Towards Verifying Ethereum Smart Contract Bytecode in Isabelle/HOL

Sidney Amani
Data61 (CSIRO)
NSW, Australia
Sidney.Amani@data61.csiro.au

Myriam Bégel*
ENS Paris-Saclay, Universitée Paris-Saclay
France
Myriam.Begel@ens-paris-saclay.fr

Maksym Bortin
Data61 (CSIRO)
NSW, Australia
Maksym.Bortin@data61.csiro.au

Mark Staples
Data61 (CSIRO) & School of CSE, UNSW
NSW, Australia
Mark.Staples@data61.csiro.au

Abstract

Blockchain technology has increasing attention in research and across many industries. The Ethereum blockchain offers *smart contracts*, which are small programs defined, executed, and recorded as transactions in the blockchain transaction history. These smart contracts run on the Ethereum Virtual Machine (EVM) and can be used to encode agreements, transfer assets, and enforce integrity conditions in relationships between parties. Smart contracts can carry financial value, and are increasingly used for safety-, security-, or mission-critical purposes. Errors in smart contracts have led and will lead to loss or harm. Formal verification can provide the highest level of confidence about the correct behaviour of smart contracts. In this paper we extend an existing EVM formalisation [7] in Isabelle/HOL by a sound program logic at the level of bytecode. We structure bytecode sequences into blocks of straight-line code and create a program logic to reason about these. This abstraction is a step towards control of the cost and complexity of formal verification of EVM smart contracts.

Keywords formal verification, blockchain, smart contracts, Ethereum, Isabelle/HOL

1 Introduction

Blockchain technology emerged to support financial transactions in the Bitcoin system, but has become increasingly important in many industries, with potential use in legal, medical, and supply chain industries. The Ethereum blockchain provides a general-purpose computational mechanism called *smart contracts*. These are programs that run on the Ethereum Virtual Machine (EVM) [18]. They and their effects are recorded in the blockchain history. They can be used to encode or execute agreements between parties. For example, a party invoking a smart contract could cause digital currency to be transferred to another party, or could record a state change which makes the other party eligible to invoke other transactions. Smart contracts provide new ways to implement multi-party relationships within blockchain-based applications. In addition to the direct financial value in these applications, blockchains are increasingly being used for safety-critical applications such as in pharmaceutical supply chains, or for mission-critical application such as in electrical power grids. For such reasons, it is highly desirable to know that smart contract implementations do not violate critical requirements, and formal modelling and verification can be applied in this context. To address these questions we seek:

(i) a trustworthy logical framework capable of expressing complex safety and security requirements;
(ii) a valid formal model of the EVM within the framework;
(iii) a sound program logic defined within the framework and able to reason about properties of smart contracts.

In our setting, we use the logical framework Isabelle/HOL, and an existing EVM formal model [7]. Thus, our remaining goal is a sound program logic. In this paper we propose such a logic for EVM bytecode. We target unstructured bytecode rather than a high-level programming language for the following reasons. First, our approach is independent of any high-level language (e.g. Solidity [5]) compiler, making our work more general and significantly less reliant on the correctness of higher-level tools. Second, bytecode is the actual programming language of Ethereum as all smart contracts appear only in this form on the blockchain. Altogether, this gives us the motivation to focus on reasoning about EVM bytecode, despite the absence of convenient programming constructs like conditionals, which we can take for granted in structured languages.

The main contributions of the paper are:

*work done while at Data61 (CSIRO)

PL'17, January 01–03, 2017, New York, NY, USA
2017.

(i) an extension to the EVM formalisation [7] in the Isabelle/HOL theorem prover, covering smart contract correctness properties, and which gives a separate universal treatment of termination based on Ethereum's concept of execution 'gas';
(ii) a sound program logic to verify smart contracts at the bytecode level; and
(iii) Isabelle tactics to support automated generation of verification conditions using the rules of the logic.

Our development is entirely formalised in Isabelle/HOL and available online[1]. The termination proof of EVM bytecode and some of our proof automation tactics have been accepted in the official EVM semantics repository[2] maintained by the Ethereum foundation.

The paper is structured as follows. Section 2 describes the background for the presented work. Section 3 describes how we can capture correctness properties in a pre/postcondition style for EVM bytecode programs. Section 4 is devoted to our program logic, and Section 5 shows the soundness of the logic w.r.t. the correctness property. Section 6 presents a case study, which outlines the specification and verification of bytecode output of the Solidity compiler, as well as how we automate generation of verification conditions using Isabelle tactics. Finally, Section 7 outlines some related work and Section 8 summarises the results and gives an outlook.

2 Background

The EVM is described in the Ethereum 'Yellow Paper' [18], which provides a clear foundation not only for its implementation, but also for its formalisation in logic. One such formalisation has been done [7] using the 'meta-tool' *Lem* [11]. Lem supports a variety of theorem provers including Isabelle/HOL. *Isabelle* [13] is a logical framework in form of a generic interactive theorem prover, whereas Isabelle/HOL encodes higher-order logic and is the most important and most developed part of the framework. Based on a small (meta)-logical inference kernel, Isabelle's LCF-style architecture ensures very high confidence about its soundness as a theorem prover.

However, the EVM model [7] needs to be validated to provide confidence that it meets the specification [18]. To this end, a validation test suite accompanies the model in Lem. Using this, the actual EVM, regarded as the reference implementation of [18], and the OCaml code generated by Lem are both applied to a large collection of contracts, cross-checking their outputs.

As we entirely focus on the Lem output in Isabelle/HOL, we wanted to have an additional validation of this particular EVM formalisation. To this end, we also invoked Isabelle's code generator and run the test suite on the OCaml code generated by Isabelle.

