

ERC-20R and ERC-721R: Reversible Transactions on Ethereum

Kaili Wang
kkwang@cs.stanford.edu

Qinchen Wang
qinchenw@cs.stanford.edu

Dan Boneh
dabo@cs.stanford.edu

September 9, 2022

Abstract

Blockchains are meant to be persistent: posted transactions are immutable and cannot be changed. As a result, when a theft takes place, there are limited options for reversing the disputed transaction, and this has led to significant losses in the blockchain ecosystem.

In this paper we propose reversible versions of ERC-20 and ERC-721, the most widely used token standards. With these new standards, a transaction is eligible for reversal for a short period of time after it has been posted on chain. After the dispute period has elapsed, the transaction can no longer be reversed. Within the short dispute period, a sender can request to reverse a transaction by convincing a decentralized set of judges to first freeze the disputed assets, and then later convincing them to reverse the transaction.

Supporting reversibility in the context of ERC-20 and ERC-721 raises many interesting technical challenges. This paper explores these challenges and proposes a design for our ERC-20R and ERC-721R standards, the reversible versions of ERC-20 and ERC-721. We also provide a [prototype implementation](#). Our goal is to initiate a deeper conversation about reversibility in the hope of reducing some of the losses in the blockchain ecosystem.

1 Introduction

Since their inception, cryptocurrencies have been plagued by thefts and accidental losses [24]. Victims include end users [7], DAOs [18], bridges [16, 15, 23, 9], and exchanges [10, 1, 8, 20, 21]. Usually, the stolen assets are first transferred from the victim’s address to an address controlled by the attacker. From there the assets are laundered by transferring them to other addresses and eventually to an offramp. In a few cases, the assets are seized at the offramp [4].

The annual losses can be quite high. In 2020, \$7.8 billion was stolen, and in 2021 that amount doubled to \$14 billion [5, 19]. A recent well publicized attack on the Ronin bridge resulted in a theft of over \$600 million [16]. The attacker transferred ETH and USDC from the Ronin contract on Ethereum to an address controlled by the attacker. From there, some of the funds were moved to the Tornado mixer [17]. Other bridges have experienced similar thefts [15, 23, 9]. In many of these attacks, the assets stolen were held in ERC-20 contracts.

Similarly, NFTs held in ERC-721 and ERC-1155 contracts have seen an uptick in thefts. In a twelve months period following July 2021, an estimated \$100 million were stolen in NFTs using phishing and other attacks [12, 5].

These attacks were often discovered soon after the theft took place. Had there been a way to reverse the offending transaction(s) – as in traditional finance –

the damage could have been greatly reduced.

Beyond theft, transaction finality has also worked against us when funds are accidentally sent to a wrong address. In May 2022, a Cosmos-based blockchain called JUNO passed a proposal to move \$36 million USD to a specific address. The address contained a typo [14] and consequently the funds were lost. The funds could have been recovered had there been a way to reverse that transaction.

Reversible transactions. In a 2018 tweet, Vitalik Buterin wrote that

Someone should come along and issue an ERC20 called “Reversible Ether” that is 1:1 backed by ether but has a DAO that can revert transfers within N days.

Nowadays, this can be applied to any ERC-20 token (not just wrapped ETH), as well as to NFTs.

Enabling reversible transactions is not easy and introduces many fascinating technical challenges. The main contribution of this paper is to explore how to support reversibility within the ERC-20 and ERC-721 framework. We propose two new standards that allow transaction reversal within a limited time window, say four days. We call these standards ERC-20R and ERC-721R, the reversible versions of ERC-20 and ERC-721, respectively.

We envision the following high-level workflow for reversing a posted transaction (see Figure 1):

- *Request freeze.* The victim posts a freeze request to a governance contract, along with the relevant evidence, and some stake.
- *Freeze assets.* A decentralized set of judges decides to accept or reject the request. If accepted, the judges instruct an on-chain governance contract to call the *freeze* function on the impacted ERC-20R or ERC-721R contract. Subsequently, the assets in question are frozen and can no longer

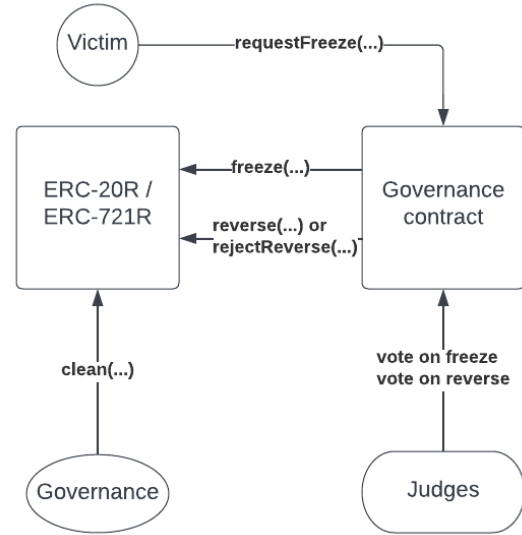


Figure 1: The process for reversing a transaction.

be transferred. For NFTs, this is a simple matter of freezing the disputed NFT. For ERC-20 tokens this is more complicated, as we explain below. We discuss the operation of the governance contract, and the selection of judges, in Section 3. We envision the freeze process being relatively quick, taking the judges at most one or two days to make a decision.

- *Trial.* Both sides can then present evidence to the decentralized set of judges. Eventually the judges reach a decision, at which point they instruct the governance contract to call either the *reverse* or *rejectReverse* functions on the impacted ERC-20R or ERC-721R contract. The *reverse* function transfers the disputed (frozen) assets to their original owner. The *rejectReverse* releases the freeze on the disputed assets and leaves them where they are. The trial may be lengthy, possibly taking several weeks or months.

