Resolving the Multiple Withdrawal Attack on ERC20 Tokens

Reza Rahimian, Jeremy Clark, Shayan Eskandari Concordia University

Abstract—Custom tokens are an integral component of decentralized applications (dapps) deployed on Ethereum and other blockchain platforms. For Ethereum, the ERC20 standard is a widely used token interface and is interoperable with many existing dapps, user interface platforms, and popular web applications (e.g., exchange services). An ERC20 security issue, known as the multiple withdrawal attack, was raised on GitHub and has been open since October 2017. The issue concerns ERC20's defined method approve () which was envisioned as a way for token holders to give permission for other users and dapps to withdraw a capped number of tokens. The security issue arises when a token holder wants to adjust the amount of approved tokens from N to M. If malicious, a user or dapp who is approved for N tokens can front-run the adjustment transaction to first withdraw N tokens, then allow the approval to be confirmed, and withdraw an additional M tokens afterwards. In this paper, we evaluate 10 proposed mitigations for this issues and find that no solution is fully satisfactory. We then propose 2 new solutions that mitigate the attack, one of which fully fulfills constraints of the standard, and we also show a general limitation in addressing this issue from ERC20's approve method.

Index Terms—Ethereum; ERC20 tokens; Blockchain;

1. Introduction

Ethereum is a public blockchain proposed in 2013, deployed in 2015 [1], and has the second largest market cap at the time of writing¹. It has a large development community which track enhancements and propose new ideas.² Ethereum enables decentralized applications (dapps) to be deployed and executed. dapps can accept and transfer Ethereum's built-in currency (ETH) or might issue their own custom currency-like tokens for specific purposes. Tokens might be currencies with different properties than ETH, they may be required for access to a dapp's functionality or they might represent ownership of some off-blockchain asset. It is beneficial to have interoperable tokens with other dapps and off-blockchain webapps, such as exchange services that allow tokens to be traded.

Toward this goal, the Ethereum project accepted a proposed standard (called ERC20 [2]) for a set of methods which standard-compliant tokens should implement. In terms of object oriented programming, ERC20 is an interface that defines abstract methods (name, parameters, return types) and provides guidelines on how the methods should be implemented, however it does not provide an actual concrete implementation (see Figure 1).

Figure 1. The ERC20 standard defines 6 methods and 2 events that must be implemented. Using approve and transferFrom in the existence of a race condition may lead to a "multiple withdrawal attack"

Since the introduction of ERC20 in November 2015. several vulnerabilities have been discovered. In October 2017, a security issue called "Multiple Withdrawal Attack" was opened on GitHub [3], [4]. The attack originates from two methods in the ERC20 standard for approving and transferring tokens. The use of these functions in an adverse environment (e.g., front-running [5]) could result in more tokens being spent than what was intended. This issue is still open and several solutions have been made to mitigate it. The authors of the ERC20 standard [2] reference two sample implementations: OpenZeppelin [6] and ConsenSys [7]. OpenZeppelin mitigates the attack by introducing two additional methods to replace the approve method, and the ConsenSys implementation does not attempt to resolve the attack. Additional implementations have a variety of different trade-offs in mitigating the issue (see Section 3).

Contributions. In this paper, we evaluate 10 proposed mitigations for the multiple withdrawal attack. We develop a set of criteria that encompass backwards compatibility, interoperability, adherence to the ERC20 standard, and attack mitigation. The summary is provided in Figure 2. Since no mitigation is fully satisfactory, we develop two additional solutions based on the Compare and Set (CAS) pattern[8]—a widely used lock-free synchronization strategy that allows comparing and setting values atomically. We study in detail possible implementations of ERC20's approve and transferFrom methods. We argue that a CAS-based approach can never adequately deploy a secure approve method while adhering to the ERC20 standard. We then propose a secure implementation of the transferFrom method that mitigates the attack and fully satisfies the ERC20 standard.

^{1. [2019-02-11]} https://coinmarketcap.com/currencies/ethereum/

^{2. [2019-02-11]} https://www.coindesk.com/data

#	Proposed	Is it backward	Does it remediate vulnerable	Does it allow non-zero	Does it allows zero	Does it mitigate the		
#	solution	compatible?	functions?	allowances?	token transfers?	attack?		
	Suggested Solutions							
1	Enforcement by UI	It does not change any code	It does not change any code	By default approve method	By default transferFrom method	Race condition can occur		
2	MiniMe Token	It adds only one line to approve method	Only forces allowance to be zero before non-zero values	If it is already zero, otherwise in two calls	By default transferFrom method	Race condition can occur		
3	Minimum viable token	It does not implement vulnerable functions	It does not implement approve function	It does not implement approve function	It does not implement transferFrom function	By not addressingvulnerable functions		
4	Approving trusted parties	It does not change any code	It depends on code verification	By default approve method	By default transferFrom method	It is not a universal solution		
5	Monolith DAO	It adds two new functions	It does not change any code	It adjusts allowance	By default transferFrom method	By using two new methods		
6	Alternate approval function	It adds one new function	It does not change any code	By using new method	By default transferFrom method	By using new method		
7	Changing ERC20 API	It adds new overloaded approve method	By new method with three parameters	By using new method	By default transferFrom method	By using new method		
8	New token standards	It introduces new API	✓	√	✓	√		
9	Detecting token transfers	It adds two lines to approve method	√	It locks down allowance in case of any token transfer	By default transferFrom method	It blocks legit and non- legit allowances as well		
10	Keeping track of remaining tokens	It adds three lines to the approve method	√	✓	By default transferFrom method	Race condition can occur		
			New Pro	posals				
11	Proposal 1: Securing approve method	It adds new codes to the approve method	√	It adjusts the allowance	√	√		
12	Proposal 2: Securing transferFrom method	It adds new codes to transferFrom method	It secures transferFrom method	By default approve method	√	√		

Figure 2. Comparison of 10 proposals and 2 contributed by this paper. In our first proposal, CAS is used to mitigate the attack by comparing transferred tokens with new allowance. It is not fully ERC20-compliant since the allowance result does not always match what is requested. In our second proposal, a new local variable is defined to keep track of transferred tokens and prevents transfers in the case of already transferred tokens.