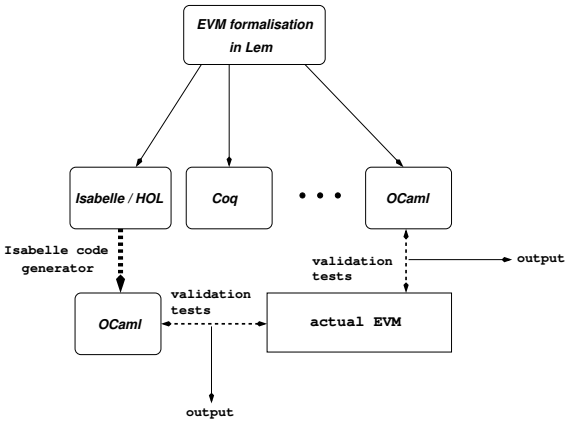

Figure 1. Validation of EVM models

Our 'double-validation' process is outlined in Figure 1. The use of the test suite from Isabelle has required certain efforts, mainly because of different representations of machine words. OCaml code from Lem uses efficient native modules, whereas the Isabelle side invokes a formally verified theory of machine words. Because of this, our suite needs much more time to pass the tests. Nonetheless, we gain a complementary indication that all three models of EVM follow the specification and behave equally.

3 Total Correctness of EVM Bytecode Programs

In his PhD thesis [12], Myreen introduced a general method of formal verification of machine code with a particular application to ARM. In this section we show how this general method can be adapted to EVM, with additional consideration given to EVM specific properties rooted in gas consumption.

In the general model, a state carries all the information needed to execute a program, including instructions (with respective reference numbers) constituting the program itself, a program counter that refers to the current instruction, a stack and so on. All these elements are treated uniformly as sets of so-called *state elements* and separated in a state using separation logic conjunctions $\wedge^*$ (denoted by $*$ in [16]). A single machine step is captured by a function *next* that takes the current instruction via program counter from the state and transforms the state in accordance with the instruction's specified behaviour. Of course, *next* might not always be able to pick an instruction as an execution can have terminated properly or with an exception. This is indicated within a state by the *not-continuing* flag, which is just an abbreviation for the state element *ContinuingElm False*, as opposed

[1] https://github.com/seed/eth-isabelle/tree/dispatcher_ex_mbegel
[2] https://github.com/pirapira/eth-isabelle/

to *continuing* flag abbreviating *ContinuingElm True*. If *not-continuing* is present in a state, *next* leaves the state unchanged.

An input/output property of a program c is captured by a triple $\models \{P\}\, c\, \{Q\}$, where P and Q are separation logic predicates on the state. The triple is true iff for any state s and predicate F such that

$$(P \wedge^* \mathit{code}(c) \wedge^* F)\ s$$

holds, there exists a natural number k such that

$$(Q \wedge^* \mathit{code}(c) \wedge^* F)\ \mathit{next}^k(s)$$

holds. The predicate F is usually called in this context a *frame*, and keeping it allows us to reason about parts of states locally. $\mathit{code}(c)$ is another element specifying that the program code is present in the state, and next^k denotes k-times iteration of *next*. Such triples are highly generic and we cannot conclude much from many of these, except that under given preconditions the program c will pass a state satisfying Q. This changes immediately in cases where Q is of the form $\mathit{not\text{-}continuing} \wedge^* Q'$, now stating that c has reached a terminating state satisfying Q'. Hence, showing $\models \{P\}\, c\, \{\mathit{not\text{-}continuing} \wedge^* Q'\}$ amounts to showing termination of the program c in a Q'-state.

This generic technique applies seamlessly to EVM programs as shown by Hirai [7]. However, we realised that showing termination of a contract individually is an unnecessary burden because Ethereum was designed in such a way that all smart contracts are guaranteed to terminate (either successfully or due to an 'out-of-gas' exception). More specifically, to ensure that miners get compensated for their costs incurred by operating the Ethereum blockchain, each EVM instruction has a gas fee. When invoking a contract, the initiator provides a gas budget proportional to how computationally expensive the execution is expected to be and every step of execution is deducted from the budget. If the gas consumption exceeds the budget, an 'out-of-gas' exception is raised, the miner keeps all of the gas and the state of the contract prior to the invocation is restored.

These blockchain specifics give us a termination order. We augment the EVM formalisation [7] by the function next^* which consequently assigns to each state the terminal element from the reflexive-transitive closure of *next*. The essential property of next^* is that for any state s there exists $k \geq 0$ such that

(i) $\mathit{next}^*(s) = \mathit{next}^k(s)$ holds;
(ii) for any l, such that $0 \leq l < k$, the state $\mathit{next}^l(s)$ contains *continuing*;
(iii) for any l, such that $l \geq k$, the state $\mathit{next}^l(s)$ contains *not-continuing*.

In other words, $\mathit{next}^*(s) = \mathit{next}^k(s)$ where k is the least number such that $\mathit{next}^k(s)$ reaches a state with the *not-continuing* element.

Now, regarding the contract correctness, we strengthen $\models \{P\}\, c\, \{Q\}$ to a total correctness property $\models [P]\, c\, [Q]$ which is true iff for any state s and frame F

$$(P \wedge^* \mathit{code}(c) \wedge^* F)\ s$$

implies

$$(Q \wedge^* \mathit{code}(c) \wedge^* F)\ \mathit{next}^*(s)$$

To sum up, what we achieved so far is to factor out the termination part which we have shown once and for all, thus removing this obligation from the verification process completely. In the next section we will present our program logic handling EVM bytecode, which (based on soundness presented in Section 5) will give us a sound device to derive verification conditions for contract properties of the form $\models [P]\, c\, [Q]$.

