SAFEVM: A Safety Verifier for Ethereum Smart Contracts

Elvira Albert, Jesús Correas, Pablo Gordillo
Complutense University of Madrid
Spain
Guillermo Román-Diez
Universidad Politécnica de Madrid
Spain
Albert Rubio
Complutense University of Madrid
Spain

ABSTRACT

Ethereum smart contracts are public, immutable and distributed and, as such, they are prone to vulnerabilities sourcing from programming mistakes of developers. This paper presents SAFEVM, a verification tool for Ethereum smart contracts that makes use of state-of-the-art verification engines for C programs. SAFEVM takes as input an Ethereum smart contract (provided either in Solidity source code, or in compiled EVM bytecode), optionally with assert and require verification annotations, and produces in the output a report with the verification results. Besides general safety annotations, SAFEVM handles the verification of array accesses: it automatically generates SV-COMP verification assertions such that C verification engines can prove safety of array accesses. Our experimental evaluation has been undertaken on all contracts pulled from etherscan.io (more than 24,000) by using as back-end verifiers CPAchecker, SeaHorn and VeryMax.

CCS CONCEPTS

• Theory of computation → Program analysis; • Software and its engineering → Software verification and validation.

KEYWORDS

Smart contracts, Ethereum blockchain, Safety verification.

ACM Reference Format:

1 OVERVIEW OF SAFEVM

Each blockchain provides its own programming language to implement smart contracts. Solidity, a Turing complete language, is the most popular language to write smart contracts for the Ethereum platform that are then compiled to EVM (Ethereum Virtual Machine [22]) bytecode. Each instruction executed by the EVM has an associated gas consumption specified by Ethereum. Being security a main concern of Ethereum, the Solidity language contains the verification-oriented functions, assert and require, to check for safety conditions or requirements and terminate the execution if they are not met. As usual, the assert function can be used for verification purposes (e.g., to check invariants), while the require function is used to specify preconditions (e.g., to ensure valid conditions on the inputs or contract state variables, or to validate return values from calls to external contracts). When the Solidity code is compiled into EVM bytecode, the require condition is transformed into a test that checks the condition and invokes a REVERT bytecode if it does not hold. REVERT aborts the whole execution of the smart contract, reverts the state and all remaining gas is refunded to the caller. The assert checks the condition and invokes an INVALID bytecode if it does not hold. When executing INVALID, the state is reverted but no gas is refunded, and hence it has more serious consequences than REVERT: besides the economic consequences of losing the gas, the only information given to the transaction is an out-of-gas error message. The treatment of array accesses is done similarly as for the assert, when an array position is accessed, the generated EVM bytecode checks if the position accessed is within the array bounds and otherwise the INVALID bytecode is executed. Division and related bytecodes like MOD, SMOD, ADDMOD, MULMOD, also lead to executing INVALID when the denominator is zero.

Therefore, the INVALID bytecodes are key for the verification of the Ethereum smart contracts, as they capture both assertion violations and several sources of fatal operations (e.g., out-of-bounds access, division by zero). In essence, our approach to the verification of smart contracts consists in decompiling the EVM bytecode for the smart contract into a C program with ERROR annotations (following the SV-COMP format, https://sv-comp.sosy-lab.org/2019/rules.php) to enable their verification using existing tools for the verification of C programs. Developing the verifier from the low-level EVM has
important advantages: (i) sometimes the source code is not available (e.g., the blockchain only stores the bytecode), (ii) the INVALID bytecodes are visible at the level of bytecode and we can give a uniform treatment to the various safety concerns described above, (iii) our analysis works for any other language that compiles to EVM (e.g., Vyper), and it is not affected by changes in the source language, or by compiler optimizations. Luckily, there are a number of open-source tools that help us in the decompilation process and that we have integrated within our tool-chain.

Fig. 1 depicts the main components of SAFEVM that are as follows (shaded boxes are off-the-shelf used systems not developed by us): (1) Input. SAFEVM takes a smart contract, optionally with assert and require verification annotations. The smart contract can be given in Solidity source code or in EVM compiled code. In the latter case, the annotations have been compiled into bytecode as described above. (2) CFG. In either form, the code is given to Oyente [2], a symbolic execution engine that has been extended to compute the complete CFG from the given smart contract. As Oyente does not handle recursive functions, they are already discarded at this step. The CFG generation phase is not described in the paper, we refer to [2, 3]. (3) EthIR. The decompilation of the EVM bytecode into a higher-level rule-based representation (RBR) is carried out from the generated CFG by EthIR [3]. Technical details of this phase are not described in the paper, we refer to [3]. (4) C+SV-COMP translator. We have implemented a translator for the recursive RBR representation into an abstract Integer C program (i.e., all data is of type Integer) with verification annotations using the SV-COMP format. Features of the EVM that we cannot handle yet (e.g., bit-wise operations) are abstracted away in the translation (see Sec. 2). INVALID instructions are transformed into ERROR annotations in the C program following the SV-COMP format. (5) Verification. Any verification tool for Integer C programs that uses SV-COMP annotations can be used to verify the safety of our C-translated contracts. We have evaluated our approach using three state-of-the-art C verifiers, CPAchecker [6], VeryMax [9], and SeaHorn [14], and the verification report they produce is processed by us to report the results in terms of functions of the smart contract.