2. Preliminaries

2.1. How the multiple withdrawal attack works

According to the ERC20 API definition, the approve function³ allows a spender (*e.g.*, user, wallet or other smart contracts) to withdraw up to an allowed amount of tokens from token pool of the approver. If this function is called again, it overwrites the current allowance with the new input value. On the other hand, the transferFrom function allows the spender to actually transfer tokens from the approver to anyone they choose (importantly: not necessarily themselves). The contract updates balance of transaction parties accordingly.

An adversary can exploit the gap between the confirmation of the approve and transferFrom functions since the approve method replaces the current spender allowance with the new amount, regardless of whether the spender already transferred any tokens or not. This functionality of the approve method is shaped by the language of the standard and cannot be changed. Furthermore, while variables change and events are logged, this information is ambiguous and cannot fully distinguish between possible traces. Consider the following illustration:

3. We use the term method and function interchangeably.

- Alice allows Bob to transfer N tokens on her behalf by broadcasting approve (_Bob, N).
- Later, Alice decides to change Bob's approval from N to M by broadcasting approve (_Bob, M).
- Bob notices Alice's second transaction after it is broadcast to the Ethereum network but before it is added to a block.
- 4) Bob front-runs (using an asymmetric insertion attack [5]) the original transaction with a call to transferFrom(_Alice, _Bob, N). If a miner is incentivized (e.g., by Bob offering high gas) to add this transaction before Alice's, it will transfer N of Alice's tokens to Bob.
- Alice's transaction will then be executed which changes Bob's approval to M.
- 6) Bob can call transferFrom method again and transfer M additional tokens by broadcasting transferFrom (_Alice, _Bob, M).

In summary, in attempting to change Bob's allowance from N to M, Alice makes it possible for Bob to transfer N+M of her tokens. We operate on the assumption that a secure implementation would prevent Bob from withdrawing Alice's tokens multiple times when the allowance changes from N to M (see Figure 3).

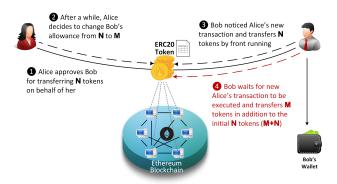


Figure 3. Possible multiple withdrawal attack in ERC20 tokens when Alice changes Bob's allowance from N to M. By front-running, Bob is able to move N+M tokens from token pool of Alice. Solutions consider Bob's initial transfer of N tokens (in step 3) to be legitimate and try to prevent the second transfer of the additional M tokens in step 4.

2.2. Why mitigation is important

ERC20 tokens are important component of Ethereum's supplementary financial system. Read two FC papers and add some trivia from them here, with citations. ERC20 tokens have many financial (as well as non-financial) uses and could hold considerable value (potentially exceeding the value of Ether itself). Furthermore, ERC20 tokens that have already been issued cannot easily migrate to a new secure implementation and should these tokens appreciate in value in the future. Finally, firms that hold ERC20 tokens require assurance of their security, particularly in the case that they require their financial statements to be audited—an issue like this could lead to further hesitation by auditors.

2.3. Where to prevent this attack

In general, there are a few logical places to address this attack. Ideally the token author (instead of the token holders) would mitigate the attack within the ERC20 implementation itself. Since two methods are involved in the attack, it could be addressed within the approve and/or transferfrom method. By contrast, token owners have no control over the implementation of the contract, and so token owners are relegated to mitigating the attack by carefully monitoring the contract around the time allowance adjustments are made, and possibly adopt new approaches to adjusting allowances. We start with this latter case.

Prevention by token owners. Consider Alice, a token holder using a web app (e.g., a user interface deployed with web3) to adjust Bob's allowance. If this UI is written to the ERC20 standard, Alice will only have approve available to her. The ERC20 authors [2] advise Alice against directly changing Bob's allowance from N to M. Instead, she should set the approval to 0 and then it to M. Presumedly, Alice will not do the second approval (setting it from 0 to M) if she sees that Bob withdraws N tokens before her N to 0 adjustment is confirmed. How will Alice know whether Bob withdraws first? The answer depends on how deeply her webapp can monitor the blockchain. One option is to monitor the variable that records Bob's

allowance (technically an on-blockchain helper dapp could also do this). However if she sees Bob's allowance at 0 after initiating the first adjustment, it does not tell her that Bob did not withhdraw N-both methods result in Bob having an allowance of 0 so either (or both) could have run. Next, she might rely on events passed from the contract to her web3 app. Transfer events as specified in ERC20 to record the transfer parameters (i.e., address _from, address _to, uint256 _tokens). Alice's webapp could filter the transfers and display only transfers matching her address in the from field. The displayed transfers will include any transfer Bob makes, however Bob can provide any address in _to, not just his own, and the event does not report who authorized the transaction (i.e., it does not log msg.sender). Therefore Alice cannot always unambiguously rely on events alone to determine Bob did not transfer funds. If Alice's web app is beyond web3 and runs a full node maintaining blockchain state, it could reliably detect Bob's transfer. The takeaway from all these options is that prevention by token owners has some undesirable properties: (1) it splits adjustments into two transactions, (2) because the first transaction needs to be confirmed before the second is initiated, it takes time to complete, and (3) to precisely mitigate this attack, a heavyweight web app is needed to inspect deeper than variable state changes and events. For these reasons, we concentrate on mitigating the attack in the contract itself. If successful, adjustments can be made with a single function call from any lightweight user interface without any additional monitoring of the contract.

Prevention by token author in approve. Securing approve method could be a solution and prevent the attack. By using compare and set (CAS) pattern [8], approve method can change spender allowance from N to M atomically (i.e., comparing new allowance with transferred token and set it accordingly). Comparison part of CAS requires knowledge of previously transferred tokens that will reveal any token transfer in case of frontrunning. Although this is promising approach, setting new allowance in approve method must satisfy ERC20 constraint that dictates "If this function is called again it overwrites the current allowance with _value" [2]. This does not allow any adjustments in allowance while it is a prerequisite for securing approve method. For example, considering front-running by Bob when Alice changes his allowance form 100 to 110, the approve method can reveal 100 transferred tokens by Bob and has to set Bob allowance to 10 (110-100=10). However, based on ERC20 constraints, it must not adjust the allowance and has to set it literally to 110. Consequently, this allows Bob to transfer 100+110=210 tokens in total. We implemented this approach in 4.1 and concluded that securing approve method cannot prevent the attack while adhering constraints of the ERC20 standard.