4 Program Logic

A Hoare-style program logic comprises a collection of rules that allow us to derive semantic properties of compound programs from properties of its parts. In the case of structured languages we usually have, for instance, a rule telling us that a program **if** C **then** p_1 **else** p_2 exhibits a certain input/output behaviour if p_1 and p_2 do so, however, with the additional precondition C available for p_1 and $\neg C$ for p_2. The situation is not that simple when we have to reason about EVM bytecode. At this level, a conditional compound construct appears merely as a jump instruction that transfers the flow of execution to another part of the program. In this sense, a program logic that treats the entire bytecode program simply as a list of instructions would be in principle feasible, but intricate. To this end significantly more sophisticated techniques are available, such as decompilation via extraction of *Control Flow Graphs* (CFG), in particular applied to the Java Virtual Machine (JVM) code [19]. The aim of CFG extraction is to split a program into *basic blocks*, i.e. sequences of instructions without jumps, and connect them using edges corresponding to jumps. The essential property of basic blocks is that they comprise straight-line code, i.e. the control flow always enters it at its first instruction and leaves only after the last one has been executed.

However, full CFG extraction poses more advanced challenges in EVM context than for JVM, especially from the formal modelling perspective. This is because JVM jump instructions take their target address as an immediate value argument which can be determined statically, whereas in EVM jump destinations must be obtained from the stack, i.e. dynamically. For that reason, our bytecode preprocessing currently addresses basic

block extraction only, presented in the next section. Then, Sections 4.2, 4.3 and 4.4 present how our logic handles programs, blocks and instructions, respectively.

4.1 Extraction of Basic Blocks

We divide EVM instructions into three groups:

(i) `JUMPDEST` indicates a jump destination and hence beginning of a basic block;
(ii) `JUMP`, `JUMPI`, `UNKNOWN` and all of *Misc*-instructions[3] indicate the end of a basic block (`UNKNOWN` and Misc-instructions interrupt program execution);
(iii) all remaining instructions.

Furthermore, we classify basic blocks with the following four types:

(i) *Terminal* – if the last instruction of the block interrupts execution;
(ii) *Jump* – if the last instruction is `JUMP`;
(iii) *Jumpi* – if the last instruction is `JUMPI`;
(iv) *Next* – otherwise, i.e. when control passes from the last instruction of the block to the instruction with the successor address.

block index	address	instruction	block type
0	0	OR	*Next*
	1	ADD	
	2	SWAP1	
3	3	JUMPDEST	*Jump*
	4	MLOAD	
	5	POP	
	6	JUMP	
7	7	DUP3	*Jumpi*
	8	PUSH1 0	
	10	ISZERO	
	11	JUMPI	
12	12	POP	*Terminal*
	13	RETURN	

Figure 2. A program split into four basic blocks, where grey instructions appear in the original code but are removed from the list of instructions of their block.

Figure 2 illustrates how we split EVM bytecode into basic blocks of different types, thereby indexing the blocks with addresses of their first instruction and removing all the jumps from the block contents. The entire extraction process is captured in the Isabelle development by means of the function *build-blocks* which maps a list of instructions to a list of tuples (i, xs, t), where i is the block index, xs — the list of instructions of the block, and t — the type of the block.

[3]RETURN, STOP, SUICIDE, CREATE, CALL, CALLCODE, DELEGATECALL

By splitting bytecode into basic blocks no information gets lost since we can connect the produced blocks in the right order and insert jumps back in accordance with block types. More precisely, we also have the function *connect-blocks* such that for any bytecode program c the identity

$$\mathit{connect\text{-}blocks}(\mathit{build\text{-}blocks}\ c) = c$$

holds.

Based on these preparations, the following predicates will be defined inductively in the following three sections:

(i) $\mathit{blocks} \vdash_{\texttt{prog}} [P]\ (n, xs, t)\ [Q]$

(ii) $\vdash_{\texttt{block}} [P]\ xs\ [Q]$

(iii) $\vdash_{\texttt{instr}} [P]\ x\ [Q]$

where P, Q are state predicates, x — an instruction, xs — a list of instructions, *blocks* — a list of basic blocks, and (n, xs, t) — a basic block.

4.2 Program Rules

We start at the program level, where we have the following rules for each block type.

$$\text{(i)}\quad \frac{\vdash_{\texttt{block}} [P]\ xs\ [Q]}{\mathit{blocks} \vdash_{\texttt{prog}} [P]\ (n, xs, \mathit{Terminal})\ [Q]}$$

That is, a *Terminal*-block is simply passed to the level of blocks as we do not need to look at any other block after this has been processed.

The rule for a *Next*-block is different in this sense:

$$\text{(ii)}\quad \frac{\begin{array}{l}\vdash_{\texttt{block}} [P]\ xs\ [pc\ m \wedge^* R] \\ (m, ys, t) \in \mathit{blocks} \\ \mathit{blocks} \vdash_{\texttt{prog}} [pc\ m \wedge^* R]\ (m, ys, t)\ [Q]\end{array}}{\mathit{blocks} \vdash_{\texttt{prog}} [P]\ (n, xs, \mathit{Next})\ [Q]}$$

The state element $pc\ m$ determines the program counter after xs has been processed. Then we need to retrieve the next block associated to the index m from the structure *blocks*, which is expressed by $(m, ys, t) \in \mathit{blocks}$, and proceed with (m, ys, t).