This simplistic design can create opportunities for an

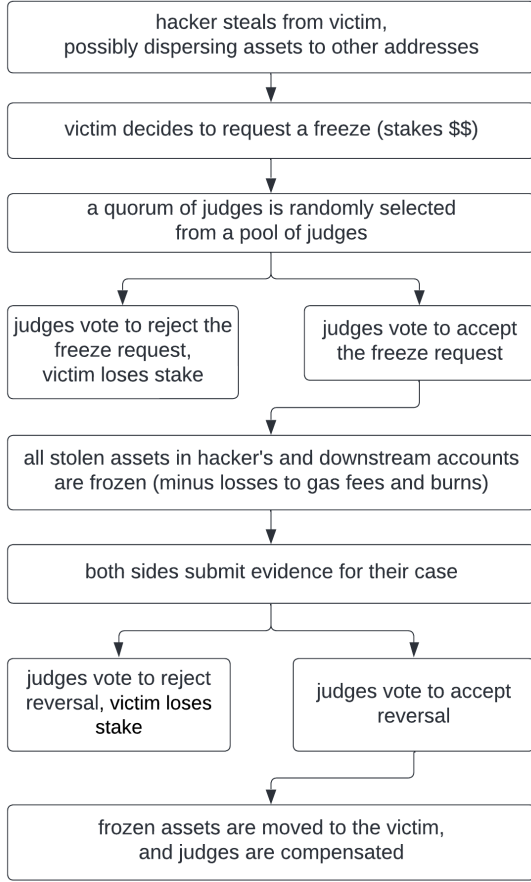


Figure 2: Overview of reversal process

attacker to freeze the assets of an honest user, and possibly even to reverse honest transactions. A more complete workflow is presented in Figure 2, and we discuss this further in Section 3.

Locating the stolen assets. By the time the victim submits a freeze request, the attacker may have already moved the stolen assets through multiple accounts. In fact, the attacker can monitor the mempool, and move the assets as soon as it sees a request to freeze the stolen assets. In the case of an NFT, the

attacker may have sold the stolen NFT to an unsuspecting honest user. In the case of an ERC-20 token, the attacker may have divided the stolen tokens across multiple accounts; it may have exchanged a portion of the tokens for another ERC-20 token using an honest on-chain exchange; it may have burned a portion of the tokens; or it may have sent the tokens to a mixer. The new reversible standards must properly handle all these cases.

In case of a dispute over an ERC-721 NFT, the freeze is applied to the current holder of the NFT: either the original attacker, or an honest user who purchased the stolen NFT from the attacker. If the judges decide that a theft took place, then the ERC-721R contract sends the NFT back to the pre-theft owner. The current owner of the NFT loses the NFT. This policy is consistent with tort law in many countries, but of course, other policies can be implemented.

In case of a dispute over stolen ERC-20 tokens things are more complicated. By the time the freeze is executed, the funds may have been dispersed across many downstream accounts, some honest and some dishonest. In Section 2 we present an example algorithm that assigns fractional responsibility to each of the downstream accounts that received a portion of the stolen funds. The partial freeze is then applied to these accounts. Implementing this freeze strategy requires the ERC-20 contract to maintain a transaction log during the dispute window so that the *freeze* function can trace the funds when it is called by the governance contract. If the judges decide that a theft took place, the ERC-20R contract moves the frozen tokens from the obligated accounts to the pre-theft account. We discuss this in more detail in Section 2.

Implementation. We provide a reference implementation of these new standards. Our Solidity implementation is split into two parts: (i) the main ERC-20R and ERC-721R contracts that keep track of all the balances and transactions, and (ii) a governance contract

that selects judges and gathers votes. Our ERC-20R and ERC-721R implementations are extensions of the OpenZeppelin non-reversible contracts.

The new API functions in the ERC-20R and ERC-721R contracts are *freeze*, *reverse*, *rejectReverse*, and *clean*. Let us describe this API in more detail:

- *freeze()*: calculates the amounts to freeze on the attacker’s address as well as potential downstream addresses, and freezes those amounts. For ERC-20R it returns a *claimID* that points to an on-chain list of (account,amount) pairs. The list identifies all the accounts that contain frozen assets associated with the complaint, and the amount frozen in each. For ERC-721R it returns a boolean success flag. The inner workings of the *freeze* function is explained in Section 2.
- *reverse()*: sends all frozen assets associated with the claimed theft back to the original owner. For ERC-20R, takes as argument a valid *claimID*. For ERC-721R, takes in arguments (*tokenId*, *index*) where *index* identifies the transaction being reversed.
- *rejectReverse()*: unfreezes all amounts associated with the claim. For ERC-20R, takes as argument a *claimID*; for ERC-721R, takes as argument a *tokenId*.
- *clean()*: These contracts store some transaction data on chain. The *clean* function removes on-chain information for transactions whose dispute window has elapsed. For ERC-20R, takes in an array of addresses and an epoch; for ERC-721R, takes in a list of token IDs. The data structure that is being updated is explained in Section 2.1.

The *freeze*, *reverse* and *rejectReverse* functions can only be called by the governance contract. Anyone can call the *clean* function.

1.1 Related Work

In the fall of 2018, one of the co-creators of the ERC-20 standard proposed the concept of a *reversible ICO*, where investors would be able to get a refund amount inversely proportional to how recently they invested [22]. Although this safeguards against a single token being a scam at launch, it does not protect against malicious transactions.

Eigenmann [11] drafted a contract for a reversible token that extends the ERC-20 standard. It used an escrow method, where the escrow period was 30 days, during which the sender could recall the money at any time. This is problematic because Bob could pay Alice for a service, and then reverse the payment 28 days later, after Alice completed the service. A similar approach is used in a proposal for refunds in ERC-721 mints [13], where Bob can get his money back within a certain time window after buying an NFT. In our proposal, Bob would need to present sufficient evidence to a committee of judges for a transaction to be reversed. This protects counter-parties to the disputed transaction.

The *Reversecoin* project from 2015 launched as a layer 1 blockchain [6]. It introduced a timeout period between transaction initiation and confirmation. Each account has an offline key pair that enables the owner to either reverse a transaction or immediately confirm it. This may not prevent many of the modern hacks: the attacker would either steal the confirmation key, or trick the user into using the confirmation key to confirm a malicious transaction.

