

A Formal Verification Tool for Ethereum VM Bytecode

Daejun Park
Yi Zhang
University of Illinois at
Urbana-Champaign
Runtime Verification, Inc.

Manasvi Saxena
Grigore Rosu
University of Illinois at
Urbana-Champaign
Runtime Verification, Inc.

Philip Daian
Cornell Tech, IC3
Runtime Verification, Inc.

ABSTRACT

In this paper, we present a formal verification tool for the Ethereum Virtual Machine (EVM) bytecode. To precisely reason about all possible behaviors of the EVM bytecode, we adopted KEVM, a complete formal semantics of the EVM, and instantiated the K-framework’s reachability logic theorem prover to generate a correct-by-construction deductive verifier for the EVM. We further optimized the verifier by introducing EVM-specific abstractions and lemmas to improve its scalability. Our EVM verifier has been used to verify various high-profile smart contracts including the ERC20 token, Ethereum Casper, and DappHub MakerDAO contracts.

Demo Video URL: <https://youtu.be/4XBcAclq0Vk>

ACM Reference Format:

Daejun Park, Yi Zhang, Manasvi Saxena, Grigore Rosu, and Philip Daian. 2018. A Formal Verification Tool for Ethereum VM Bytecode. In *Proceedings of The 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018)*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

The Ethereum smart contract is a safe-critical system whose failures have caused millions of dollars of lost funds,¹ and it requires rigorous formal methods to ensure correctness and security properties of the contracts.² The smart contract is usually written in a high-level language such as Solidity³ or Vyper⁴, and then it is compiled down to the Ethereum Virtual Machine (EVM) bytecode⁵ that actually runs on the blockchain.

In this paper, we present a formal verification tool for the EVM bytecode. We chose the EVM bytecode as the verification target language so that we can directly verify what is actually executed without the need to trust the correctness of the compiler. To precisely reason about the EVM bytecode without missing any EVM quirks, we adopted KEVM [4], a *complete* formal semantics of the

EVM, and instantiated the K-framework’s reachability logic theorem prover [7] to generate a *correct-by-construction* deductive program verifier for the EVM. While it is sound, the initial out-of-box EVM verifier was relatively slow and failed to prove many correct programs. We further optimized the verifier by introducing custom abstractions and lemmas specific to EVM that expedite proof searching in the underlying theorem prover. We have been using the EVM verifier to verify the *full functional correctness* of high-profile smart contracts including multiple ERC20 token contracts [9], Ethereum’s Casper⁶ contract, and DappHub’s MakerDAO⁷ contract. The verification artifact is publicly available at [8].

Contributions. We describe our primary contributions:

- We present a formal verification tool for the EVM bytecode that is capable and scalable enough to verify various high-profile, safe-critical smart contracts. Moreover, our verifier is the first tool, to the best of our knowledge, that adopts a complete formal semantics of EVM, being able to completely reason about all possible corner-case behaviors of the EVM bytecode. See Section 5 for comparison to other tools.
- We enumerate important, concrete challenges in verifying the EVM bytecode, and propose EVM-specific abstractions and lemmas to mitigate the challenges. (Section 2 & 3)
- We present a case study of completely verifying multiple high-profile ERC20 token contracts. We enumerate divergent behaviors we found across these tokens, illuminating potential security vulnerabilities for any API clients assuming consistent behavior across ERC20 implementations. (Section 4)

2 EVM VERIFICATION CHALLENGES

Verifying the EVM bytecode is challenging, especially due to the internal byte-manipulation operations that require the non-linear integer arithmetic reasoning which is undecidable in general [5]. Here we provide a few examples of the challenges in verifying EVM bytecodes.