Further, in case of a *Jump*-block we additionally need to retrieve the address of the jump destination from the stack after the block has been processed. This yields the following, slightly more involved, rule:

$$\text{(iii)}\quad \frac{\begin{array}{l}\vdash_{\texttt{block}} [P]\ xs\ [R_1] \\ (j, ys, t) \in \mathit{blocks} \\ \mathit{head}\ ys = (j, \texttt{JUMPDEST}) \\ \mathit{blocks} \vdash_{\texttt{prog}} [R_2]\ (j, ys, t)\ [Q]\end{array}}{\mathit{blocks} \vdash_{\texttt{prog}} [P]\ (n, xs, \mathit{Jump})\ [Q]}$$

where R_1 and R_2 abbreviate the conditions

$$\langle h \leq 1023 \wedge g \geq 8 \rangle \wedge^* \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ g \wedge^* \mathit{pc}\ i\ \wedge^* \mathit{stack\text{-}height}\ (h+1) \wedge^* \mathit{stack}\ h\ j \wedge^* R$$

and

$$\begin{array}{l} \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ (g-8) \wedge^* \\ \mathit{pc}\ j\ \wedge^* \mathit{stack\text{-}height}\ h \wedge^* R \end{array}$$

respectively. Regarding the stack, $\mathit{stack\text{-}height}(h{+}1)$ and $\mathit{stack}\ h\ j$ specify a state where index h refers to the top of the stack containing j — the jump destination we are looking for. Regarding gas, $\mathit{gas\text{-}pred}\, g$ binds the available amount of gas to g, whereas the part $\langle h \leq 1023 \wedge g \geq 8\rangle$ is a *pure* condition, i.e. not dependent on state, and sets an upper bound for the height of the stack and a lower bound for the amount of gas, namely 8 units: as much as EVM requires to perform a jump, which is deducted by $\mathit{gas\text{-}pred}\ (g-8)$. Note that here and in the rules below we need to carry the *continuing* state element because the EVM model [7] imposes having either *continuing* or *not-continuing* in a state to process instructions.

The case of a conditional jump. i.e. a *Jumpi*-block, is similar, except that we also need to retrieve from the stack the value c to be compared to 0 and jump only if $c \neq 0$:

$$\text{(iv)}\quad \frac{\begin{array}{l} \vdash_{\texttt{block}}\ [P]\ \mathit{xs}\ [R_1] \\ (j, \mathit{ys}, t) \in \mathit{blocks} \\ (i, \mathit{zs}, t') \in \mathit{blocks} \\ \mathit{head}\ \mathit{ys} = (j, \texttt{JUMPDEST}) \\ \mathit{blocks} \vdash_{\texttt{prog}}\ [R_2]\ (j, \mathit{ys}, t)\ [Q] \quad \bullet \text{ if } c \neq 0 \\ \mathit{blocks} \vdash_{\texttt{prog}}\ [R_3]\ (i, \mathit{zs}, t')\ [Q] \quad \bullet \text{ if } c = 0 \end{array}}{\mathit{blocks} \vdash_{\texttt{prog}}\ [P]\ (n, \mathit{xs}, \mathit{Jumpi})\ [Q]}$$

where R_1, R_2, R_3 abbreviate the conditions

$$\begin{array}{l} \langle h \leq 1022 \wedge g \geq 10\rangle \wedge^* \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ g \wedge^* \\ \mathit{pc}\ (i-1)\ \wedge^* \mathit{stack\text{-}height}\ (h+2)\ \wedge^* \mathit{stack}\ (h+1)\ j \wedge^* \\ \mathit{stack}\ h\ c\ \wedge^* R \end{array}$$

and

$$\begin{array}{l} \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ (g-10) \wedge^* \\ \mathit{pc}\ j\ \wedge^* \mathit{stack\text{-}height}\ h \wedge^* R \end{array}$$

and

$$\begin{array}{l} \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ (g-10) \wedge^* \\ \mathit{pc}\ i\ \wedge^* \mathit{stack\text{-}height}\ h \wedge^* R \end{array}$$

respectively.

4.3 Block Rules

At the level of basic blocks we need only two simple rules that handle the cases of non-empty and empty lists of instructions to be processed:

$$\text{(i)}\quad \frac{\vdash_{\texttt{instr}}\ [P]\ x\ [R] \qquad \vdash_{\texttt{block}}\ [R]\ \mathit{xs}\ [Q]}{\vdash_{\texttt{block}}\ [P]\ x{::}\mathit{xs}\ [Q]}$$

and

$$\text{(ii)}\quad \frac{P \Rightarrow Q}{\vdash_{\texttt{block}}\ [P]\ \mathit{Nil}\ [Q]}$$

where $P \Rightarrow Q$ means $P\ s$ implies $Q\ s$ for any state s.

4.4 Instruction Rules

To be able to verify any possible EVM bytecode program we need to provide a rule for each of 70 EVM instructions. Presently we have rules for 30 most commonly used instructions, and extend this set gradually 'on-demand'. The following rule for `PUSH1`, the operation pushing one byte on the stack, is quite representative as it shows how we specify the necessary conditions for the instruction to be performed by the EVM in the precondition as well as the effect of the operation on state elements in the postcondition:

$$\vdash_{\texttt{instr}}\ [P]\ (n, \texttt{PUSH1}\ x)\ [Q]$$

where P stands for

$$\begin{array}{l} \langle h \leq 1023 \wedge g \geq 3\rangle \wedge^* \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ g \wedge^* \\ \mathit{pc}\ n\ \wedge^* \mathit{stack\text{-}height}\ h\ \wedge^* F \end{array}$$

and Q for

$$\begin{array}{l} \mathit{continuing} \wedge^* \mathit{gas\text{-}pred}\ (g-3) \wedge^* \\ \mathit{pc}\ (n+1)\ \wedge^* \mathit{stack\text{-}height}\ (h+1) \wedge^* \mathit{stack}\ h\ x \wedge^* F \end{array}$$

In particular, we have stated that the height of the stack increases by 1 ($\mathit{stack\text{-}height}\ (h+1)$) and that the top index h of the stack points to the value x ($\mathit{stack}\ h\ x$) among the effects of `PUSH1`. It is also worth noting that we incorporate frames into such instruction-specific rules by carrying a variable F in pre- and postconditions, as shown above. This is in contrast to the more common way, which is introducing a generic *frame* rule (cf. [16]) of the form

$$\frac{\vdash_{\texttt{instr}}\ [P]\ i\ [Q]}{\vdash_{\texttt{instr}}\ [P \wedge^* F]\ i\ [Q \wedge^* F]}$$

and removing F from all instruction-specific rules. Although the rule is sound, this treatment leads to a considerable overhead in the verification process, since we would need to apply the frame rule each time an instruction-specific rule is applied.