Prevention by token owner in transferFrom.

Based on ERC20 specifications, "approve functions allows _spender to withdraw from your account multiple times, up to the _value". Therefore, spender must not be able to transfer more than allowed tokens. That being said, transferFrom method can prevent transferring

of new tokens in case of already transferred tokens up to allowed amount. By comparing transferred tokens in transferFrom method, spender will be restricted to move solely the remaining tokens of his allowance. In case of trying to transfer more tokens than allowed, the transaction fails. For example, Alice's new transaction for increasing Bob allowance from 100 to 110, sets Bob allowance to 110 (approve method sets allowance regardless of transferred tokens). However, transferFrom method prevent the attack by not allowing Bob to transfer more than 10 tokens if he had already transferred 100 tokens. We implemented this approach in 4.2 and it mitigates the attack effectively.

2.4. What an ideal solution looks like

An important criterion for a solution is to adhere the specifications of ERC20 standard. Conforming with the standard ensures that new tokens are backward-compatible with already deployed smart contracts or web applications. We summarized ERC20 constraints [2] that must be satisfied by any sustainable solution:

- Calling approve function has to overwrite current allowance with new allowance.
- approve method does not adjust allowance, it sets new value of allowance.
- 3) Transferring 0 values by transferFrom method MUST be treated as normal transfers and fire the Transfer event as non-zero transactions.
- Spender will be allowed to withdraw from approver account multiple times, up to the allowed amount.
- 5) Transferring initial allowed tokens is considered as legitimate transfer. It could happen right after approval or before changing it.
- Race condition MUST not happen in any case to prevent multiple withdrawal from the account.
- 7) Introducing new methods violate ERC20 API and MUST be avoided for having compatible token. In other words, introducing new secure methods that are not defined in the initial ERC20 API, make them unusable with already deployed smart contracts. Therefore, smart contracts cannot use them to prevent the attack.

3. Evaluating Proposed Solutions

Several solutions have been proposed by Ethereum community—mostly from developers on GitHub[4]—to address the attack. There would be some trad-offs for each solution that needs to be evaluated in term of conforming with constraints of the standard and attack mitigation. We have examined each solution and explained possible ERC20 constraint violation.

3.1. Enforcement by User Interface (UI)

ERC20 standard recommends to (1) set allowance to zero before any non-zero values and (2) enforce approval processing check in UI instead of smart contract (see approve

Allows _spender_to withdraw from your account multiple times, up to the _value_amount. If this function is called again is

NOTE: To prevent attack vectors like the one described here and discussed here, clients SHOULD make sure to create use interfaces in such a way that they set the allowance first to a before setting it to another value for the same spender. THOUGH The contract itself shouldn't enforce it, to allow backwards compatibility with contracts deployed before

function approve(address _spender, uint256 _value) public returns (bool success)

Figure 4. Recommendation of ERC20 standard to mitigate multiple with drawal attack by enforcement in UI. $\label{eq:continuous}$

screenshot 4). This approach can not prevent the attack in the below scenario [19]:

- 1) Bob is allowed to transfer N Alice's tokens.
- 2) Alice publishes a new transaction that changes Bob's allowance from N to 0.
- 3) Bob front runs Alice's transaction and transfers N Alice's tokens (transferFrom sets Bob's allowance to 0 respectively).
- 4) Alice's transaction is mined and sets Bob's allowance to 0.
- 5) Now Alice publishes a new transaction that changes Bob's allowance from 0 to M.
- Alice's second transaction is mined, Bob now is allowed to move M Alice's tokens.
- 7) Bob transfers M Alice's tokens and in total N+M.

At step 3, Bob is able to transfer N tokens and consequently his allowance becomes 0 by transferFrom method. This is considered as a legitimate transaction since Alice has already approved it. The issue occurs after Alice's new transaction in step 5 to set Bob's allowance to M. In case of front-running by Bob, Alice needs to check Bob's allowance for the second time before setting any new value. However, she finds out Bob's allowance 0 in either case. In other words, she can not distinguish whether Bob's allowance is set to 0 because of her transaction in step 2 or Bob already transferred tokens by front-running is step 3. Someone may point out that Alice notices this by checking Transfer event logged by transferFrom function. However, if Bob had transferred tokens to someone else (like Carol), then Transfer event will not be linked to Bob, and, if Alice's account is busy and many people are allowed to transfer from it, Alice may not be able to distinguish this transfer from a legitimate one performed by someone else. Overall, this recommendation does not prevent the attack while tries to follow ERC20 recommendations for setting Bob's allowance to zero before any non-zero value. Hence, enforcement should be considered at contract level not UI to remove the gap between transactions and mitigate the attack. Interestingly, OpenZeppelin example implements a workaround in contract level that makes it inconsistent with the standard UI recommendations.

3.2. Minimum viable token

As suggested by Ethereum Foundation[17], we can boil down ERC20 standard to a very basic functionalities by implementing only essential methods. This prevents effecting of the attack by skipping implementation of vulnerable functions. While removing approve and transferFrom functions prevent the attack, it

```
// To change the approve amount you first have to reduce the addresses
// allowance to zero by calling `approve(_spender,0)` if it is not
// already 0 to mitigate the race condition described here:
// https://github.com/ethereum/FIPs/issues/20#issuecomment-263524729
require((_amount == 0) || (allowed[msg.sender][_spender] == 0));
```

Figure 5. MiniMeToken added code to approve method for allowing non-zero allowance values if it is already set to zero.

makes the token partially-ERC20-compliant. Golem Network Token (GNT⁴) is one of these examples since it does not implement the approve, allowance and transferfrom functions. According to ERC20 specifications[2], these methods are not OPTIONAL and must be implemented. Moreover, ignoring them will cause failed function calls from standard smart contracts that expect to interoperate with these methods. Therefore, we would not consider it as a backward compatible solution although mitigates the attack.