Finally, centralized exchanges maintain the ability to freeze and remove assets. For example, *Binance USD* (BUSD), issued by Paxos, states that (link):

Paxos also has the ability to create and burn BUSD tokens at will, as well as freeze and remove funds from people who exhibit nefarious or illicit activity.

The same holds for *USD Coin* issued by Circle. In these centralized systems, the operator acts as a centralized judge that can reverse transactions.

2 The transaction reversal process

In this section we describe the details of the ERC-20R and ERC-721R process of reverting a transaction. We describe the data structures and algorithms needed to track the funds that will be frozen and later sent back to the victim, if the victim prevails.

2.1 Freezing assets

After a theft from an ERC-20R or an ERC-721R contract, the victim posts an on-chain freeze request to the governance contract. The request includes the offending transaction ID, a link to evidence that an unauthorized transfer took place, and some stake. If the judges are convinced by the evidence (see Section 3) then they instruct the governance contract to call the *freeze* function on the relevant ERC-20R or ERC-721R contract. Once the assets are frozen, any attempt to transfer them will fail. If the judges are unconvinced, they instruct the governance contract to reject the request, and the victim loses the stake.

In this section we explain what happens when the governance contract calls the *freeze* function on the relevant ERC-20R or ERC-721R contract.

2.1.1 An ERC-721R freeze

The *freeze* function on an ERC-721R contract is quite simple. First, we add two structures to the contract:

```
mapping(uint256 => bool) _frozen;
mapping(uint256 => Queue) _owners;
```

The `_frozen` structure indicates if a particular `tokenId` is frozen. If `_frozen[tokenID]` is true then the asset is frozen, and cannot be transferred.

The `_owners` structure keeps track of the recent list of owners for the asset identified by `tokenId`. On every transfer, a pair (*newOwner*, *blockNumber*) is appended to the queue at `_owners[tokenID]` to record that at this block number, the new owner of `tokenId` became *newOwner*. This `_owners` structure looks as:

$$\begin{cases} tokenID_0 \rightarrow [(owner_0, bn_0), (owner_1, bn_1), \dots] \\ tokenID_1 \rightarrow [(owner_0, bn_0), (owner_1, bn_1), \dots] \\ tokenID_2 \rightarrow [(owner_0, bn_0), (owner_1, bn_1), \dots] \\ \dots \end{cases}$$

where for each *tokenId* we have $bn_0 \leq bn_1 \leq \dots$

The governance contract freezes an asset by calling `freeze(tokenID, index)`

where `index` points to an entry in the queue `_owners[tokenID]`. The function first verifies that the disputed transfer took place within the dispute window by using the block number at position `index+1` in the queue `_owners[tokenID]`. The function also ensures that the asset is not already frozen. If the checks succeed, the function sets `_frozen[tokenID]` to true, indicating that the asset is frozen. That's it. Doing these checks on-chain ensures that even malicious judges cannot reverse a transaction outside of the dispute window.

If at a later time the ERC-721R contract is asked to revert the disputed transfer (that is, the governance contract calls the *reverse* function), the contract simply transfers `tokenId` to the address written at position `index` in the queue `_owners[tokenID]`. This was the owner before the disputed transaction.

Calling the *clean* function on this contract with a list of token IDs removes data from the `_owners` structure that is no longer needed. This removes transfers that can no longer be reverted because the dispute window has elapsed. It is done to save on chain storage. Note that a frozen `tokenId` cannot be cleaned

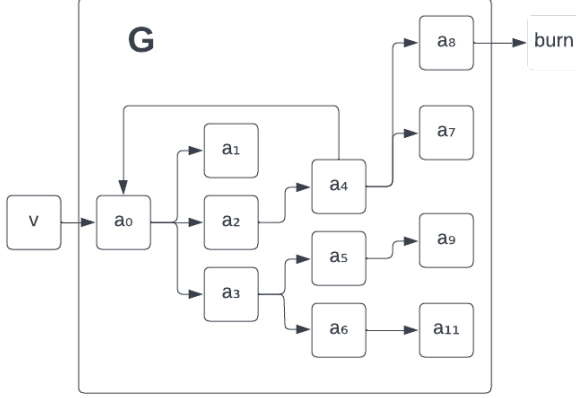


Figure 3: An example graph G of transfers following the disputed transaction from v to a_0 .

until it becomes unfrozen; this preserves the data needed to reverse a disputed transaction, if needed.

2.1.2 An ERC-20R freeze

The freeze function on an ERC-20 contract is much more complicated. The problem is that the tokens might have been transferred to multiple accounts between the time of the theft and the freeze request.

Figure 3 shows a transaction from a victim address v to an attacker address a_0 . Subsequent transfers from a_0 that took place after the disputed transaction, but prior to a freeze request, are indicated as directed edges in the graph. All these downstream addresses may hold stolen funds that may need to be frozen.

Another complication is that tokens can be burned and simply vanish from the system. For example, in a stablecoin contract that is collateralized by a fiat currency, the account a_8 may choose to redeem its coins for fiat currency, at which point the coins at a_8 are burned. Consequently, there might not be sufficient assets to freeze at address a_8 , and the stolen funds are permanently lost. We discuss this issue further in Section 4.

Yet another complication is that a single address, say address a_1 , might own funds obtained from multiple thefts. When the funds are frozen, the system needs to remember the amount frozen for each reversion request, and if a reversion request is approved, only the funds associated with that request should be taken out of account a_1 .

If there are multiple freeze events on a single account, the frozen amount is cumulative. In particular, if account a has a total of s frozen coins, then the ERC-20R contract will reject any transfer that causes the balance of account a to drop below s .

Calculating suspect addresses. Let us begin by describing a specific strategy for calculating the amount to freeze on each account in the graph G from Figure 3. Accounts that are subject to a freeze are called *suspect* addresses. Our freeze algorithm will chase the stolen funds across the transaction graph and freeze funds across suspected addresses. The algorithm will freeze assets that are as close as possible to a_0 in G .