Apart from the frame rule we still have two generic rules at the instruction level:

$$\text{(i)}\quad \frac{\begin{array}{l} \vdash_{\texttt{instr}}\ [P']\ i\ [Q'] \\ P \Rightarrow P' \\ Q' \Rightarrow Q \end{array}}{\vdash_{\texttt{instr}}\ [P]\ i\ [Q]}$$

$$\text{(ii)}\quad \vdash_{\texttt{instr}}\ [\langle \mathit{False}\rangle]\ i\ [Q]$$

The rule (i) is the usual 'consequence' rule, allowing us to adjust pre- and postconditions, whereas (ii) is needed

to discharge trivial proof obligations having unsatisfiable preconditions. Such obligations arise frequently from conditional jumps (rule (iv), Section 4.2) where the condition is fully evaluated prior to the actual jump, such that we need to follow only one of the emerging branches.

The following section puts the program logic and the results of Section 3 together by means of a soundness property and outlines its proof.

5 Soundness

As our program logic is separated in three layers, we establish its soundness in three steps.

At the level of instructions, soundness basically amounts to the property

$$\frac{\vdash_{\text{instr}} [P]\, x\, [Q] \qquad (P \wedge^* \mathit{code}([x]) \wedge^* F)\ s}{(Q \wedge^* \mathit{code}([x]) \wedge^* F)\ \mathit{next}(s)} \tag{1}$$

which we prove by structural induction on $\vdash_{\text{instr}} [P]x[Q]$. By this, we need to show that the pre- and postconditions, as specified in each rule for individual instructions, are indeed covered by the behaviour of the respective instruction. Note that $\mathit{code}([x])$ ensures that $x = (\mathit{addr}, \mathit{instr})$ is present in the code-element of s, such that next executes precisely the instruction instr, if $\mathit{pc}\ \mathit{addr}$ is present in s as well. This, in turn, is obtained from the preconditions of instruction rules, such as the `PUSH1`-rule from the previous section.

Next, at the level of blocks we show

$$\frac{\vdash_{\text{block}} [P]\ \mathit{xs}\ [Q] \qquad (P \wedge^* \mathit{code}(\mathit{xs}) \wedge^* F)\ s}{(Q \wedge^* \mathit{code}(\mathit{xs}) \wedge^* F)\ \mathit{next}^{|\mathit{xs}|}(s)} \tag{2}$$

where the usage of $\mathit{next}^{|\mathit{xs}|}$ is justified, since we consider a basic block xs which requires precisely $|\mathit{xs}|$ steps to be processed completely. By induction on $\vdash_{\text{block}} [P]\ \mathit{xs}\ [Q]$ we need to consider the cases when xs is non-empty or empty. In case $\mathit{xs} = x :: \mathit{zs}$ we can assume $\vdash_{\text{instr}} [P]\, x\, [R]$ and $\vdash_{\text{block}} [R]\ \mathit{zs}\ [Q]$ such that

$$\frac{(R \wedge^* \mathit{code}(\mathit{zs}) \wedge^* F)\ s}{(Q \wedge^* \mathit{code}(\mathit{zs}) \wedge^* F)\ \mathit{next}^{|\mathit{zs}|}(s)}$$

holds by the induction hypothesis. Furthermore, from (1) and $\vdash_{\text{instr}} [P]\, x\, [R]$ we can further conclude

$$\frac{(P \wedge^* \mathit{code}([x]) \wedge^* F)\ s}{(R \wedge^* \mathit{code}([x]) \wedge^* F)\ \mathit{next}(s)}$$

which combined establish (2). In case xs is empty, we can assume $P \Rightarrow Q$ which immediately gives us (2).

The ultimate soundness statement is at the program level:

$$\frac{\mathit{build\text{-}blocks}\ c \vdash_{\text{prog}} [P]\ \mathit{first\text{-}block}\ [Q] \qquad 0 < |c| < 2^{256}}{\models [P]\, c\, [Q]} \tag{3}$$

where *first-block* is a shorthand for the block with the smallest index in *build-blocks* c, and the assumption $0 < |c| < 2^{256}$ is necessary to avoid dealing with empty programs as well as programs with more than 2^{256} instructions (imposed by the design of EVM). In other words, in order to establish an input/output property specified by $\models [P]\, c\, [Q]$, we can transform c into its basic blocks bs, pick the first block b from bs, and apply the rules of our program logic to derive

$$\mathit{bs} \vdash_{\text{prog}} [P]\ b\ [Q]$$

However, in order to show (3) we need some preparations to be able to apply structural induction on the program logic rules. To this end we deploy our function *connect-blocks* and state the proposition

$$\frac{\mathit{bs} \vdash_{\text{prog}} [P]\ b\ [Q] \qquad b \in \mathit{bs} \qquad \mathit{wf\text{-}blocks}\ \mathit{bs}}{\models [P]\ \mathit{connect\text{-}blocks}\ \mathit{bs}\ [Q]} \tag{4}$$

where *wf-blocks* is our well-formedness predicate capturing all necessary technical details about block structure, essentially retaining the property

$$\frac{0 < |c| < 2^{256}}{\mathit{wf\text{-}blocks}(\mathit{build\text{-}blocks}\ c)}$$

for any program c.

Thus, the proposition (4) is a generalisation of (3), since for any c we can instantiate bs by *build-blocks* c and b by *first-block* in (4), and use the identity

$$\mathit{connect\text{-}blocks}(\mathit{build\text{-}blocks}\ c) = c$$

from Section 4.1. to obtain (3).