Byte-Manipulation Operations. The EVM provides three types of storage structures: a local memory, a local stack, and the global storage. Of these, only the local memory is byte-addressable (i.e., represented as an array of bytes), while the others are word-addressable (i.e., each represented as an array of 32-byte words). Thus, a 32-byte (i.e., 256-bit) word needs to be split into 32 chunks of bytes to be stored in the local memory, and those 32 chunks need to be merged back to be loaded in either the local stack or the global storage. These byte-wise splitting and merging operations can be

¹<https://blog.ethereum.org/2016/06/19/thinking-smart-contract-security/>

²<https://blog.ethereum.org/2016/09/01/formal-methods-roadmap/>

³<http://solidity.readthedocs.io/en/v0.4.24/>

⁴<https://vyper.readthedocs.io/en/latest/index.html>

⁵<http://yellowpaper.io/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ESEC/FSE 2018, 4–9 November, 2018, Lake Buena Vista, Florida, United States

© 2018 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

⁶<https://eips.ethereum.org/EIPS/eip-1011>

⁷<https://makerdao.com/>

formalized using the non-linear integer arithmetic operations, as follows.⁸ Suppose x is a 256-bit integer. Let x_n be the n^{th} byte of x in its two’s complement representation, where the index 0 refers to the LSB, defined as follows:

$$x_n \stackrel{\text{def}}{=} (x/256^n) \bmod 256$$

Let *merge* be a function that takes as input a list of bytes and returns the corresponding integer value under the two’s complement interpretation, recursively defined as:

$$\begin{aligned} \text{merge}(x_i \cdots x_{j+1}x_j) &\stackrel{\text{def}}{=} \text{merge}(x_i \cdots x_{j+1}) * 256 \pm x_j \quad \text{when } i > j \\ \text{merge}(x_i) &\stackrel{\text{def}}{=} x_i \end{aligned}$$

where $*$ and \pm are multiplication and addition over words (modulo 2^{256}). If the byte-wise operations are blindly encoded as SMT theorems, then Z3, a state-of-the-art SMT solver, times out attempting to prove “ $x = \text{merge}(x_{31} \cdots x_0)$ ”. The SMT query can be simplified to allow Z3 to efficiently terminate, for example, by omitting the modulo reduction for multiplication and addition in *merge* with additional reasoning about the soundness of the omission. Despite these improvements, the merge operation still incurs severe performance penalties as solving the large formula is required for every load/store into memory, an extremely common operation.

Arithmetic Overflow. Since EVM arithmetic instructions perform the modular arithmetic (i.e., $+$, $-$, $*$, $/ \bmod 2^{256}$), detecting the arithmetic overflow is critical for preventing potential security holes due to an unexpected overflow. Otherwise, for example, increasing a user’s token balance could trigger an overflow, resulting in loss of the funds as the balance wraps around to a lower-than-expected value. There is no standard EVM-level overflow check, so the overflow detection varies across compilers and libraries. For example, the Vyper compiler inserts the following runtime check for an addition $a + b$ over the 256-bit unsigned integers a and b :

$$b == 0 \mid \mid a \pm b > a$$

where \pm represents addition modulo 2^{256} . It seems straightforward to show that the above formula is equivalent to $a + b < 2^{256}$ (where $+$ is the pure addition *without* modulo reduction), but it is no longer trivial once the above is compiled down to EVM. The compiled EVM bytecode of the above conditional expression can be encoded in the SMT-LIB format as follows:

```
(not (= (chop (+ (bool2int (= b 0))
                (bool2int (> (chop (+ a b)) a)))) 0))
```

where `(chop x)` denotes $(x \bmod 2^{256})$, and `(bool2int x)` is defined by `(ite x 1 0)`. However, Z3 fails (timeout) to prove that the above SMT formula is equivalent to $a + b < 2^{256}$.

Gas Limit. Reasoning about gas consumption is important since an exception will be thrown when the gas runs out of during the execution, and such an exception could result in tokens being frozen and getting stuck,⁹ which is as dangerous as token being stolen or vaporized. Analyzing the precise gas consumption is not trivial

⁸It is also possible to formalize the byte-manipulation using the bit-vector theory, but the formalization using the mathematical integer theory has an advantage of the functional specifications being succinct. Indeed, the KEVM semantics adopted the integer formalization because of the advantage.