Before we describe the detailed algorithm, let us first see a few examples. Suppose that s coins are stolen from the victim v and transferred to account a_0 at time T_0 . At time $T_f > T_0$ the ERC-20R contract receives a request to freeze that transaction.

Example 1. If at time T_f the balance at a_0 is s or more, then s coins will be frozen at a_0 and the process concludes.

Example 2. Suppose that at some time between T_0 and T_f a transaction transfers $s/4$ coins from a_0 to a_1 and an additional $s/4$ coins from a_0 to a_2 . The remaining balance at a_0 is $s/2$ and no subsequent transactions apply to a_0 . Then at the freeze time, the entire $s/2$ balance at a_0 will be frozen. Moreover, if the balance at a_1 and a_2 is exactly $(s/4)$, then $(s/4)$ coins will be frozen at each of those accounts. Now that a total of s coins has been frozen, the process terminates. Note that a_1 or a_2 might be the addresses of an honest exchange or a mixing service, in which case

a portion of the coin's liquidity pool at the exchange or mix will become frozen.

Example 3. Suppose that following the disputed transaction from v to a_0 , there is a second transaction $a_0 \rightarrow a_1$. Clearly if there are insufficient funds at a_0 at the time of the freeze, then some of the freeze obligations should pass on to a_1 , as explained in the previous paragraph. However, suppose that at some time *before* the $a_0 \rightarrow a_1$ transaction, there is a transaction $a_1 \rightarrow a_2$ that transfers funds from a_1 to a_2 . We propose that none of the freeze obligations will pass on to a_2 , because the $a_1 \rightarrow a_2$ transaction was posted before the disputed funds arrived at a_1 . The funds sent to a_2 are not directly involved in the dispute. One can give examples where this policy can lead us astray, but by in large, we claim that a_2 should not be involved in the freeze.

These examples suggest that the freeze process is an iterative procedure that freezes the maximum amount possible at every step. If the balance at the current node is insufficient, then the obligation is passed to the descendants of that node. The process terminates once s coins are frozen, or once there are no more descendants to process.

We stress that the freezing algorithm in its entirety runs in a *single* transaction. This ensures that account balances cannot change while the freeze process is executing.

Terminology. Suppose that the freeze function is called to freeze a transaction t_0 that transferred s coins from address v to address a_0 . Let t_f be the posted freeze transaction on chain. To describe the freeze algorithm, we use the following notation:

- $toFreeze(a)$ is the number of coins that the freeze transaction t_f will freeze at address a . At the start of the algorithm $toFreeze(a) = 0$ for all a , with the exception of a_0 , where

$$toFreeze(a_0) := \min(s, Bal(a_0)). \quad (1)$$

The quantity $Bal(a)$, for an address a , is the available balance at a at the time of t_f . This $Bal(a)$ is calculated as the balance at a at the beginning of the freeze transaction minus the amount of coins already frozen at a due to a prior dispute. Thus (1) will freeze the maximum amount possible at a_0 .

- For a transaction $t = (a \rightarrow b)$, from a to b , let $val(t)$ be the value transferred from a to b .
- $burnedAt(a)$ is set to the number of coins burned at address a between the transaction where a portion of the disputed funds were first sent to a and the freeze transaction t_f .

Now, consider the graph G that is defined by the set of transactions that took place after the disputed transaction and before the freeze transaction. The nodes in the graph are addresses, and every directed edge represents a transfer from one address to another. The graph only includes an edge $b \rightarrow c$ if there is directed path of transactions from a_0 to address b that all took place after t_0 .

The algorithm. We first describe a freeze algorithm that applies when the graph G rooted at a_0 is a directed *acyclic* graph (DAG). In Appendix A we extend the algorithm to handle cycles by introducing a pre-processing phase that eliminates cycles.

The freeze algorithm is implemented in the function $CalcFreeze$ shown in Figure 4. This function is called as

$$CalcFreeze(t_0)$$

where t_0 is the disputed transaction.

The algorithm begins by constructing a **topological sort** of the vertices of G starting at a_0 . Recall that a topological sort is a list L of the vertices in G , where every vertex in G appears exactly once in L , such that for every edge $(a \rightarrow b)$, the vertex a appears in L before b . Every DAG has a topological sort L , and the

list L can be constructed in linear time in the number of edges.

For each address a in G the algorithm builds a value $oblig(a)$ that indicates the obligation amount that is passed to address a and its descendants as a result of the disputed transaction. The value of $oblig(a)$ can increase whenever the algorithm process an address that sent funds to a . For all a , the array $oblig(a)$ is initially set to zero.

The algorithm attempts to freeze the maximal amount possible at every node a , starting with the root a_0 . Line (3) calculates τ' , the amount left to freeze after the available balance at a is frozen. Moreover, the algorithm subtracts the amount burned at a because that obligation should not transfer to the descendants of a . The remaining amount, τ' , should pass as an obligation to the descendants of a . In Line (6) the algorithm loops over all the transactions that take funds out of a . The loop proceeds in reverse chronological order, namely from the most recent transaction to the oldest transaction. Then Line (8) obligates the recipient of the most recent transaction from a to the maximal possible amount. This continues until all of τ' is passed as an obligation to the children of a . We analyze this algorithm and its properties in Appendix A.

Note that every address a is only visited once, and only after all its parents have been processed. Consequently, the running time is linear in the number of edges in the graph. More precisely, if there are V nodes and E edges in the graph then the running time is $O(V + E)$ thanks to the data structure we use that keeps edges in a chronologically sorted order.