Our tool SAFEVM has a very large (potential) user base, as Ethereum is currently the most advanced platform for coding and processing smart contracts. As we will describe in Sec. 3, using SAFEVM we have automatically verified safety of around 20% of all functions (depending on the verifier) that might execute INVALID bytecodes from the whole set of contracts pulled from etherscan.io (more than 24,000 contracts), and we have found potential vulnerabilities in functions that could not be verified.

2 TRANSFORMATION TO C WITH SV-COMP ANNOTATIONS

As motivating example, we use a Solidity contract that implements a lottery system called SmartBillions (available at https://smartbillions.com/). We illustrate the safety verification of its internal function commitDividend (an excerpt of its code appears to the left of Fig. 2) that commits remaining dividends to the user wh. We have shortened the variable names by removing the vowels from the names. Lines marked with ⋄ might lead to executing different sources of INVALID: Line 16 (L16 for short) to a division by zero when t11sapply is 0; at L19 when 1st  \geq  dvdnd.length and thus it is accessing a position out of the bounds of the array; and at L21 when the condition within the assert does not hold. In order to be able to verify its safety (i.e., absence of INVALID executions), we add the lines marked with ⋄ that introduce error-handling functions require and assert in the verification process.

The starting point of our translator is the RBR produced by EthIR [3]. The RBR is composed of a set of rules containing decompiled versions of bytecode instructions (e.g., LOAD and STORE are decompiled into assignments) and whose structure of rule invocations is obtained from the CFG produced by Oyente. The RBR might contain two kinds of rules: sequences of instructions referred to as blockX, and conditional jump rules, named jumpX, whose first instruction is the Boolean condition used to select between the rules of the function definition. Rule parameters include: the operand stack flattened in variables named s\(_j\), the state of the contract (this is the global data), named g\(_i\), and the local memory (represented by local variables), named l\(_i\). To the right of Fig. 2 we show the fragment of the RBR produced by EthIR for the loop of L18-L20. At rule block759 we show the transformation of some EVM bytecodes (the original bytecodes appear in comments //) into higher-level RBR instructions. The RBR is already abstract in the sense that when variables refer to state or memory locations that are not known they become fresh variables (see variable fresh0 in block759) so that a posterior analysis will not assume any value for them (details are in [3]). Observe that the fragment of the RBR contains an INVALID instruction within block758 and such block can be executed when \(g_{eq}(s_9, s_8)\) (see rule jump745). By tracking variable assignments, we can infer that s\(_9\) contains the value of 1st and s\(_8\) the size of dvdnd, hence the comparison is checking out-of-bounds array access. The remaining of the section explains the main four phases of the translation from the RBR to an abstract Integer C program.

(1) C functions: Our translation produces, for each non-recursive rule definition in the RBR, a C function without parameters that returns void. Recursive rules produced by loops are translated into iterative code. For this part of the translation, we compute the SCC
(4) **SV-COMP annotations**: The verification of Ethereum smart contracts is done in SAFEVM by guaranteeing the unreachability of the INVALID operations in the C-translated code. Following the SV-COMP rules, we translate INVALID operations into calls to the \_VERIFIER\_error function so that its unreachability can be proven by any verification tool compatible with the SV-COMP annotations. An example of an INVALID operation can be seen in L38. Verification tools return that the program in Fig. 2 cannot be verified as the INVALID instruction could be executed. This is due to the fact that contract state values are unknown, that is: t1 Smap is not guaranteed to be different from 0 at L16 and the size of the array dvxds is not guaranteed to be greater than the value of 1st at L19. Lines L11 and L12 contain the Solidity instructions needed to guarantee that L16 and L19, respectively, will never execute an INVALID instruction. The assert at L21 can be verified by using the require at L14. The inclusion of the require annotation also improves the contract as, if it is violated, a REVERT rather than an INVALID bytecode will be executed, not causing a loss of gas of the transaction (while the gas needed to check it is negligible).

### 3 EXPERIMENTAL EVALUATION

All components of SAFEVM, except for the C verifiers, are implemented in Python and are open-source. SAFEVM accepts smart contracts written in versions of Solidity up to 0.4.25 and bytecode for the Ethereum Virtual Machine v1.8.18. This section reports the results of our experimental evaluation using SAFEVM with CPAchecker, SeaHorn and VeryMax as verification back-ends. An artifact to try our tool can be downloaded from [http://costa.fdi.ucm.es/papers/costa/safevm.ova](http://costa.fdi.ucm.es/papers/costa/safevm.ova).