⁹https://www.reddit.com/r/ethereum/comments/4ghzhv/governmentals_1100_eth_jackpot_payout_is_stuck/

because it depends on both the program and its input state — the parameters and the storage. For example, *SSTORE*, an instruction that stores a given value in a given location of the storage, consumes an amount of the gas depending on both the new value to be stored and the existing value to be overwritten — a larger amount of gas is charged for storing a value for the first time than updating the existing value, which incentivizes recycling the storage entries that are not used any longer.

Hash Collision. Precise reasoning about the SHA256 hash is critical. Since it is not practical to consider the hash algorithm details every time the hash function is called in EVM bytecodes, an abstraction for the hash function is required. Designing a sound but efficient abstraction is not trivial because while the SHA256 hash is not cryptographically collision-free, the contract developers assume collisions will not occur during normal execution of their contracts.¹⁰ A naive way of capturing the assumption would be to simply abstract the SHA256 hash as an injective function. However, it is not sound simply because of the pigeonhole principle, and thus we need to be careful when abstracting the hash function.

3 EVM-SPECIFIC ABSTRACTIONS

K’s reachability logic theorem prover can be seen as a symbolic model checker equipped with coinductive reasoning about loops and recursions (for more details of the underlying theory and implementation, refer to [7]). The prover, in its current form, often delegates domain reasoning to SMT solvers. The performance of the underlying SMT solvers is critical for the overall performance. The domain reasoning involved in the EVM bytecode verification is not tractable in many cases, especially due to the non-linear integer arithmetic. We had to design custom abstractions and lemmas to avoid the non-tractable domain reasoning and improve the scalability.

Abstraction for Local Memory. We present an abstraction for the EVM local memory to allow the word-level reasoning. As mentioned in Section 2, since the local memory is byte-addressable, the load and store operations involve the conversion between a word and a list of bytes, which is not tractable to reason about in general. Our abstraction helps to make the reasoning easier by abstracting away the byte-manipulation details of the conversion. Specifically, we introduce uninterpreted function abstractions and lemmas for the word-level reasoning as follows.

The term $\text{nthByteOf}(v, i, n)$ represents the i^{th} byte of the two’s complement representation of v in n bytes (0 being MSB), with discarding high-order bytes when v is not fit in n bytes. Precisely it is defined as follows:

$$\begin{aligned} \text{nthByteOf}(v, i, n) &= \text{nthByteOf}(\lfloor v/256 \rfloor, i, n-1) \quad \text{when } n > i+1 \\ \text{nthByteOf}(v, i, n) &= v \bmod 256 \quad \text{when } n = i+1 \end{aligned}$$

However, we want to keep it uninterpreted (i.e., do not unfold the definition) when the arguments are symbolic, to avoid the non-linear arithmetic reasoning.

We introduce lemmas over the uninterpreted functional terms. The following lemmas are used for symbolic reasoning about *MLOAD*

¹⁰The assumption is not unreasonable, as virtually all blockchains rely heavily on the collision-resistance of hash functions.

and MSTORE instructions. They capture the essential mechanisms used by the two instructions: splitting a word into a list of bytes and merging it back into the word. First, we have the bound of $\text{nthByteOf}(v, i, n)$ by the definition: $0 \leq \text{nthByteOf}(v, i, n) < 256$. Then we have the following lemma for the merging operation over the terms:

$$\text{merge}(\text{nthByteOf}(v, 0, n) \cdots \text{nthByteOf}(v, n-1, n)) = v$$

$$\text{if } 0 \leq v < 2^{8n} \text{ and } 1 \leq n \leq 32$$

Refer to [8] for the other lemmas of the memory abstraction.

Abstraction for Hash. We do not model the hash function as an injective function simply because it is not true due to the pigeonhole principle. Instead, we abstract it as an uninterpreted function, hash, that takes as input a list of bytes and returns an (unsigned) integer:

$$\text{hash} : \{0, \dots, 255\}^* \rightarrow \mathbb{N}$$

Note that this abstraction captures the possibility of the hash collision, even if the probability is very small.