Once this algorithm completes, the contract will add the quantity $toFreeze(a)$ to the number of coins frozen at address a (recall that a might already have frozen coins due to a prior dispute). This means that a subsequent transfer out of address a will fail if it causes the balance of a to drop below the commula-

```

CalcFreeze( $t_0$ ):      // freeze trans.  $t_0=(v \rightarrow a_0)$ 
 $L :=$  (topological sort of the graph rooted at  $a_0$ )
                        // assuming the graph  $G$  is a DAG
 $oblig(a_0) := val(t_0)$ 
                        // the obligation of  $a_0$  due to  $t_0$ 
for each  $a$  in  $L$  in order do:      // start at  $a_0$ 
1:   $\tau := oblig(a)$ 
    // total obligation at  $a$  from parents
2:   $toFreeze(a) := \min(\tau, Bal(a))$ 
    // amount to freeze at  $a$ 
3:   $\tau' := \tau - toFreeze(a) - burnedAt(a)$ 
    // the amount left to freeze, but do not
    // pass burned amount downstream
4:  if  $\tau' \leq 0$ : continue    // all done with  $a$ 
5:   $E(a) := \text{sort}(\{t=(a \rightarrow b)\})$ 
    // the set of trans. from  $a$  sorted
    // in reverse chronological order
6:  for each  $t=(a \rightarrow b)$  in  $E(a)$  in order do:
    // for each outgoing edge from  $a$ ,
    // starting with the most recent one
7:     $ob := \min(\tau', val(t))$ 
8:     $oblig(b) += ob$  ;  $\tau' -= ob$ 
    // obligate recipient  $b$  to  $ob$  tokens
9:    if  $\tau' \leq 0$ : break
    // all done with  $a$ ,
    // terminate the inner loop on line (6)

```

Figure 4: The freeze algorithm for a DAG

tive frozen amount.

Implementing the algorithm. The ERC-20R contract needs to maintain enough state to support the freeze process. We introduce a new Solidity data

structure called `_spenditures` that serves two functions: (i) when asked to freeze a transaction, the contract needs to verify that the transaction took place within the dispute window, and (ii) for an address a , the contract needs to identify all downstream addresses that received funds from a after the disputed transaction took place.

The new Solidity data structure called `_spenditures` is illustrated in Figure 5.

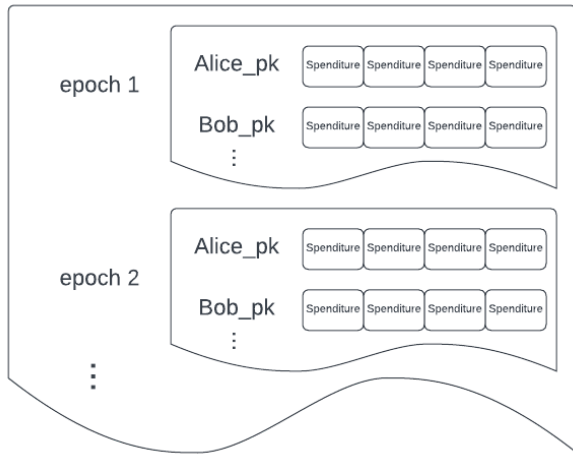


Figure 5: Spenditures nested map.

We refer to each sequence of Δ blocks as an epoch (e.g., $\Delta = 1000$). For each epoch, the `_spenditures` data structure lists all the transactions that took place during that epoch for each source address (Alice and Bob in the figure). Each Spenditure entry in Figure 5 contains the triple $(to, amount, blockNumber)$. Every time the contract processes a transfer request, it appends a Spenditure entry at the appropriate array in the `_spenditures` structure. A burn transaction causes a Spenditure triple to be added where the to field is null.

The freeze function takes in arguments $(epoch, from, index)$,

where $epoch$ identifies the epoch in which the disputed transaction took place, $from$ identifies the payer, and $index$ identifies the specific Spenditure in the data structure. The $index$ argument is provided by the governance contract by running an off-chain search over the `_spenditures` structure to locate the disputed transaction.

The freeze function checks that the requested transaction is within the dispute window, and if so, runs the freeze algorithm using the `_spenditures` data structure to determine the edges of the graph G from Figure 3. The final calculated freeze amounts per address are frozen. In addition, all these freeze amounts are recorded in a secondary array called `_claims` that is indexed by a newly generated 256-bit *ClaimID*. If the victim prevails in trial, this array is used to transfer the correct amount from every suspect address to the victim. The freeze function returns the *ClaimID* that indicates how to reverse the transaction.

The clean function. To minimize on-chain storage, our ERC-20R includes a `clean` function that takes in an epoch and an array of addresses. The function deletes all array entries in `_spenditures` with the specified epoch and sender addresses for epochs for which the dispute window has elapsed. Anyone can call the `clean` function and free up on-chain storage that is no longer needed.

2.2 Reversing a transaction

Once the trial over a disputed transaction concludes, the governance contract will call the `rejectReverse` function or the `Reverse` function on the ERC-20R or ERC-721R contract.

- An ERC-20R reversal takes as input a *ClaimID* which is an index into the `_claims` array. The entry indicates the obligation of each suspect account in the disputed transfer. The function transfers the specified amount from suspect accounts

to the original owner, and clears the freezes.

- An ERC-721R reversal takes as input *TokenID* and an *index* into `_owners` array. This indicates the owner prior to the disputed transaction. The asset is then transferred to the original owner, and the asset is unfrozen.

3 The Arbitration Process

We now turn to the governance process where judges examine the evidence on a disputed transaction and decide if the transaction should be reversed. Some of the key issues in designing this process are: (i) how judges are selected, (ii) how judges are compensated, and (iii) how to discourage misbehaving judges, such as judges who take bribes or make bad decisions on disputed transactions.

This process is orchestrated by a governance contract. A single governance contract can govern many 20R and 721R contracts. Judges vote on cases by either using an on chain or off chain voting system. Once enough votes on a case are collected, the governance contract calls the appropriate function on the 20R or 721R contract to either execute the reversal or dismiss the case. To ensure that judges make independent decisions, it is important that votes remain secret until sufficiently many votes are cast. This could be done, for example, using some flavor of a commit-and-reveal voting scheme, or any other semi-private voting scheme.

3.1 Selecting judges

We envision a large pool of available judges who will be compensated for their work. When a freeze request is submitted, the governance contract selects a *random and unpredictable* quorum of n judges from the pool. The value of n can be fixed, say $n = 12$, or it can increase with the size of the disputed transaction,

so that deciding on a large transaction requires more judges. This random selection is best implemented using a randomness beacon [2]. This quorum of judges decides on the initial freeze request, and later decides whether to approve or reject the reversal request.