Unfolding the definition of $\models [P]\ \mathit{connect\text{-}blocks}\ \mathit{bs}\ [Q]$ in (4) we further obtain

$$\frac{\mathit{bs} \vdash_{\text{prog}} [P]\ b\ [Q] \qquad b \in \mathit{bs} \qquad \mathit{wf\text{-}blocks}\ \mathit{bs} \qquad (P \wedge^* \mathit{code}(\mathit{connect\text{-}blocks}\ \mathit{bs}) \wedge^* F)\ s}{(Q \wedge^* \mathit{code}(\mathit{connect\text{-}blocks}\ \mathit{bs}) \wedge^* F)\ \mathit{next}^*(s)} \tag{5}$$

and can proceed by induction on $\mathit{bs} \vdash_{\text{prog}} [P]\ b\ [Q]$. By this, we have to consider four cases: one for each type of the block b. So, for instance, if $b = (n, \mathit{xs}, \mathit{Terminal})$, i.e. a terminal block, we have $\vdash_{\text{block}} [P]\ \mathit{xs}\ [Q]$. Since b is a part of the block list bs by assumption, we can separate some bs' such that

$$(P \wedge^* \mathit{code}(\mathit{connect\text{-}blocks}\ \mathit{bs}) \wedge^* F)\ s \quad = \quad (P \wedge^* \mathit{code}(\mathit{xs}) \wedge^* \mathit{code}(\mathit{connect\text{-}blocks}\ \mathit{bs}') \wedge^* F)\ s$$

Hence, we can instantiate frame F in (2) by

$$code(\textit{connect-blocks } bs') \wedge^* F$$

and consequently obtain

$$(Q \wedge^* code(\textit{connect-blocks } bs) \wedge^* F)\ next^{|xs|}(s)$$

As we consider a terminal block, the state $next^{|xs|}(s)$ is the first one containing the *not-continuing* element, i.e.

$$next^{|xs|}(s) = next^*(s)$$

holds, concluding this case.

Although slightly more involved, the proof of the remaining three cases follows the same principles, making however additional use of the induction hypothesis.

6 Case study

Our development provides the ground work for full functional correctness of Ethereum smart contracts. These contracts are typically implemented in a language called Solidity, which provides high level abstractions to facilitate structured development. In Solidity, the functionality of a smart contract is encapsulated into a *contract interface*, which, analogously to a class in object oriented programming (OOP), has a well-defined interface with public and private elements. Creating a contract on the blockchain instantiates the contract interface. Just like class instantiations in OOP, contracts on the blockchain are stateful objects, where the storage is persistent across contract function calls.

```
contract MyContract {
  function dispatch1() public returns (uint) {
    return (1);
  }
  function dispatch2() public returns (uint) {
    return (2);
  }
}
```

Figure 3. Contract with two functions: *dispatch1* that returns 1; *dispatch2* that returns 2.

Figure 3 shows an example of a Solidity contract with two public functions. Each public function of a contract gets assigned a unique hash and the Solidity compiler produces EVM bytecode with a single entry point. At the beginning of the bytecode a *dispatcher* is introduced that inspects the first arguments passed when the contract gets invoked, compares it with the function hashes and calls the appropriate function. The binary format for computing function hashes and packing arguments is standardised in an ABI (abstract binary interface), enabling interoperability between contracts implemented using various EVM languages.

Since the contract abstraction is provided by the Solidity language, it is the job of the compiler to implement the dispatcher. This means that the code can only be verified at the level of bytecode, hence it is a great application to showcase our framework. In this case study we show how we specified and verified the dispatcher produced for the code in Figure 3.

6.1 Specification

```
definition
  spec_MyContract :: 32 word ⇒ contract_action
where
  spec_MyContract z =
  if z = dispatch1_hash
  then ContractReturn (word_rsplit 1))
  else (if z = dispatch2_hash
        then ContractReturn (word_rsplit 2))
        else ContractFail [ShouldNotHappen])

theorem verify_dispatcher:
 ∃r. ⊨ [pc 0 ∧* stack_height 0 ∧*
        sent_data (word_rsplit arg) ∧*
        sent_value 0 ∧* memory_usage 0 ∧*
        continuing ∧* gas_pred 3000 ∧*
        ... ]
         (build_blocks bytecode_MyContract)
       [action (spec_MyContract arg) ∧* r]
```

Figure 4. Specification and functional correctness theorem of MyContract.

Figure 4 shows our handwritten specification of *MyContract* in Isabelle/HOL. The Solidity compiler can print a contract ABI description which includes the signatures of all public functions as well as their hashes. We used this feature to obtain *dispatch1_hash* and *dispatch2_hash*. The rest of the specification is straightforward. The argument passed to *spec_MyContract* corresponds to the first 32-bit machine word passed as an argument when the contract gets invoked. *ContractReturn* and *ContractFail* specify the action resulting from the contract execution. Upon successful execution, contracts return a byte array, hence we use *word_rsplit* to convert a word into a list of bytes encoded in big-endian format. When none of the function hashes match the first contract argument, the dispatcher hits an `INVALID`[4] instruction, which corresponds to raising a *ShouldNotHappen* exception in our bytecode semantics.

The theorem at the bottom of Figure 4 states the input/output property we proved. It has the form of Hoare triples introduced in Section 3. The precondition

[4] `INVALID` is the mnemonic of an EVM opcode intentionally kept undefined in order to raise an exception.

initialises the machine state, e.g. program counter is 0, stack is empty, etc. Importantly, *arg* is the 32-bit word contract argument used by the dispatcher code as well as the specification, and *gas-pred* specifies the gas budget for the execution. *bytecode_MyContract* is a list of deeply embedded EVM instructions which we convert into a list of basic blocks with *build-blocks*. In the postcondition, we are only interested in the resulting action satisfying *spec_MyContract*, hence we quantify existentially on the rest of the state to leave it unspecified.

6.2 Verification

Smart contract bytecode is divided into two sections: the pre-loader and runtime code. The pre-loader bootstraps the contract by deploying it on the Ethereum network and running its constructor (the contract in Figure 3 has no constructor). The runtime code contains the core functionality of the contract that can be invoked by other blockchain agents. Our verification considers only the runtime code of MyContract.

