

Electronics and Computer Science
Faculty of Engineering and Physical Sciences
University of Southampton

Gherman Nicolisin

April 20, 2024

**Folidity - Formally Verifiable
Smart Contract Language**

Project supervisor: Prof. Vladimiro Sassone
Second examiner: Dr. Indu Bodala

A project report submitted for the award of
BSc Computer Science

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
ELECTRONICS AND COMPUTER SCIENCE

A project report submitted for the award of BSc Computer Science

By Gherman Nicolisin

This paper addresses the long-lasting problem involving the exploits of Smart Contract vulnerabilities. There are tools, such as in the formal verification field and alternative Smart Contract languages, that attempt to address these issues. However, neither approach has managed to combine the static formal verification and the generation of runtime assertions. Furthermore, this work believes that implicit hidden state transition is the root cause of security compromises. In light of the above, we introduce Folidity, a formally verifiable Smart Contract language with a unique approach to reasoning about the modelling and development of Smart Contract systems. Folidity features explicit state transition checks, a model-first approach, and built-in formal verification compilation stage.

Statement of Originality

- I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.
- I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

You must change the statements in the boxes if you do not agree with them.

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

I have acknowledged all sources, and identified any content taken from elsewhere.

If you have used any code (e.g. open-source code), reference designs, or similar resources that have been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

I have not used any resources produced by anyone else.

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching

staff before you seek outside assistance, e.g. a proofreading service, and declare it here.

I did all the work myself, or with my allocated group, and have not helped anyone else

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

The material in the report is genuine, and I have included all my data/code/designs.

We expect that you have not previously submitted any part of this work for another assessment. You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

I have not submitted any part of this work for another assessment.

If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

My work did not involve human participants, their cells or data, or animals.

Contents

Statement of Originality	i
Contents	iii
1. Introduction	1
2. Security and Safety of Smart Contracts	2
2.1. Overview	2
2.2. Vulnerability classification	3
2.3. Setting the scene	3
3. Related Work	5
3.1. Overview	5
3.2. Formal Verification Tools	5
3.3. Safe Smart Contract Languages	6
3.4. Problem Statement	7
4. Proposed Solution	8
4.1. Outline	8
4.2. Language design	10
4.2.1. Primitives, Expressions and Statements	10
4.2.2. Declarations	12
4.3. Formal Verification	15
4.3.1. Model consistency	15
4.3.2. Proving constraint satisfiability	16
4.3.3. Other techniques	19
5. Implementation	20
5.1. Outline	20
5.2. Scope	21
6. Project Planning	22
A. Folidity Grammar	23
B. Gannt Chart	23
Bibliography	25

1. Introduction

The concept of “smart contract” (SC) was first coined by Nick Szabo as a computerised transaction protocol [1]. He later defined smart contracts as observable, verifiable, privacy-applicable, and enforceable programs [2]. In other words, smart contracts were envisioned to inherit the natural properties of traditional “paper-based” contracts.

In 2014 SCs were technically formalised at the protocol level by Dr. Gavin Wood as an arbitrary program written in some programming language (Solidity) and executed in the blockchain’s virtual machine of Ethereum [3].

Ethereum Virtual Machine (EVM) iterated over the idea of Bitcoin Scripting [4], allowing developers to deploy general-purpose, Turing-Complete programs that can have their own storage, hence the state, written in Solidity [5]. This enabled sophisticated applications that grew beyond the simple fund transfers among users.

Overall, SC can be summarised as an *immutable, permissionless, deterministic* computer program that is executed as part of state transition in the blockchain system [6], [3].

After a relatively short time, SCs have come a long way and allowed users to access different online services, also known as Decentralised Applications (DApps), in a completely trustless and decentralised way. The applications have spanned financial, health, construction [7], and other sectors.

2. Security and Safety of Smart Contracts

2.1. Overview

With the increased adoption of DApps and total value locked in them, there have been numerous attacks focused on extracting funds from SCs. Due to the permissionless nature of SCs, the most common attack vector exploits the mistakes in the SC's source code. Specifically, the attacker can not tamper with the protocol code due to consensus mechanisms. Instead, they can interact with the publicly accessible parameters and interfaces to force the SC into an unexpected state, essentially gaining partial control of it.

A notorious example is the DAO hack when hackers exploited unprotected re-entrance calls to withdraw **\$50 million worth of ETH**. This event forced the community to hard-fork the protocol to revert the transaction, provoking a debate on the soundness of the action [8].

Another less-known example is the “King of the Ether” attack, which was caused by the unchecked low-level Solidity `send` call to transfer funds to a contract-based wallet [9]. The “King of the Ether Throne” contract could not recognise the failed transaction on the wallet side. Instead, the contract proceeded with the operation, incorrectly mutating its internal state.

Other issues involve the *safety* and *liveness* of SCs. The term *safety* is used to describe the *functional safety* and *type safety*. *Functional safety* refers to the guarantees that the system behaves according to the specification irrespective of the input data [10], whereas *type safety* refers to the guarantees that the language provides a sound type system [11]. The two are often used interchangeably with the *security* of code as compromising the former affects the latter. When talking about *liveness*, we describe the business logic of a DApp, particularly whether it transitions into the expected new state [12]. This is particularly important for the execution of mission-critical software in a distributed context.