Who can join the pool of judges and preside over cases? One can envision a process that requires a real world identity, a requirement for qualifications, and a statement of conflicts, much like real-world judge selection. We will leave the details of how to admit applicants into the pool of judges to future policy discussions of the exact mechanism.

3.2 Compensating judges

Every freeze request to the governance contract is accompanied by some stake from the party requesting the freeze. This stake can be used to compensate judges for their work. We state the following principle in deciding how to compensate judges: When judges submit their vote on a freeze request or on a reversal decision, they are compensated for their work. However, *the compensation amount is independent of their voting decision.*

If judges approve a freeze request, then the claimant's staked amount, minus fees, remains locked in the governance contract. If later the claimant loses the trial, the staked amount can compensate the defendant for their effort. If the claimant wins the trial, they could get back their excess stake. Importantly, if the judges reject the initial freeze request, then the claimant's staked amount minus the judges' fixed fees, should be burned.

Priority fees. While victims must provide a minimal stake along with a freeze request, they can optionally provide additional stake if they so choose. This extra "tip" from the victim could potentially indicate the priority of the case. Judges could rule on cases by priority rather than by chronological order. The tip

could either be burned, or given to the prevailing party, or some combination of the two.

3.3 Discouraging judicial misbehavior

A judge might fail to vote in a timely manner, or they might frequently vote with the extreme minority, potentially indicating an issue with their decision making. In either case, the governance contract could remove such a judge from the pool.

A more interesting question is how to prevent bribery, where a party bribes judges to vote in its favor. There are several technical solutions that could help partially mitigate this issue. One approach is to use a secrecy-preserving process for selecting the quorum of judges to preside over a case: a selected judge will learn that they were selected; however, on their own, they will not know who the other judges are, nor will they be able to prove to anyone that they were selected. Once all the selected judges vote, the set of judges is revealed along with a proof that the correct set voted. Importantly, the proof is revealed by the posted data from the selected judges, not by a set of trustees. This way, a party who wishes to bribe a judge will not know who to bribe because the set of judges is only revealed after they all voted. This makes it more difficult (but not impossible) to bribe judges. This mechanism can be implemented using techniques related to single secret leader election (SSLE) [3].

4 Discussion and Extensions

Safe token burns. Consider a contract that supports token burns, such as a wrapped asset. The contract could choose to only process a burn request for tokens for which the dispute window has elapsed. This will prevent losses due to burns, but at the cost of delaying some burn requests.

Reversibility is viral. An exchange that processes a swap of one 20R token for another 20R token is safe to do so without delay: if one side of the transaction is later claimed to have come from stolen funds, the exchange can initiate a reversal on the other side of the transaction. However, exchanging a 20R token for a non-reversible ERC-20 token is more dangerous. The exchange might only approve such a swap after the dispute window has elapsed on the 20R side of the transaction. This introduces a delay when swapping a reversible token for a non-reversible one. Thus, once some key tokens become reversible, other tokens are incentivized to do the same to eliminate this delay. In other words, reversibility is viral.

Appeals. Future work may consider an appeals process, in which the party who loses the trial can request another group of judges to rule on the disputed transaction. We do not explore this here.

Partial reversals. In some cases judges may opt for a partial reversal, where part of the funds go back to the owner, and part stay with the recipient. While we focused on complete reversals, the API can support partial reversals.

Prioritizing freeze requests. To speed up the processing of freeze requests, validators could optionally prioritize freeze transactions in the mempool over other transactions.

Grouping disputes. A transaction that sends coins from address A_1 to A_2 is often accompanied by a transaction from A_2 to A_1 . For example, A_1 might send DAI to A_2 , and in exchange A_2 sends an NFT to A_1 . Suppose that both DAI and the NFT support our reversible standards. Then, if A_1 asks to reverse the DAI transaction, A_2 will likely ask to reverse the NFT transaction. As an optimization, both claims can be decided jointly by the same quorum of judges.

5 Conclusion

In this paper, we proposed extensions to ERC-20 and ERC-721 that introduce a short time window when a transaction can be reversed. If adopted widely, these standards can protect the blockchain community from large financial losses. We hope that this paper will generate more discussion of reversibility as well as new designs to address the challenges that it raises.

Acknowledgments. This work was supported by NSF, the Simons Foundation, NTT Research, Coinbase, and UBRI.

A Details of the freeze algorithm

In this section we extend and further analyze the freeze algorithm in Figure 4.

A.1 Properties of the algorithm

First, let us examine some properties of the freeze algorithm in Figure 4. Consider the two transaction graphs in Figure 6. Both show the flow of funds, starting with a disputed transaction from v to a_0 . Let us assume that when *freeze* is called, the balance at a_0 and a_1 is zero in both graphs. As a result, in both graphs, calling *freeze* on the transaction from v to a_0 will transfer the obligation to addresses a_2 and a_3 , which is where the funds reside at the time of the freeze.

In the graph G_1 the transaction from a_1 to a_3 is the most recent, and therefore the algorithm will freeze 10 tokens at a_3 and zero tokens at a_2 . This may seem unfair because both a_2 and a_3 may have received a portion of the disputed funds. We stress, however, that in this case there is no “correct” answer because it is not possible to determine how exactly a_1 split the 20 tokens at its disposal (10 from the disputed transaction and 10 from a prior balance). Our algorithm’s

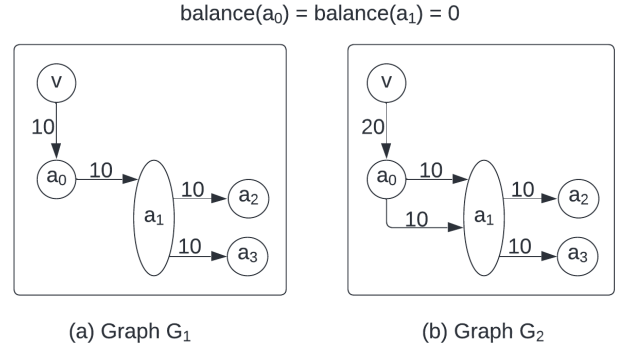


Figure 6: Two graphs where transactions are ordered chronologically from top to bottom. In both graphs, the balance at a_0 and a_1 is zero at the time of the freeze. In G_1 the starting balance at a_1 is 10.

decision is as “correct” as any other answer.