3.3. Approving trusted parties

Approving token transfer to parties that we trust, reduces risk of impacting by the attack. Non-upgradable smart contracts or trusted accounts can be considered safe for delegating token transfers on behalf of us. Because they do not contain any logic to take advantage of this vulnerability. However, upgradable smart contracts may add new code to their new versions that needs code reverification before approving token transfers. Similarly, approving token transfer to people that we trust could be considered as a mitigation plan. Nonetheless, this solution would have limited use cases and it could not be considered as a universal preventive approach.

3.4. MiniMeToken

MiniMeToken[14] follows ERC20 recommendation for setting allowance to zero before non-zero values. Token holder needs to execute two transactions, setting allowance from N to zero and then from 0 to M. MiniMeToken added a new line of code to the approve method for preventing the attack by enforcing this rule (see Figure 5). The red clause in line 225 (_amount == 0) allows setting of approval to 0 and blue condition checks allowance of _spender to be 0 before setting to nonzero values (i.e., If _spender allowance is 0 then allows non-zero values). Similar to 3.1, this solution will not prevent Bob from transferring N+M tokens. Because Alice would not be able to distinguish whether N tokens have been already drained by Bob or not. Considering the below scenario:

- 1) Alice decides to set Bob's allowance to 0.
- 2) Bob front-runs Alice's transaction and his allowance sets to 0 after transferring N tokens.
- 3) Alice's transaction is executed and sets Bob's allowance to 0 (Red clause passes sanity check).
- 4) Alice checks Bob's allowance and she will find it 0, so, she can not determine whether this

State	Input value (_value)	Current spender's allowance	Approve function result	New spender's allowance
1	Zero	Non-zero	Set to _value	0
2	Zero	Zero	Set to _value	0
3	Non-zero	Zero	Set to _value	_value
4	Non-zero	Non-zero	No result	No change

Figure 6. Functionality of default approve method in MonolithDAO token that enforces setting spender's allowance to zero before any non-zero values. It implements ERC20 recommendation for changes allowance from N to M in two steps $(N \rightarrow 0 \rightarrow M)$.

- was because of her transaction or Bob already transferred N tokens.
- Alice considers that Bob has not transferred any tokens and allows him for transferring new M tokens.
- Bob is able to transfer new M tokens and N+M in total.

3.5. MonolithDAO

MonolithDAO Token[18] implements two additional functions for allowance increase⁵ or decrease⁶. The default approve function has additional codes to enforce the owner for setting allowance to zero before nonzero values. It allows non-zero spender's allowance if it is already set to zero (see Figure 6). If the current spender's allowance is non-zero, decreaseApproval and increaseApproval functions have to be used instead of standard approve method as explained below:

- 1) Alice allows Bob to transfer N tokens by calling approve (_Bob, N). Alice used the default approve function consciously since current Bob's allowance is 0. So, he checked Bob's allowance before calling approve method.
- 2) After a while, Alice decides to decrease Bob's approval by M and runs decreaseApproval (_Bob, M).
- 3) Bob notices Alice's second transaction and front runs it by executing transferFrom(_Alice, _Bob, N).
- 4) Bob's transaction will be executed first and transfers N token to his account and the his allowance becomes 0 as result of this transfer.
- 5) Alice's transaction is mined after Bob's transaction and tries to decrease Bob's allowance by M. If Bob had already transferred more than M tokens, new Bob's allowance becomes negative and it fails the transaction. So, the transaction does not change Bob's remaining allowance and he would be able to transfer the rest (which is legitimate transfer since Alice has already approved it). If Bob had transferred less than M tokens, the new allowance will be applied and reduces Bob's allowance by M.

^{5.} increaseApproval(address _spender, uint addedValue)

^{6.} decreaseApproval(address _spender, uint _subtractedValue)

Although these two new complementary functions prevent the attack, they have not been defined in the initial specifications of ERC20 standard. Therefore, they can not be used by smart contracts that are already deployed on the Ethereum network and still call approve method for setting new allowance—and not increaseApproval or decreaseApproval. Moreover, ERC20 specifications does not define any increase or decrease of allowance. It only allows setting new allowances without adjustment. For example, if Alice has approved Bob for 100 tokens and wants to set it to 80, the new allowance should be 80 while using decrease methods will set it to 20 (100 - 80 =20). Comparatively, increase method sets new allowance to 180 while it has to set it to 80 again to be in-compliant with ERC20 specification. For these reasons, this solution would not be compatible with ERC20 standard and only is usable if approver or smart contract are aware of these supplementary methods.

3.6. Alternate approval function

Another suggestion[10] is to move security checks to another function called safeApprove⁷ that compare current and new allowance value and sets it if has not been already changed. By using this function, Alice uses the standard approve function to set Bob's allowance to 0 and for new approvals, she has to use safeApprove function. safeApprove takes the current expected approval amount as input parameter and calls approve method if previous allowance is equal to the current expected approval. By using this function, Alice will have one step more to read the current allowance and pass it to the new safeApprove method. Although this approach mitigates the attack by using CAS pattern[8], however, it is not backward compatible with already implemented smart contracts due to their unawareness of this new complementary function. In other words, the new safeApprove method is not defined in ERC20 standard and existing smart contracts would not be able to use this new safety feature.

3.7. Detecting token transfers

In order to set new allowance atomically, tracking of transferred tokens is required to detect token transfers before setting new allowances. If approve method reveals any transferred tokens due to front running, it throws an exception without setting new allowance. As suggested by [20], a flag can be used to detect whether any token has been transferred or not. transferFrom method sets this flag to true in case of any token transfer. approve method checks the flag to be false before allowing new approvals. This approach requires new data structure to keep track of used/transferred tokens for each spender. It can prevent front-running in the below scenario:

1) Alice runs approve (_Bob, N) to allow Bob for transferring N tokens. Since Bob's initial allowance is 0 and his corresponding flag=false, then sanity check passes and Bob's allowance sets to N (line 16).

```
7. safeApprove(address _spender, uint256 _currentValue, uint256 _value
```

Figure 7. approve method needs to be modified by adding a line of code like allowed [msg.sender] [_spender].used = false; between lines 16 and 17 to unlock spender flag for the next legitimate change. However, this change makes attack mitigation ineffective.