However, one can avoid reasoning about the potential collision by assuming all of the hashed values appearing in each execution trace are collision-free. This can be achieved by instantiating the injectivity property only for the terms appearing in the symbolic execution, in a way analogous to the universal quantifier instantiation.

Arithmetic Simplification Rules. We introduce simplification rules, specific to EVM, that capture arithmetic properties, which reduce a given term into a smaller one. These rules help to improve the performance of the underlying theorem prover’s symbolic reasoning. For example, we have the following simplification rule:

$$(x \times y) / y = x \quad \text{if } y \neq 0$$

where $/$ is the integer division.¹¹ We also have a rule for the masking operation, $0\text{xff} \cdots f \& n$, as follows:

$$m \& n = n \quad \text{if } m + 1 = 2^{1+\log m} \text{ and } 0 \leq n \leq m$$

where $\&$ is the bitwise AND operator, and m denotes a bitmask $0\text{xff} \cdots f$. Refer to [8] for the other simplification rules.

4 CASE STUDY: ERC20 VERIFICATION

In this section, we present a case study of completely verifying high-profile, practically deployed implementations of the ERC20 token contract [9], one of the most popular Ethereum smart contracts that provides the essential functionality of maintaining and exchanging tokens.

4.1 Formal Specification

The ERC20 standard [9] informally specifies the correctness properties that ERC20 token contracts must satisfy. Unfortunately, however, it leaves several corner cases unspecified, which makes it less than ideal to use in the formal verification of token implementations.

We specified ERC20-K, a complete formal specification of the high-level business logic of the ERC20 standard in the K framework.

¹¹Note that Z3 fails to prove this seemingly trivial formula at the time of this writing. Indeed, this issue has been fixed in the develop branch, once reported by the author of this paper: <https://github.com/Z3Prover/z3/issues/1683>

```
[transfer-success]
callData: #abiCallData("transfer", #address(TO), #uint256(VALUE))
statusCode: _ => EVMC_SUCCESS
output: _ => #asByteStackInWidth(1, 32)
log: ... (. => #abiEventLog(FROM, "Transfer",
    #indexed(#address(FROM)), #indexed(#address(TO)), #uint256(VALUE)))
storage:
    #hashedLocation({BALANCES}, FROM) |-> (BAL_FROM => BAL_FROM -Int VALUE)
    #hashedLocation({BALANCES}, TO) |-> (BAL_TO => BAL_TO +Int VALUE)
    ...
requires:
    andBool FROM !=Int TO
    andBool VALUE <=Int BAL_FROM
    andBool BAL_TO +Int VALUE <Int (2 ^Int 256)

[transfer-failure]
callData: #abiCallData("transfer", #address(TO), #uint256(VALUE))
statusCode: _ => EVMC_REVERT
output: _ => _
log: ...
storage:
    #hashedLocation({BALANCES}, FROM) |-> BAL_FROM
    #hashedLocation({BALANCES}, TO) |-> BAL_TO
    ...
requires:
    andBool FROM !=Int TO
    andBool ( VALUE >Int BAL_FROM
    orBool BAL_TO +Int VALUE >=Int (2 ^Int 256) )
```

Figure 1: Formal specification of transfer function

ERC20-K clarifies what data (e.g., balances and allowances) are handled by the various ERC20 functions and the precise meaning of those functions on such data. ERC20-K also clarifies the meaning of all the corner cases that the ERC20 standard omits to discuss, such as transfers to itself or transfers that result in arithmetic overflows, following the most natural implementations that aim at minimizing gas consumption. The complete specifications are publicly available at [6].