We obtain the runtime code as a byte string of opcodes from the Solidity compiler and convert it into a list of EVM instructions via an Isabelle/HOL function. In this sense we work directly on the bytecode output of the Solidity compiler.

MyContract bytecode has 227 EVM instructions, that get split into 27 basic blocks, including 4 of type *Jumpi* (i.e. conditional jumps). Once the automation support, described in the next section reached maturity, proving correctness of the bytecode was rather a routine task.

A difficult proof arose when parsing the contract input data in order to extract the argument passed to the dispatcher. This involved dealing with operations on words of different size, which is notoriously hard. In particular, the instruction `CALLDATALOAD` reads a 256-bit machine word from the input data array. Since the hash value passed in the input data is only a 32-bit word, the dispatcher does the following word arithmetic to convert the value:

$$w32\text{-}hash = (w256\text{-}input\text{-}data \gg 224)\ \&\ \texttt{0xffffffff}$$

With the help of Isabelle's machine word library [2], we proved that when the hash is packed in the input data, the above bit-wise operations return the same hash value.

The total development of our framework is ≈5000 lines of Isabelle/HOL theories, excluding the existing formalisation of EVM model, etc. The size of the top-level specification of MyContract is ≈15 lines and the functional correctness proof of the dispatcher is ≈50 lines of proof specific to this example, which compares favourably with the ≈500 lines of reusable proof automation machinery we developed. We describe this automation next.

6.3 Automation

When reasoning about bytecode, even the verification of small smart contracts can involve long, tedious and repetitive proofs. Hence, our program logic was purposefully designed to be amenable to proof automation.

The inductively defined inference rules presented in Section 4 are all designed to be used in a syntactically-driven verification condition generator (VCG). Shaping the rules in a way that only one of them applies at each point in the proof makes it trivial to write a VCG that merely tries applying all the rules one after another. Such a VCG can be implemented with a few lines of Eisbach [10] — Isabelle/HOL's high-level tactic language. We developed a VCG for each level of our program logic. For instance, the program-level tactic looks like this:

```
method prog_vcg =
 (prog_jumpi_vcg | prog_jump_vcg
 | prog_terminal_vcg | prog_next_vcg)+
```

where e.g., *prog_jumpi_vcg* is another tactic applying the *Jumpi*-rule (iv) in Section 4.2 and solving its resulting subgoals by invoking more specific tactics we designed. *prog_vcg* tries each of the tactics separated by the | symbol, while the + sign at the end means that it will apply repetitively until none of the tactics can apply on the goal.

A common source of friction we experienced during the early development of our framework was with proving goals of the form: $\forall s.\ R\ s \longrightarrow P\ s$ where P and R are separation logic predicates comprising the same separation conjunctions but in different orders. Such proof obligations arise when we weaken a precondition to be able to apply an $\vdash_{\texttt{instr}}$ rule. Since each $\vdash_{\texttt{instr}}$ rule expects a separation logic expression with conjunctions in a specific order, we routinely have to re-order these to match a given precondition.

To ease the pain, we reuse the separation logic algebra framework [8], which provides a set of generic Isabelle tactics to manipulate separation logic terms. We instantiated the algebra with the EVM machine state and created Isabelle tactics which re-order separation conjunctions such that, e.g. the first term in P matches the first one of R. Once the first elements match, we leverage the tactics of the separation algebra framework to remove the first element from both R and P.

An issue we encountered is that when R contains variables, we have to re-order terms in P in such way that the variables in R get instantiated in the correct order. When such problem occurs, we resort to manual reordering of terms.

7 Related Work

As a consequence of the repeated exploitation of security flaws in smart contracts, a notable amount of approaches and tools have already been proposed (e.g. [1, 3, 9]).

The major trend is to apply various kinds of static analysis not only at the level of structured contract languages (e.g. Solidity's Why3 backend [4, 15]) but also at the bytecode level [1, 9, 17]. The obvious advantage of static analysis approaches is that full automation can be achieved for contract properties that can be confirmed statically, e.g. certain orders of transactions [1]. Oyente [9] and Porosity [17] decompile bytecode into a control flow graph and perform control flow analysis in order to detect common smart contract security defects such as reentrancy bugs. These tools are of great value when added to the development process so they can unveil mistakes in early stages, but they do not prove functional correctness.

Our approach is more general as we aim to specify and verify contract properties in pre/postcondition style where the conditions can comprise any higher-order logic formula describing an EVM state. For that reason, the degree of automation is limited in our case such that the user will need to interact with the proof system to discharge elaborated claims.

Bhargavan et al. [3] proposed a technique using an intermediate functional language called F^*, which is more amenable to verification. It provides not only translation of a subset of Solidity programs to F^*, but also decompilation of EVM bytecode to F^* as well. This use of decompilation makes the approach similar to ours, since our program logic, in fact, resembles decompilation. As explained in Section 4, we split bytecode program into blocks without jumps and determine the actual jump destinations dynamically 'on-the-fly', by applying logic rules. By contrast, Bhargavan et al. perform static stack analysis to this end. However, a more striking difference is that our approach is homogeneous since every step of our verification process is performed and justified within a single, trusted logical framework without any translations to or from other formalisms. Such translations must be either assumed to behave correctly in some sense or formally modelled and verified, whereas we aimed to avoid both of these options.

KEVM [6] is a formal semantics of the EVM written using the K-framework. Like the Lem semantics [7] we use, KEVM is executable and therefore can run the Ethereum foundation's validation test suite. Reasoning about KEVM programs involves specifying properties in Reachability Logic and verifying them with a separate analysis tool. K supports translation for analysis with tools of varying power, including symbolic execution, faster concrete execution, or Isabelle. As explained earlier, we preferred the option of working in a single trusted logical framework.