- 2) Alice decides to set Bob's allowance to 0 by executing approve (_Bob, 0).
- Bob front-runs Alice's transaction and transfers N tokens. Then, transferFrom turns his flag to true.
- 4) Alice's transaction is mined and passes sanity check because passed value is 0 in line 15.
- 5) Bob's allowance is set to 0 while his flag remains true. (approve method does not flip spender flags.)
- 6) Alice wants to change Bob's allowance to M by executing approve (_Bob, M). Since Bob already transferred N tokens (his flag=true), then transaction fails.
- Bob's allowance does not change and he cannot move more tokens than initially allowed.

Although this approach mitigates the attack, but it prevents any further legitimate approvals as well. Considering a scenario that Alice rightfully wants to increase Bob's allowance from N to M (two non-zero values). If Bob had already transferred number of tokens, Alice would not be able to change his approval. Because Bob's flag is set to true and line 15 does not allow changing allowance by throwing an exception (see Figure 7). Even setting allowance to 0 and then to M, does not flip the flag to false (There is no code for it in the approve method). So, It keeps Bob's allowance locked down and blocked further legitimate allowances. In fact, approve method needs a new code between lines 16 and 17 to set the flag to false. But it will cause another problem. After setting allowance to 0, spender flag becomes false and allows non-zero values even if tokens have been already transferred. Considering front-running by Bob in the below scenario:

- 1) Alice changes Bob's allowance from N to 0.
- 2) Bob transfers N tokens before allowance change and his used flag turns to true.
- 3) Alice's transaction is successful since _value=0. (The second condition is not evaluated although used flag is true).
- 4) Alice transaction turns used flag to false and sets Bob's allowance to 0.
- 5) Now Alice wants to set Bob's allowance from 0 to M, His flag is false and allowance is 0. So, Alice cannot distinguish whether Bob moved any token or not. Setting new allowance will allow Bob to transfer more tokens than Alice wanted.

In fact, resetting the flag in approve method will not fix the issue and makes attack mitigation ineffective. In short, this approach can not satisfy both legitimate

```
struct Allowance {
    uint initial;
    uint residual:
mapping(address => mapping(address => Allowance)) public allowances;
function approve(address spender, uint amount) public returns (bool) {
   Allowance storage _allowance = allowances[msg.sender][spender];
    // This test should not be necessary.
    allowance.initial = amount:
    _allowance.residual = spent < amount ? amount - spent : 0;
    Approval(msg.sender, spender, allowance.residual);
    return true:
```

Figure 8. Keeping track of remaining tokens by introducing new data

and non-legitimate scenarios. Nevertheless, it is a step forward by introducing the need for a new variable to track transferred tokens.

3.8. Keeping track of remaining tokens

This approach[16] is inspired by the previous solution and keeping track of remaining tokens instead of detecting transferred tokens. approve method uses these variables to set allowance accordingly. It uses modified version of data structure that used in the previous solution for storing residual tokens (see Figure 8). At the first look, it seems to be a promising solution by setting approval to zero before non-zero values. However, the highlighted code in approve method resembles the situation that is explained in 3.1. In case of front-running, both initial and residual variables will be zero and it would not be possible for Alice to distinguish if any token transfer have occurred due to her allowance change or Bob token transfer. To make it more clear, considering the below scenario:

- 1) Bob's allowance is initially zero (allowances [_Alice] [_Bob].initial=0) and his residual is zero as well
- Alice allows Bob to transfer N tokens that makes allowances[_Alice][_Bob].initial=N
- Alice decides to change Bob's allowance to M and has to set it to zero before any non-zero.
- Bob noticed Alice's transaction for setting his allowance to zero and transfers N tokens in advance. Consequently, transferFrom function his residual sets to zero (allowances[_Alice][_Bob].residual=0)as a sustainable solution.
- Alice's transaction for setting allowance to zero is mined and allowances[_Alice][_Bob].initial=0 and allowances [_Alice] [_Bob] .residual=0 Most of the new token standards are introduced to This is similar to step 1 that no token has been transferred. So, Alice would not be able to distinguish whether any token have been transferred or not.

```
// Standard ERC20 Approve Method
function approve(address _spender, uint256 _value) public returns (bool success)
    allowed[msg.sender][_spender] = _value;
      mit Approval(msg.sender, _spender, _value);
    return true;
// Atomic "Compare And Set" Approve Method function approve(address _spender, uint256 _currentValue, uint256 _newValue)
                                                        public returns (bool success)
   if (allowed[msg.sender][_spender] != _currentValue) { return false; }
    allowed[msg.sender][_spender] =
    emit Approval(msg.sender, _spender, _newValue);
emit Approval(msg.sender, _spender, _newValue, _currentValue);
return true;
```

Figure 9. Suggested ERC20 API Change by adding new approve method with three parameters to compare and set new allowance atomically

- Considering no token transfer by Bob, Alice approves Bob for spending new M tokens.
- Bob is able to transfer new M tokes in addition to initial N tokens.

Someone may think of using Transfer event to detect transferred tokens or checking approver balance to see any transferred tokens. As explained in 3.1, using Transfer event is not sufficient in case of transferring tokens to a third party. Checking approver balance also would not be an accurate way if the contract is busy and there are lot of transfers. So, it would be difficult for the approver to detect legitimate from non-legitimate tokens transfers. Overall, this approach cannot prevent the attack.

3.9. Changing ERC20 API

As advised by [23], changing ERC20 API could secure approve method by comparing current allowance of the spender and sets it to new value if he has not already been transferred any tokens. This allows atomic compare and set of spender allowance to make the attack impossible. This approach needs a new overloaded approve method with three parameters in addition to the standard approve method with two parameters (see Figure 9).