Figure 1, for example, shows part of the (simplified) specification of the transfer function. It specifies two possible behaviors: success and failure.¹² For each case, it specifies the function parameters (callData), the return value (output), whether an exception occurred (statusCode), the log generated (log), the storage update (storage), and the path-condition (requires). Specifically, the success case specification (denoted by [transfer-success]) specifies that the function succeeds in transferring the VALUE tokens from the FROM account to the TO account, with generating the corresponding log message, and returns 1 (i.e., true), if no overflow occurs (i.e., the FROM account has a sufficient balance, and the TO account has enough room to receive the tokens). The failure case (denoted by [transfer-failure]) specifies that the function throws an exception without modifying the account balances, if an overflow occurs.

4.2 Formal Verification

For the case study, we consider four implementations of the ERC20 token contract: the Vyper ERC20 token¹³, the HackerGold (HKG)

¹²Indeed, transfer admits four possible behaviors: success and failure of regular transfer (i.e., FROM \neq TO), and success and failure of self-transfer (i.e., FROM = TO). Here we omit the self-transfer behaviors due to the space limit. Refer to [6] for the complete specification.

¹³https://github.com/ethereum/vyper/blob/master/examples/tokens/ERC20_solidity_compatible/ERC20.vy

	Vyper	HKG	Zeppelin		Vyper	HKG	Zeppelin
totalSupply	36.4	N/A	34.3	approve	33.9	48.4	35.4
balanceOf	33.3	37.3	37.1	transfer	148.5	198.5	219.7
allowance	36.4	42.3	39.6	transferFrom	174.4	257.6	179.2

Table 1: Verification time (secs) of ERC20 token contracts

ERC20 token¹⁴, and OpenZeppelin’s ERC20 token¹⁵. Of these, the Vyper ERC20 token is written in Vyper, and the others are written in Solidity.

We compiled the source code down to the EVM bytecode using each language compiler, and executed our verifier to verify that the compiled EVM bytecode satisfies the aforementioned specification. During this verification process, we found divergent behaviors across these contracts that do not conform to the ERC20 standard. Due to the deviation from the standard, we could not verify those contracts against the original ERC20-K specification. In order to show that they are “correct” w.r.t. the original specification *modulo* the deviation, we modified the specification to capture the deviation and successfully verified them against the modified specification. Table 1 provides the performance of the verifier. We describe the verification result as follows.

Vyper ERC20 Token. The Vyper ERC20 token is successfully verified against the original specification, implying its full conformance to the ERC20 standard.

HackerGold (HKG) ERC20 Token. In addition to the well-known security vulnerability of the HKG token,¹⁶ we found that the HKG token implementation deviates from our specification as follows:

- **No totalSupply function:** No totalSupply function is provided in the HKG token, which is *not* compliant to the ERC20 standard.
- **Returning false in failure:** It returns a false instead of throwing an exception in the failure cases for both transfer and transferFrom. It does not violate the standard, as throwing an exception is recommended but not mandatory according to the ERC20 standard.
- **Rejecting transfers of 0 values:** It does not allow transferring 0 values, returning false immediately without logging any event. It is *not* compliant to the standard. This is a potential security vulnerability for any API clients assuming the ERC20-compliant behavior.
- **No overflow protection:** It does not check arithmetic overflow, resulting in the receiver’s balance wrapping around the 256-bit unsigned integer’s maximum value in case of the overflow. It does not violate the standard, as the standard does not specify any requirement regarding it. However, it is potentially vulnerable, since it will result in loss of the funds in case of the overflow as the receiver’s balance wraps around to a lower-than-expected value.

¹⁴<https://github.com/ether-camp/virtual-accelerator/blob/master/contracts/StandardToken.sol>

¹⁵<https://github.com/OpenZeppelin/openzeppelin-solidity/blob/master/contracts/token/ERC20/StandardToken.sol>

¹⁶<https://www.ethnews.com/ethercamps-hkg-token-has-a-bug-and-needs-to-be-reissued>
Note that the token contract had been manually audited by Zeppelin, but they failed to find the vulnerability, which implies the need of the rigorous formal verification.

