) remains a valid PoW now for the miner. If ever some stateful or stateless repair proposal was approved, the miner performs the repair operation that is required to be performed (as it would have been performed). This way the miner validates and re-constructs the chain. The miner rejects a chain as invalid if *any* of the following holds: (1) a block’s repair operation was not approved according to the policy, or (2) a previously approved repair was not performed on the chain.

Validating Transactions. Validating a chain involves validating a block and its contents as a subroutine. The miner validates all the transactions contained in its transactions list against the current database of UTXOs; the validation of unedited transactions is performed in the same way as in the immutable version of the Bitcoin protocol. The miner simply validates transactions in the block against their witnesses. In case of having only the hash old version of the transaction and the old witness (this is the case of removal of non-payment data entries - stateless redactions), the miner can validate the witness with respect to the new version of the transaction as the payment scripts are unchanged in a stateless redaction. This is similar to the validation in [25]. Therefore, we can ensure that all the transactions included in the block have a valid witness, or in case of redacted transactions, the old version of the transaction had a valid witness.

E.2 Integrating into Cardano (PoS)

Cardano is a cryptocurrency that runs the Ouroboros PoS consensus mechanism. Our interest in this system is to show how one can instantiate our **Reparo** protocol Figure 2 on top of a PoS based blockchain.

Similarities between Bitcoin and Cardano. Transactions in Cardano work the same way as in Bitcoin and Cardano is based on a UTXO model. The address field in Cardano has additional semantics for the staking procedure of the PoS consensus process. Block headers in Cardano are more or less the same as in Bitcoin except for consensus proof which is different from the PoW value in Bitcoin.

Consensus. The time is divided into 120 second slots and each slot has a slot leader elected to propose the new block. The slot leaders are elected with a winning probability proportional to their stake in the system. In an epoch which lasts for 20 hours, the slot leaders for each slot of the next epoch are determined but not revealed. As a proof of election, the elected slot leader generates a signature proving his stake in the system, which can be verified by everyone else.

Policy for PoS. The repair policy requirements is more or less the same as discussed previously for systems like Ethereum and Bitcoin. However, now we do not allow *redactions* that affect the state of the system. In case of a *stateful* redaction, as the state of the chain has changed, the data point that causes this state change is erased/redacted. In PoS based blockchain systems, such a stateful redaction causes failure in chain validation as an honest new user can no longer verify the stakeholder consensus proof. This is because the stake distribution has changed, but the transaction resulting in the older stake distribution is no longer stored and therefore

consensus proofs based on the older stake distribution can no longer be verified.

Repairing and Chain Validation. The only difference in terms of proposing and finally performing a repair operation is that the voting period for a repair proposal begins from start of the immediately next epoch and spans throughout that epoch. This is to ensure that chain validation procedure is able to validate consensus after repair operations have been performed. In more detail, recall that the chain validation as described for Bitcoin, we *go back in time* and validate the blocks in the state in which they were mined and perform repair operations just the way they were performed. When following this procedure, one must be able to validate the consensus proof, PoW in Bitcoin and PoS in Cardano. As mentioned earlier, the slot leaders for the current epoch are determined by the end of the previous epoch. Consider a case where there is a repair proposal whose voting period starts in the middle of an epoch and ends in the middle of the next epoch and gets approved. The miner performs the repair operation that could potentially change the state and thereby the stakes. This makes the verification of the elected slot leaders for the rest half of the slots in the epoch inefficient and time consuming, as these slot leaders were determined by the state before the repair operation was performed. In order to avoid this inefficiency, we let the voting period to be synchronized with the epoch period of Cardano. Chain validation can now proceed as in Bitcoin, except that the consensus is PoS [31].

Algorithm 5: validateBlock

input : Chain $\mathcal{C} = (B_1, \dots, B_n)$, repair layer $\text{Rdb} = (\text{Rdb}_1, \dots, \text{Rdb}_n)$, block B_{n+1} , repair approved Adb_{n+1} .
output: $\{\perp, (\mathcal{C}', \text{Rdb}')\}$

- 1: Parse $B_{n+1} := \langle \text{header}_{n+1}, TX_{n+1} \rangle (ACC_{n+1})$, where $\text{header}_{n+1} = (pt_{n+1}, G(TX_{n+1}), hd_{n+1})$;
- 2: Parse $B_n := \langle \text{header}_n, TX_n \rangle (ACC_n)$, where $\text{header}_n = (pt_n, G(TX_n), hd_n)$;
- 3: Validate transactions x_{n+1} , if *invalid* **return** \perp ;
- 4: **if** $pt_{n+1} \neq H(\text{header}_n)$ **then return** \perp ;
- 5: **if** $\text{Adb}_{n+1} = \emptyset \wedge ACC_{n+1} = \delta(y_n, x_{n+1}) \wedge \text{chk_pow}(\text{header}_{n+1})$ **then** Set $\mathcal{C}' \leftarrow \mathcal{C} \parallel B_{n+1}$, and $\text{Rdb}' \leftarrow \text{Rdb} \parallel \emptyset$, and **return** $(\mathcal{C}', \text{Rdb}')$;
 \triangleright Validate Blocks where repairs were approved
- 6: Initialize $\mathcal{C}' \leftarrow \mathcal{C}, \text{Rdb}' \leftarrow \text{Rdb}$;
- 7: **for all** $TX_j^* \in \text{Adb}_{n+1}$ **do**
 - \triangleright Perform all the repair operations
 - 8: Parse $B_j := \langle \text{header}_j, TX_j \rangle (ACC_j)$, where $\text{header}_j = (pt_j, G(TX_j), hd_j)$;
 - 9: **if** $\text{chkApproval}(\mathbb{P}, H(G_{tx}(TX_j), G_{tx}(TX_j^*))) \neq \text{approve}$ **then return** \perp ;
 - 10: Set $ACC_j^* := \delta(ACC_{j-1}, TX_j^*)$;
 - \triangleright Perform the repairs as originally performed
 - 11: $\mathcal{C}', \text{Rdb}' \leftarrow \text{repairChain}(\mathcal{C}', \text{Rdb}', rp_j^*)$;
- 12: Parse $\mathcal{C}' := (B'_1, \dots, B'_n)$ and $B'_n := \langle \text{header}'_n, TX'_n \rangle$;
 \triangleright Check the state transition after repair
- 13: **if** $ACC_{n+1} = \delta(ACC'_n, TX_{n+1}) \wedge \text{chk_pow}(\text{header}_{n+1})$ **then** Set $\mathcal{C}' \leftarrow \mathcal{C} \parallel B_{n+1}$, and $\text{Rdb}' \leftarrow \text{Rdb} \parallel \emptyset$ and **return** $(\mathcal{C}', \text{Rdb}')$;
- 14: **return** \perp ;

Table 7: Prominent Smart Contract Bugs

Bug	Start Block	ETH Affected
DAO ¹²	1,428,757	3,600,000
QuadrigaCX ¹³	1,952,428	67,317
Parity Multisig ¹⁴	4,049,249	513,736
REXmils ¹⁵	4,066,859	6,687
No Code Contracts	—	6533.17