Interestingly, an algorithm that tries to split the obligation evenly between a_2 and a_3 in the graph G_1 can lead to paradoxical results. Consider the graph G_2 . The algorithm needs to obligate a_2 and a_3 to a total of 20 tokens. By the discussion in Section 2 (Example 3), address a_3 bears the full responsibility for the *second* transaction from a_0 to a_1 , and will therefore be obligated to 10 tokens as a result of this transaction. Now, if a_3 were to bear a portion of the responsibility for the *first* transaction from a_0 to a_1 , then the total obligation passed to a_3 would exceed 10 tokens. Since the balance at a_3 may only be equal to 10, this strategy will cause the algorithm to fail to freeze 20 tokens in total. Consequently, splitting the obligation between a_2 and a_3 in such cases may lead to freezing a reduced number of tokens, thereby preventing the victim from recovering the disputed funds.

Our algorithm does not suffer from this issue: Theorem 1 below shows that our freeze algorithm in Figure 4 always freezes the correct amount: 10 tokens in the graph G_1 and 20 tokens in the graph G_2 .

A.2 Graphs with cycles

The freeze algorithm in Figure 4 relies on a topological sort of the nodes in the graph G . The topological sort only exists when the graph is acyclic. Here we extend the algorithm to any directed graph, even one that includes cycles. We use a pre-processing phase that eliminates cycles.

Let us first see an example. Suppose the graph contains a transaction $t = (a_0 \rightarrow a_1)$ for five coins, and a subsequent transaction $t' = (a_1 \rightarrow a_0)$ for three coins. This simple cycle means that a_0 sent five coins to a_1 , and a_1 subsequently sent three coins back to a_0 . For the purpose of our freeze algorithm, we can replace both transactions by a single transaction from a_0 to a_1 for two coins, and eliminate the cycle.

More generally, let $t_0 = (v \rightarrow a_0)$ be the disputed transaction. The pre-processing step scans the graph rooted at a_0 to look for a directed cycle of transactions c_0, \dots, c_{k-1} where $c_i = (b_i \rightarrow b_{i+1 \bmod k})$ that were all posted after the disputed transaction. Let c be the transaction of smallest value along the cycle, breaking ties arbitrarily. Let $d := \text{val}(c)$. We can eliminate the cycle by removing transaction c from the graph, and subtracting d from the value of every other transaction in the cycle. The algorithm can repeat this process until the graph is free of cycles. At this point, the algorithm from Figure 4 can be applied to the resulting graph.

A.3 Two correctness theorems

Finally, we argue that the algorithm will freeze the correct amount. The first theorem shows that the algorithm will freeze the correct total amount. The second theorem shows that the algorithm will not over-freeze funds at any single address.

Theorem 1. *Suppose that no burn transactions are processed between the disputed transaction t_0 and the freeze transaction t_f . Moreover, there are no*

prior freezes in the system. Then, if the disputed transaction t_0 sends s tokens to address a_0 , then the freeze algorithm in Figure 4 will freeze a total of exactly s tokens. In particular, when the algorithm terminates we have

$$s = \sum_a \text{toFreeze}(a). \quad (2)$$

Proof sketch. First, let us assume that the directed graph rooted at a_0 is a DAG. We argue that the algorithm satisfies two properties:

- *Property 1.* When an address a is processed in the main loop, let us call the quantity τ' computed on Line (3) the **excess obligation at a** , namely the obligation not covered by a 's current balance. We claim that when the algorithm completes processing address a , the quantity τ' is added to the total obligation of the children of a .
- *Property 2.* For a transaction $t : a \rightarrow b$, let $ob(b, t)$ be the obligation that the algorithm passes to b as a result of t . We claim that for every transaction $t : a \rightarrow b$ we have

$$ob(b, t) \leq \text{val}(t).$$

This means that the funds that transaction t sends to b (and its descendants) is sufficient to cover the full obligation that the algorithm passes to b .

Together these two properties prove the lemma: at every address, the exact excess obligation is passed to its children, and there are sufficient funds at the descendants to cover these obligations.

Let us prove these two properties. Property 2 is immediate from Line (8) in the algorithm. This line ensures that for every transaction t , at most $\text{val}(t)$ is added to the obligation at b .

Property 1 follows from the fact that no coins are burned. In particular, consider the node a in Figure 7. The total obligation passed to a is $\tau := x_0 + \dots + x_m$.

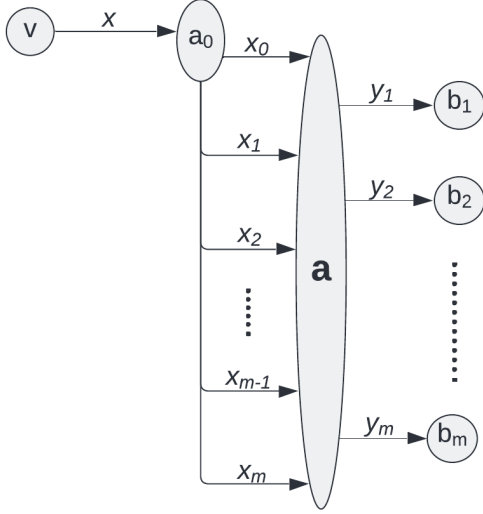


Figure 7: A sequence of $m + 1$ incoming transactions to a from its parent a_0 , along with m outgoing transactions from a , interleaved in time. Incoming transaction number i obligates a_0 to x_i tokens, and outgoing transaction number j sends y_j tokens to b_j .