8 Conclusions and Future Work

In this paper we have presented our approach to the verification of Ethereum smart contracts at the level of EVM bytecode. Building strictly on the thoroughly validated formal EVM model [7] in Isabelle/HOL, we have augmented it using the fact that EVM gas consumption allows us to state properties of contracts in pre/postcondition style with all termination considerations discharged. Further, we have outlined how we split contracts into a structure of basic blocks as well as how a sound program logic proceeds from such blocks down to the level of instructions. The presented case study has demonstrated the applicability of our program logic to real bytecode, verifying dispatch code that exhibits a structure that appears in a similar form in each Solidity-generated bytecode program. Moreover, the case study outlined how we use Isabelle tactics to automate large parts of verification condition generation process.

To further foster formal verification of Ethereum smart contract, we could restore Solidity's control structures and function calls. For instance, restoring loops would require using heuristics to detect them and the program logic would be proved sound only for the subset of EVM bytecode accepted by this heuristic. Similarly detecting function calls in EVM bytecode requires complex stack analysis because the EVM provides no support to call subroutines and for stack unwinding. Thus, heuristics must be used to distinguish call sites and stack unwinding from other stack-manipulating instructions.

The current framework leaves inter-contract reasoning outside the program logic. When a contract A makes an external call to contract B, the semantics terminates with an *InstructionToEnvironment* action. Hence, within our program logic, we can only prove properties about the state of A right before it calls B. To reason further about such interactions, the environment must be modelled at a meta-level to capture the behaviour of B. We could imagine a framework where the program logic is parametrised by a contract environment which gets invoked when an external call occurs.

Another promising avenue for research is certifying compilers. We believe, for example, that the Ethereum community would greatly benefit from an EVM backend for the CakeML [14] verified compiler. Since CakeML is a functional programming language, it would greatly simplify the verification of high-level properties of smart contracts compared to reasoning about bytecode directly. Yet the compiler correctness proof would imply the same level of guarantee as bytecode verification.

References

[1] 2017. Securify. https://securify.ch. (2017).

[2] Joel Beeren, Matthew Fernandez, Xin Gao, Gerwin Klein, Rafal Kolanski, Japheth Lim, Corey Lewis, Daniel Matichuk, and Thomas Sewell. 2016. Finite Machine Word Library. *Archive of Formal Proofs* (June 2016). http://isa-afp.org/entries/Word_Lib.html, Formal proof development.

[3] Karthikeyan Bhargavan and *et al.* 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, 91–96.

[4] Jean-Christophe Filliâtre and Andrei Paskevich. 2013. Why3 – where programs meet provers. In *European Symposium on Programming*. Springer, 125–128.

[5] Ethereum foundation. 2017. Solidity documentation. https://solidity.readthedocs.io/en/develop/. (2017).

[6] Everett Hildenbrandt, Manasvi Saxena, Xiaoran Zhu, Nishant Rodrigues, Philip Daian, Dwight Guth, and Grigore Rosu. 2017. *KEVM: A Complete Semantics of the Ethereum Virtual Machine*. Technical Report.

[7] Yoichi Hirai. 2017. Defining the Ethereum Virtual Machine for Interactive Theorem Provers. In *WTSC'17, 1st Workshop on Trusted Smart Contracts, International Conference on Financial Cryprography and Data Security*.

[8] Gerwin Klein, Rafal Kolanski, and Andrew Boyton. 2012. Mechanised Separation Algebra. In *International Conference on Interactive Theorem Proving*, Lennart Beringer and Amy Felty (Ed.). Springer, Princeton, USA, 332–337.

[9] Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor. 2016. Making smart contracts smarter. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 254–269.

[10] Daniel Matichuk, Toby Murray, and Makarius Wenzel. 2016. Eisbach: A Proof Method Language for Isabelle. *Journal of Automated Reasoning* 56, 3 (March 2016), 261–282. https://doi.org/10.1007/s10817-015-9360-2

[11] Dominic P. Mulligan, Scott Owens, Kathryn E. Gray, Tom Ridge, and Peter Sewell. 2014. Lem: reusable engineering of real-world semantics. In *Proceedings of the 19th ACM SIGPLAN international conference on Functional programming, Gothenburg, Sweden, September 1-3, 2014*, Johan Jeuring and Manuel M. T. Chakravarty (Eds.). ACM, 175–188.

[12] Magnus Oskar Myreen. 2009. *Formal verification of machine-code programs*. Ph.D. Dissertation. University of Cambridge, UK.

[13] Tobias Nipkow, Lawrence Paulson, and Markus Wenzel. 2002. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*. Lecture Notes in Computer Science, Vol. 2283. Springer.

[14] Scott Owens, Michael Norrish, Ramana Kumar, Magnus O. Myreen, and Yong Kiam Tan. 2017. Verifying efficient function calls in CakeML. *Proceedings of the ACM on Programming Languages* 1, ICFP (2017), 18.

[15] Christian Reitwiessner. 2016. Formal Verification of Smart Contracts. https://chriseth.github.io/notes/talks/formal_ic3_bootcamp. (2016).

[16] John C Reynolds. 2002. Separation logic: A logic for shared mutable data structures. In *Logic in Computer Science, 2002. Proceedings. 17th Annual IEEE Symposium on*. IEEE, 55–74.

[17] Matt Suiche. 2017. Porosity: A Decompiler For Blockchain-Based Smart Contracts Bytecode. https://github.com/comaeio/porosity/blob/master/defcon2017/dc25-msuiche-Porosity-Decompiling-Ethereum-Smart-Contracts-wp.pdf. (2017).

[18] Gavin Wood. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper* 151 (2014).

[19] Jianjun Zhao. 1999. Analyzing Control Flow in Java Bytecode. In *Proc. 16th Conference of Japan Society for Software Science and Technology*. 313–316.