OpenZeppelin ERC20 Token. The OpenZeppelin ERC20 token is a high-profile ERC20 token library developed by the security audit consulting firm Zeppelin¹⁷. We found that the OpenZeppelin token deviates from the ERC20-K specification as follows:

- **Rejecting transfers to address 0:** It does not allow transferring to address 0, throwing an exception immediately. It does not violate the standard, as the standard does not specify any requirement regarding it. However, it is questionable since while there are many other invalid addresses to which a transfer should not be made, it is not clear how useful rejecting the single invalid address is, at the cost of the additional gas consumption for every transfer transaction to check if the given address is zero.

5 RELATED WORK

While there exist several static analysis tools tailored to check certain predefined properties, here we consider, due to the space limit, only the verification tools backed by a full-fledged theorem prover that allows to reason about arbitrary (full functional correctness) properties. Specifically, Bhargavan et al. [2] and Grishchenko et al. [3] presented a verification tool based on the F* proof assistant, and Amani et al. [1] presented a tool based on Isabelle/HOL. These tools, however, adopt only a partial, incomplete semantics of EVM, and thus may miss certain critical corner-case behaviors of the EVM bytecode, which could undermine the soundness of the verifiers. Our EVM verifier, on the other hand, is a verification tool derived from a *complete* formal semantics of EVM, for the first time to the best of our knowledge.

REFERENCES

- [1] Sidney Amani, Myriam Bégel, Maksym Bortin, and Mark Staples. 2018. Towards Verifying Ethereum Smart Contract Bytecode in Isabelle/HOL. In *Proceedings of the 7th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2018)*. ACM, New York, NY, USA, 66–77. <https://doi.org/10.1145/3167084>
- [2] Karthikeyan Bhargavan, Antoine Delignat-Lavaud, Cédric Fournet, Anitha Gollamudi, Georges Gonthier, Nadim Kobeissi, Natalia Kulatova, Aseem Rastogi, Thomas Sibut-Pinote, Nikhil Swamy, and Santiago Zanella-Béguelin. 2016. Formal Verification of Smart Contracts: Short Paper. In *Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security (PLAS ’16)*. ACM, New York, NY, USA, 91–96. <https://doi.org/10.1145/2993600.2993611>
- [3] Ilya Grishchenko, Matteo Maffei, and Clara Schneidewind. 2018. A Semantic Framework for the Security Analysis of Ethereum Smart Contracts. In *Principles of Security and Trust - 7th International Conference, POST 2018, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2018, Thessaloniki, Greece, April 14–20, 2018, Proceedings*. 243–269. https://doi.org/10.1007/978-3-319-89722-6_10
- [4] Everett Hildenbrandt, Manasvi Saxena, Xiaoran Zhu, Nishant Rodrigues, Philip Daian, Dwight Guth, Brandon Moore, Yi Zhang, Daejun Park, Andrei Ștefănescu, and Grigore Roșu. 2018. KEVM: A Complete Semantics of the Ethereum Virtual Machine. In *Computer Security Foundations Symposium (CSF 2018)*.
- [5] Yuri V. Matiyasevich. 1993. *Hilbert’s Tenth Problem*. MIT Press, Cambridge, MA, USA.
- [6] Grigore Rosu. 2017. ERC20-K: Formal Executable Specification of ERC20. <https://github.com/runtimeverification/erc20-semantics>.
- [7] Andrei Ștefănescu, Daejun Park, Shijiao Yuwen, Yilong Li, and Grigore Roșu. 2016. Semantics-based Program Verifiers for All Languages. In *Proceedings of the 2016 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 2016)*. ACM, New York, NY, USA, 74–91. <https://doi.org/10.1145/2983990.2984027>
- [8] Formal Verification Team. 2018. Verified Smart Contracts. <https://github.com/runtimeverification/verified-smart-contracts/>.
- [9] Fabian Vogelsteller and Vitalik Buterin. 2015. ERC-20 Token Standard. <https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md>.

¹⁷<https://zeppelin.solutions/security-audits>