In order to experimentally evaluate SAFEVM, we pulled from etherscan.io all Ethereum contracts whose source code was available on January 2018. This ended up in 10,796 files. From those, we have searched for those files that contain EVM code with INVALID instructions, in total 7,323. The first phase of SAFEVM that performs the decompilation into the RBR fails for 1,000 files (this 13.65% is larger but quite aligned with the failing rates of other tools e.g. [1, 8]) and reaches a timeout of 60s for 22 files. Thus, our results are on the remaining 6,301 files, that contain 24,294 contracts with 44,046 public functions that can reach an INVALID instruction and 177,549 INVALID-free functions. We have tested both the translation to type int and unsigned int for defining C variables, as mentioned in Sec. 2 for those 44,046 functions. We get the following results by using 60s of timeout (Error denotes an error output by the verifier):

<table>
<thead>
<tr>
<th>Results</th>
<th>CPAchecker</th>
<th>VeryMax</th>
<th>SeaHorn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verified</td>
<td>19.48%</td>
<td>19.13%</td>
<td>20.32%</td>
</tr>
<tr>
<td>Non-Verified</td>
<td>77.04%</td>
<td>79.82%</td>
<td>73.32%</td>
</tr>
<tr>
<td>Timeout</td>
<td>5.21%</td>
<td>0.82%</td>
<td>6.29%</td>
</tr>
<tr>
<td>Error</td>
<td>0.27%</td>
<td>0.23%</td>
<td>0.07%</td>
</tr>
</tbody>
</table>

The results for all verifiers are quite aligned, although VeryMax verifies a slightly lower number of functions, and SeaHorn verifies more functions and less reach a timeout. The interpretation made by the tools regarding the Integer semantics (bounded or unbounded) leads to the only relevant difference in the number of functions verified between both translations. We have manually inspected, out of the 7,323 files, those files whose addresses start with 0x00 and 0x01 in order to understand...
the cases that could not be verified. This is a sample of 29 files (243 public functions) that are available at https://github.com/costa-group/EthIR/tree/master/examples/safevm. The manual inspection on the subset gives 54 false alarms (22.2%), namely: 49 functions were verified by CPAChecker; 140 are correct alarms, most of them produced by asserts introduced by the programmers for safety to abort the execution (e.g. 83 come from SafeMath); 54 are false alarms (many related to enum accesses and other imprecisions in the decompilation phase). More in detail, we have identified four types of situations: (1) false alarms due to inaccuracy of our tool: some assert statements contain non-integer types (e.g., strings, enum, etc.) which cannot be verified as we need a more accurate decompilation (see Sec. 4); (2) correct alarms that require conditional verification: some assert statements can only be verified for concrete contexts, e.g., we found asserts to prevent from under/overflow integer arithmetic operations in a widely used library SafeMath that can only be verified for given inputs. In the future we plan to integrate conditional verification [9] to infer the preconditions for the asserts to hold; (3) Correct alarms detecting potential vulnerabilities: we have detected several INVALID operations that could represent a vulnerability in the code (e.g., functions that access an array element without checking the boundary) and we have protected them adding require statements that enable subsequent verification; and (4) four functions whose verification results depend on the different semantics used for Integers.

As final observations, we notice that assert is overused (contradicting the best practices recommendations of Solidity) and that some contracts can be improved by using require to avoid the loss of gas when the assert statement does not hold. Finally, we argue that although there is much room for improving the accuracy, the results of our experimental evaluation are very encouraging: we have verified safety w.r.t. INVALID bycodes for around 20% of the functions that might reach INVALID fully automatically by using state-of-the-art verifiers.

4 CONCLUSIONS

Verification of Ethereum smart contracts for potential safety and security vulnerabilities is becoming a popular research topic with numerous tools being developed, among them, we have tools based on symbolic execution [13, 15, 17, 18, 20, 21], tools based on SMT solving [16, 19], and tools based on certified programming [5, 7, 12]. There are some tools also that aim at detecting, analyzing and verifying non-functional properties of smart contracts, e.g., those focused on reasoning about the gas consumption [4, 10, 11, 19].

To the best of our knowledge, SAFEMVM is the first tool that uses existing verification engines developed for C programs to verify low-level EVM code. This opens the door to the applicability of advanced techniques developed for the verification of C programs to the new languages used to code smart contracts. Although our tool is still in a prototypical stage, it provides a proof-of-concept of the transformational approach, and we argue that it constitutes a promising basis to build verification tools for EVM smart contracts. Some of the aspects that we aim at improving in future work is the handling of the data stored in the memory, as it is abstracted away by the EthIR component that SAFEMVM is using as soon as there are storage operations on memory. Developing a memory analysis for EVM smart contracts can be crucial for the accuracy of verification. We also aim at handling bit-wise operations in the future that are extensively used in the EVM bytecode. Advanced reasoning for arrays and maps (the only data structures available in Ethereum smart contracts) can be also added to the framework to gain further accuracy. This requires also further work on the decompilation side. Along the same line, learning information on the types of variables during decompilation will have an impact in the accuracy of the verification process.

ACKNOWLEDGMENTS

This work was funded partially by the Spanish MINECO project TIN2015-69175-C4-2-R and MINECO/FEDER, UE project TIN2015-69175-C4-3-R, by Spanish MICINN/FEDER, UE projects RTI2018-094403-B-C31 and RTI2018-094403-B-C33, by the CM projects S2018/TCS-4314 and S2018/TCS-4339, co-funded by EIE Funds of the European Union, and by the UCM CT27/16-CT28/16 grant.

REFERENCES