Let $Bal(a)$ be the balance at a just prior to the freeze transaction. Then, by Property 2 and because no tokens were burned we know that

$$\tau \leq Bal(a) + \underbrace{(y_1 + \dots + y_m)}_{\text{amount that left } a}.$$

Moreover, by Line (3) we know that $\tau - \tau' = Bal(a)$ and therefore

$$\tau' \leq y_1 + \dots + y_m. \quad (3)$$

But when (3) holds, the loop in Lines (6) to (9) is guaranteed to add τ' to the total obligations of the children of a . This proves Property 1.

To complete the proof we observe that the cycle elimination process from Section A.2 does not affect the number of frozen tokens. Hence, the lemma applies equally well to non-DAG graphs, assuming the cycle elimination process is applied first. \square

Next, we show that the algorithm will not over-obligate an address. This is where we rely on the reverse chronological order loop in Line (6). Consider again the transactions out of a in Figure 7. We show that for every $j \in \{1, \dots, m\}$, the address b_j will bear no obligation for funds that arrived at a after transaction number j . To define this, for a transaction $t : a \rightarrow b$ we use the following terminology:

- $ob(b, t)$ is the obligation passed on to b as a result of transaction t . In other words, $ob(b, t)$ is the value added to $oblig(b)$ when the algorithm processes the transaction t .
- $obsum(a, t)$ is the sum of all the obligations that were added to $oblig(a)$ as a result of the transactions that sent funds to a prior to t . For example, for a transaction $t_j : a \rightarrow b_j$ in Figure 7 we have

$$obsum(a, t_j) = x_0 + \dots + x_{j-1}$$

The following theorem shows that for a transaction $t : a \rightarrow b$ we have $ob(b, t) \leq obsum(a, t)$. This means that the obligation passed to b will not exceed the obligations passed to a due to transactions that preceded t . In other words, b will not bear responsibility for an amount transferred to a after transaction t .

Theorem 2. Suppose that no burn transactions are processed between the disputed transaction t_0 and the freeze transaction t_f . Then when the algorithm terminates, for every transaction $t : a \rightarrow b$ we have

$$ob(b, t) \leq obsum(a, t) \quad (4)$$

Proof. Let us prove the theorem for the graph in Figure 7. The exact same argument applies to other graphs. Let $Bal(a)$ be the balance at address a at the time of the freeze. As usual, let

$$\tau = x_0 + \dots + x_m. \quad (5)$$

Now, fix some $j \in \{1, \dots, m\}$ and let t_j be the transaction $a \rightarrow b_j$ in Figure 7. Observe that if

$$\tau \leq \text{Bal}(a) + \sum_{i=j+1}^m y_i$$

then $ob(b_j, t_j) = 0$ and the theorem follows trivially. Hence, we can assume that

$$\tau > \text{Bal}(a) + \sum_{i=j+1}^m y_i.$$

In this case, by definition of the loop in Lines (6) to (9) in the algorithm we know that

$$ob(b_j, t_j) = \min \left(y_j, \tau - \text{Bal}(a) - \sum_{i=j+1}^m y_i \right). \quad (6)$$

Since no coins were burned at a we know that

$$\sum_{i=j}^m x_i \leq \text{Bal}(a) + \sum_{i=j+1}^m y_i$$

and therefore by (6) we obtain

$$ob(b_j, t_j) \leq \min \left(y_j, \tau - \sum_{i=j}^m x_i \right).$$

It now follow by (5) that

$$ob(b_j, t_j) \leq x_0 + \dots + x_{j-1} = \text{obsum}(a, t_j)$$

as required. This completes the proof. \square

References

- [1] Sarah Ahmed. Bitgrail hack - one of the largest crypto hacks in history. [link](#), 2022.
- [2] The beacon chain. [link](#), 2018.
- [3] Dan Boneh, Saba Eskandarian, Lucjan Hanzlik, and Nicola Greco. Single secret leader election. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, page 12–24, 2020.
- [4] Ed Caesar. How a young couple failed to launder billions of dollars in stolen bitcoin. [link](#), 2022.
- [5] Chainalysis. The 2022 crypto crime report. [link](#), 2022.
- [6] Obulapathi N Challa. Reversecoin: Worlds first cryptocurrency with reversible transactions. [link](#), 2014.
- [7] Vishal Chawla. Uniswap liquidity provider hacked for \$8 million in phishing attack. [link](#), 2022.
- [8] Evelyn Cheng. Japanese cryptocurrency exchange loses more than \$500 million to hackers. [link](#), 2018.
- [9] CryptoSec. Documented timeline of DeFi exploits. [link](#), 2022.
- [10] CryptoSec. Documented timeline of exchange hacks. [link](#), 2022.
- [11] Dean Eigenmann. Reversible token. [link](#), 2018.
- [12] Elliptic. Elliptic NFT report 2022 edition: NFTs and financial crime. [link](#), 2022.
- [13] Guy Eshet. Refunds for NFTs – the new ERC721R standard. [link](#), 2022.
- [14] Sam Kessler. Typo moves \$36M in seized JUNO tokens to wrong wallet. [link](#), 2022.

- [15] Olga Kharif, Sidhartha Shukla, and Bloomberg. Hackers just stole \$100 million in crypto from harmony’s horizon bridge. [link](#), June 2022.
- [16] Ronin’s Newsletter. Community alert: Ronin validators compromised. [link](#), 2022.
- [17] Muyao Shen. Hacker moves crypto stolen from ronin breach to help cover its tracks. [link](#), 2022.
- [18] Cryptopedia Staff. What was The DAO? [link](#), 2022.
- [19] Smiljanic Stasha. Cryptocurrency hacking statistics: Facts on hacking crypto. [link](#), Feb. 2022.
- [20] Chainalysis Team. The kucoin hack: What we know so far and how the hackers are using DeFi protocols to launder stolen funds. [link](#), 2020.
- [21] Jordan Tuwiner. What was the Mt. Gox hack? [link](#), 2022.
- [22] Fabian Vogelsteller. rICO — the reversible ICO. *Medium*, 2020.
- [23] Wikipedia contributors. Poly network exploit — Wikipedia, the free encyclopedia. [link](#), 2022. [Online; accessed July-2022].
- [24] Florian Zandt. Infographic: The biggest crypto heists. [link](#), 2022.