In order to use this new method, smart contracts have to update their codes to provide three parameters instead (allowances [_Alice] [_Bob] . residual=0) of current two, otherwise any approve call will use the standard vulnerable version with two parameters. Moreover, one more call is required to read current allowance and allowances [_Alice] [_Bob].residual and pass it to the new approve method. Additionally, new event definition needs to be added to ERC20 API to log an approval events with four arguments. For backward compatibility reasons, both three-arguments and new fourarguments events have to be logged. All of these changes makes this token contract incompatible with already deployed smart contracts. Hence, we would not consider it

3.10. New token standards

improve current functionality of ERC20 tokens and do not address the attack (see Table 1). Despite these enhancements, migration from ERC20 to new token standard would not be convenient and all deployed ERC20

Token Standard	A description of non-compliance with ERC20
ERC 223 [11]	It does not implement ERC20 approve + transferFrom mechanism by assuming it potentially insecure and inefficient.
ERC 667 [12]	It solves the problem of transfer function in ERC223 (i.e., The need to implement onTokenTransfer routing in the receiving contract). So, it does address the attack and uses the same code as ERC223 with a supplementary function.
ERC 721 [21]	Unlike ERC20 tokens that share the same characteristics, ERC721 tokens represent a nonfungible tokens (NFT) which are unique and non-interchangeable with each other. In addition to this differences, ERC721 does not implement transfer method of ERC20 standard and introduces a safe transfer function defined as safeTransferFrom.
ERC 777 [13]	This standard does not use transfer and transferFrom methods of ERC20 and implements send and operatorSend instead. Moreover, costly approve + transferFrom mechanism is replaced by tokensReceived function. Therefore, it would not be backward compatible with ERC20 requirements. Nevertheless a token contract may implement both ERC20 and ERC777 in parallel which still requires attack mitigation.
ERC 827 [15]	It uses OpenZeppelin[6] ERC20 implementation and defines three new functions to allow users for transferring data in addition to value in ERC20 transactions. This feature enables ERC20 tokens to have the same functionality as Ether (transferring data and value). In fact, it extends functionality of ERC20 tokens and not addressing the attack.
ERC 1155 [22]	It is improved version of ERC721 by allowing each Token ID to represent a new configurable token type, which may have its own metadata, supply and other attributes. ERC1155 aimed to remove the need to "approve" individual token contracts separately. Therefore, it does not implement any code to address the vulnerability
ERC 1377 [9]	It implements approve method with three parameters in addition to the ERC20 default approve with two inputs. Additionally, it uses OpenZeppelin[6] approach for increasing and decreasing approvals. We would consider it as mix of MiniToken and OpenZeppelin approaches that we discussed before.

TABLE 1. COMPARISON OF ERC STANDARDS IN TERMS OF COM
PLIANCY WITH ERC20 SPECIFICATION

tokens—168,092 tokens⁸—have to be redeployed. This also means update of trading platform who already listed ERC20 tokens. Our goal is to find a backward compatible solution instead of changing current ERC20 API or migrating tokens to a new standards. Despite expanded features and improved security properties of new standards, we would not consider them as a target solutions.

4. New mitigations

By this point, we have discussed 10 solutions to the multiple withdrawal attack and we evaluated them in terms of compatibility with the standard and attack mitigation (recall the summary in Table 2). Since none of them precisely satisfy the constraints of ERC20 standard, we now propose two new solutions to mitigate the attack.

```
8. https://etherscan.io/tokens, Accessed 18-Feb-2019
```

```
function transferFrom(address _from, address _to, uint256 _tokens)
public returns (bool success) {
    require(_to! = address(@));
    require(balances[_from] >= _tokens);
    require(balances[_from] >= _tokens);
    // Checks enough tokens
    require(allowed[_from][msg.sender] >= _tokens);
    // Checks enough allowance
    balances[_from] = balances[_from].sub(_tokens);
    allowed[_from][msg.sender] = allowed[_from][msg.sender].sub(_tokens);
    transferred[_from][msg.sender] = transferred[_from][msg.sender].add(_tokens);
    balances[_to] = balances[_to].add(_tokens);
    emit Transfer(_from, _to, _tokens);
    return true;
}
```

Figure 10. Modified version of transferFrom for keeping track of transferred tokens per spender.

```
function approve(address _spender, uint256 _tokens) public returns (bool success) {
    require(_spender |= address(0));
    uint256 allowedTokens = 0;
    uint256 initiallyAllowed = allowed[msg.sender][_spender].add(transferred[msg.sender][_spender]);

//Aprover reduces allowance
    if (_tokens <= initiallyAllowed){
        if (transferred[msg.sender][_spender] < _tokens){ // If less tokens had been transferred.
        allowedTokens = _tokens.sub(transferred[msg.sender][_spender]); // Allows the rest
    }

//Approver increases allowance
    else{
        allowedTokens = _tokens.sub(initiallyAllowed);
        allowedTokens = allowed[msg.sender][_spender].add(allowedTokens);
}

allowedTokens = allowed[msg.sender][_spender].add(allowedTokens);
return true;
}</pre>
```

Figure 11. Added code block to approve function to prevent the attack by comparing and setting new allowance atomically.

4.1. Proposal 1: Securing approve method

By implementing CAS [8] in approve method, new allowance can be set atomically after comparing with transferred tokens. This tracking requires adding a new variable to transferFrom method (see Figure 10). Since this is an internal variable, it is not visible to already deployed smart contracts and keeps the transferFrom function compatible. Similarly, a block of code is added to the approve function (see Figure 11) to work in both cases with zero and non-zero allowances. Added code to the approve function, compares new allowance—passed as _tokens argument to the function—with the current allowance of the spender and already transferred tokenshighlighted as allowed[msq.sender] [_spender] transferred[msg.sender][_spender]. Then it decides to increase or decrease current allowance based on this comparison. If the new allowance is less than initial allowance—sum of allowance and transferred variables—it denotes decreasing of allowance, otherwise increasing of allowance was intended. Modified approve function prevents the attack in either increasing or decreasing of the allowance.

Unlike other solutions, there is no need to set allowance from N to 0 and then to M. Token holder can directly change the allowance from N to M which is saving one transaction accordingly. The following examples make functionality of added codes more clear when decreasing or increasing allowance:

Scenario A. Alice approves Bob for spending 100 tokens and then decides to decrease it to 10 tokens.

- 1) Alice approves Bob for transferring 100 tokens.
- 2) After a while, Alice decides to reduce Bob's allowance from 100 to 10 tokens.
- 3) Bob noticed Alice's new transaction and transfers 100 tokens by front-running.
- 4) Bob's allowance is 0 and transferred is 100 (set by transferFrom function).

	Consumed Gas by the token		
Operation	TKNv1	TKNv2	Difference
Creating smart contract	1095561	1363450	25 %
Calling Approve function	45289	46840	4 %
Calling transferFrom function	44019	64705	47 %

Figure 12. Comparison of gas consumption between standard implementation of ERC20 token (TKNv1) and secured implementation of it (TKNv2).

- 5) Alice's transaction is mined and checks initial allowance (100) with new allowance (10).
- 6) As it is reducing, transferred tokens (100) is compared with new allowance (10). Since Bob already transferred more tokens, his allowance will be set to 0.
- 7) Bob is not able to move more than initial 100 approved tokens.

Scenario B. Alice approves Bob for spending 100 tokens and then decides to increase it to 120 tokens.

- 1) Alice approves Bob for transferring 100 tokens.
- 2) After a while, Alice decides to increase Bob's allowance from 100 to 120 tokens.
- 3) Bob noticed Alice's new transaction and transfers 100 tokens by front-running.
- 4) Bob's allowance is 0 and transferred is 100.
- 5) Alice's transaction is mined and checks initial allowance (100) with new allowance (120).
- 6) As it is increasing, new allowance (120) will be subtracted from transferred tokens (100).
- 7) 20 tokens will be added to Bob's allowance.
- 8) Bob would be able to transfer more 20 tokens (120 in total as Alice wanted).

In order to evaluate functionality of the new approve and transferFrom functions, we have implemented a standard ERC20 token (TKNv19) along side the proposed ERC20 token (TKNv210) on Rinkeby test network. Result of tests for different input values shows that TKNv2 can address multiple withdrawal attack by making front-running gain ineffective. Moreover, we compared these two tokens in term of gas consumption. TokenV2.approve function uses almost the same amount of gas as TokenV1.approve, however, gas consumption of TokenV2.transferFrom is around 47% more than TokenV1.transferFrom (see Figure 12). This difference is because of maintaining a new mapping variable for tracking transferred tokens. In term of compatibility, working with standard wallets (e.g., MetaMask) have not raised any transfer issue. This shows compatibility of the token with existing wallets.

In summary, we could use CAS pattern to implement a secure approve method that can mitigate the attack effectively. However, it violates one of ERC20 specifications that says "If this function is called again it overwrites the current allowance with _value". This implementation of approve method adjusts allowance based on transferred

tokens. Essentially, it would not be possible to secure the approve method without adjusting the allowance. Considering the below scenario:

- Alice decides to change Bob's allowance from N to M (M less than N in this example).
- Bob transfers N tokens by front running and transferred variable sets to N.
- 3) Alice's transaction is mined and approve method detects token transfer. If approve method does not adjust the allowance based on transferred tokens, it has to set it to M—to conform with the standard—which is allowing Bob to transfer more M tokens.

Therefore, approve method has to adjust the allowance according to transferred tokens, not based on passed input values to the approve method. Overall, there is no solution to secure approve method while adhering specification of ERC20 standard.

4.2. Proposal 2: Securing transferFrom method

As an alternative solution to Proposal 1, we can think of securing transferFrom method instead of approve function. As specified by ERC20 standard (see figure 13), the goal here is to prevent spender from transferring more tokens than allowed. Based on this assumption, we should not consider allowance as the main prevention factor. Transferred tokens can be considered as the main variable in our calculations. For example in the below situation, we can prevent the attack by securing transferFrom method and keeping approve function untouched to set allowance as specified by the token holder:

- 1) Alice allowed Bob for transferring 100 tokens and decides to set it to 70 after a while.
- 2) Bob front runs Alice's transaction and transfers 100 tokes (legitimate transfer).
- Alice transaction is mined and sets Bob allowance to 70 by the default approve method.
- Bob noticed new allowance and tries to move new tokens by running transferFrom (_Bob, 70). Since he already transferred more than 70, his transaction fails and prevents multiple withdrawal. Additionally, Bob's allowance stays as 70, although transferred tokens shows 100.

Here, allowance value can be considered as maximum allowance. It indicates that Bob is eligible to transfer up to specified limit if he has not already transferred any tokens. This impression is completely in accordance with ERC20 standard (see figure 13). In fact, there is no relation between allowance (allowed[_from] [msg.sender]) and transferred tokens (transferred[_from] [msg.sender]). The first variable shows maximum transferable tokens by a spender and can be changed irrelative to transferred tokens (*i.e.*, approve method does not check transferred tokens). If Bob has not already transferred that much of tokens, he would be able to transfer difference of it—allowed[_from] [msg.sender].sub(transfer

^{9.} https://rinkeby.etherscan.io/address/0x8825bac68a3f6939c296a40fc8078d18c2f66ac7

^{10.} https://rinkeby.etherscan.io/address/0xf2b34125223ee54dff48f715 67d4b2a4a0c9858b

transferFrom

Transfers _value amount of tokens from address _from to address _to , and MUST fire the Transfer event

The transferfrom method is used for a withdraw workflow, allowing contracts to transfer tokens on your behalf. This can be used for example to allow a contract to transfer tokens on your behalf and/or to charge fees in sub-currence. The function SCHILD between upless the few account based deliberately subjected the sender of the pressure size of the pressure process.

Note Transfers of 0 values MUST be treated as normal transfers and fire the Transfer event

function transferFrom(address _from, address _to, uint256 _value) public returns (bool success)

Figure 13. ERC20 transferFrom method definition that emphasizes on throwing an exception when the spender is not authorized to move tokens.

Figure 14. Securing transferFrom method instead of approve method can mitigate the attack by preventing more token transfer than allowed.

red[_from] [msg.sender]). In other words, transferred variable is a life time variable that accumulates transferred tokens regardless of allowance changes. This token is implemented as TKNv3¹¹ on Rinkby network and passed compatibility checks by transferring tokens between standard wallets. In terms of gas consumption, transferFrom function needs at about 37% more gas than standard transferFrom implementation which is acceptable for having a secure ERC20 token.

5. Conclusion

While this paper is a deep dive into a specific issue with ERC20, it also illustrates a number of higher level lessons for blockchain developers. What are they?

References

- [1] Ethereum. Ethereum project repository. https://github.com/ethereum, May 2014. [Online; accessed 10-Nov-2018].
- V. B. Fabian Vogelsteller. ERC-20 Token Standard. https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md, Nov. 2015. [Online; accessed 2-Dec-2018].
- [3] M. Vladimirov. Attack vector on ERC20 API (approve/transfer-From methods) and suggested improvements. https://github.com/ethereum/EIPs/issues/20#issuecomment-263524729, Nov. 2016. [Online; accessed 18-Dec-2018].
- [4] T. Hale. Resolution on the EIP20 API Approve / TransferFrom multiple withdrawal attack #738. https://github.com/ethereum/ EIPs/issues/738, Oct. 2017. [Online; accessed 5-Dec-2018].
- [5] S. Eskandari, S. Moosavi, and J. Clark. Sok: Transparent dishonesty: front-running attacks on blockchain. 2019.
- [6] OpenZeppelin. openzeppelin-solidity. https://github.com/ OpenZeppelin/openzeppelin-solidity/blob/master/contracts/token/ ERC20/ERC20.sol, Dec. 2018. [Online; accessed 23-Dec-2018].
- [7] ConsenSys. ConsenSys/Tokens. https://github.com/ConsenSys/ Tokens/blob/fdf687c69d998266a95f15216b1955a4965a0a6d/ contracts/eip20/EIP20.sol, Apr. 2018. [Online; accessed 24-Dec-2018].
- [8] Wikipedia. Compare-and-swap. https://en.wikipedia.org/wiki/ Compare-and-swap, July 2018. [Online; accessed 10-Dec-2018].
- $11.\ https://rinkeby.etherscan.io/address/0x5d148c948c01e1a61e280c8b2ac39fd49ee6d9c6$

- [9] J. C. Atkins Chang, Noel Bao, L. Chou, and D. Goh. Service-Friendly Token Standard. https://github.com/fstnetwork/EIPs/ blob/master/EIPS/eip-1376.md, Sept. 2018. [Online; accessed 12-Jan-2019].
- [10] E. Chavez. StandardToken.sol. https://github.com/kindads/erc20-token/blob/40d796627a2edd6387bdeb9df71a8209367a7ee9/contracts/zeppelin-solidity/contracts/token/StandardToken.sol, Mar. 2018. [Online; accessed 23-Dec-2018].
- [11] Dexaran. ERC223 token standard. https://github.com/ethereum/ EIPs/issues/223, Mar. 2017. [Online; accessed 12-Jan-2019].
- [12] S. Ellis. transferAndCall Token Standard. https://github.com/ ethereum/EIPs/issues/677, July 2017. [Online; accessed 12-Jan-2019].
- [13] J. B. Jacques Dafflon and T. Shababi. EIP 777: A New Advanced Token Standard. https://eips.ethereum.org/EIPS/eip-777, Nov. 2017. [Online; accessed 12-Jan-2019].
- [14] D. N. Jordi Baylina and sophiii. minime/contracts/MiniMeToken.sol. https://github.com/Giveth/minime/blob/master/contracts/ MiniMeToken.sol#L225, Dec. 2017. [Online; accessed 23-Dec-2018].
- [15] A. Lemble. ERC827 Token Standard (ERC20 Extension). https://github.com/ethereum/eips/issues/827, Jan. 2018. [Online; accessed 12-Jan-2019].
- [16] outofgas. outofgas comment. https://github.com/ethereum/EIPs/ issues/738#issuecomment-373935913, Mar. 2018. [Online; accessed 25-Dec-2018].
- [17] E. Project. Create your own crypto-currency). https://www. ethereum.org/token, Dec. 2017. [Online; accessed 01-Dec-2018].
- [18] P. Vessenes. MonolithDAO/token. https://github.com/ MonolithDAO/token/blob/master/src/Token.sol, Apr. 2017. [Online; accessed 23-Dec-2018].
- [19] M. Vladimirov. Implementation of 'approve' method violates ERC20 standard #438. https://github.com/OpenZeppelin/ openzeppelin-solidity/issues/438#issuecomment-329172399, Sept. 2017. [Online; accessed 24-Dec-2018].
- [20] N. Welch. flygoing/BackwardsCompatibleApprove.sol. https://gist.github.com/flygoing/2956f0d3b5e662a44b83b8e4bec6cca6, Feb. 2018. [Online; accessed 23-Dec-2018].
- [21] J. E. William Entriken, Dieter Shirley and N. Sachs. ERC-721 Non-Fungible Token Standard. https://github.com/ethereum/EIPs/ blob/master/EIPS/eip-721.md, Jan. 2018. [Online; accessed 12-Jan-2019].
- [22] P. C. Witek Radomski, Andrew Cooke, J. Therien, and E. Binet. ERC-1155 Multi Token Standard. https://github.com/ethereum/ EIPs/blob/master/EIPS/eip-1155.md, June 2018. [Online; accessed 12-Jan-2019].
- [23] M. Vladimirov and D. Khovratovich. ERC20 API: An Attack Vector on Approve/TransferFrom Methods. https://docs.google.com/document/d/1YLPtQxZu1UAvO9cZ1O2RPXBbT0mooh4DYKjA_jp-RLM/edit#heading=h.m9fhqynw2xvt, Nov. 2016. [Online; accessed 25-Nov-2018].