Safety and *liveness* can be compromised due to the programmer's mistakes in the source code that can result in the SC entering the terminal in an unexpected state preventing users from interacting with it [13].

2.2. Vulnerability classification

There has been an effort in both academia and industry to classify common vulnerabilities and exploits in SCs in blockchain systems [14]–[16]. Some of the work has been recycled by bug bounty platforms, growing the community of auditors and encouraging peer-review of SCs through the websites such as *Code4rena*¹, *Solodit*², and many others.

Analysing the work mentioned above, SC vulnerabilities can be categorised into the six general groups outlined in Table 2.1. The six categories have been defined based on the analysis of the most common vulnerabilities, and how they affect the SC execution. Each category represents the general scope for a specific set of vulnerabilities that should be addressed in the SC development.

2.3. Setting the scene

Even with the raised awareness for the security and safety of SCs, recent reports from *Code4rena* still show *SCV3*, *SCV4* and *SCV5* commonly present in the recent audit reports [17], [13], [18].

In particular, in [18], a relatively simple calculation mistake resulted in other SC users being unable to withdraw their funds.

It can be seen that SC Vulnerabilities illustrated in Table 2.1 are still evident in modern SCs, resulting in opening them up to exploits of different severity levels. Looking at the mentioned reports, there is little consensus about the weight of each vulnerability. Therefore, we can not classify any particular vulnerability as more severe than the other as it solely depends on the context in the code it is present. Furthermore, it has been realised that additional tooling or alternative SCLs need to be discovered to minimise the exposure of SC code to the earlier-mentioned vulnerabilities.

¹<https://code4rena.com>

²<https://solodit.xyz>

Code	Title	Summary
<i>SCV1</i>	Timestamp manipulation	Timestamp used in control-flow, randomness and storage, can open an exploit due to an ability for validator to manipulate the timestamp
<i>SCV2</i>	Pseudo-randomness	Using block number, block hash, block timestamp are not truly randomly generated parameters, and can be manipulated by the adversary validator
<i>SCV3</i>	Invalidly-coded states	When coding business logic, control-flow checks can be incorrectly coded resulting the SC entering into invalid state
<i>SCV4</i>	Access Control exploits	This is a more broad categorisation of vulnerabilities. It occurs when an adversary calls a restricted function. This is specifically present in <i>upgradeability</i> and <i>deleteability</i> of SCs
<i>SCV5</i>	Arithmetic operations	SCs are suspected to the same arithmetic bugs as classic programs. Therefore, unchecked operations can result in underflow/overflow or deletion by zero
<i>SCV6</i>	Unchecked external calls	Unchecked re-entrant, forward, delegate calls can result in the contract entering into unexpected state

Table 2.1: Classification of SC vulnerabilities

3. Related Work

3.1. Overview

Different solutions have been presented to mitigate the consistency in the presence of vulnerabilities and programmer mistakes. We can generally categorise them into two groups: safe SCLs, which allow users to write safe and secure code, particularly described in Section 3.3, and formal verification tools used alongside traditional SCLs presented in Section 3.2.

This chapter reviews both categories of tools, allowing us to evaluate their effectiveness in correlation to usability, aiming to provide a concise framework to analyse and work with the SC tools dedicated to producing error-proof DApps.

3.2. Formal Verification Tools

Formal verification describes the assessment of the correctness of a system concerning a formal specification [19]. The specification is usually described in terms of verifiable models using mathematical proofs. There are multiple ways to verify a program formally focused on specific parts. *Model checking* utilises propositional logic to verify the mathematical abstractions of the system [20]. *Theorem proving* involves verifying relations between the model and the statements about the system [21]. Finally, *symbolic execution* focuses on the execution of the program using symbolic values instead of concrete values [19].

KEVM³ is a tool that provides executable semantics of EVM using \mathbb{K} framework. It uses reachability logic to reason symbolically about the system [22]. KEVM is a powerful tool that operates at the EVM bytecode level. Specifically, SC developers are required to write a specification in a separate file that is checked against the compiled EVM bytecode of the SC. Whilst this provides more fine-grained assurance of the safety and correctness, it requires specialised knowledge of the \mathbb{K} framework and EVM semantics, hence significantly increasing the development time.

The other interesting tool is Dafny⁴. Dafny is a general-purpose tool that checks inputs in any language using Hoare-logic and high-level annotations. Although Dafny offers compilation to some system languages, Solidity is not yet a supported

³<https://jellopaper.org/index.html>

target. Notably, work in the field suggests that the Dafny can be an effective and easy-to-use tool to produce a formal specification [23]. The syntax resembles a traditional imperative style and is substantially easier to learn and understand than KEVM.

Some tools can be used alongside Solidity code, such as Scribble⁵. Scribble enables developers to provide formal specifications of functions inside docstrings seamlessly integrating with existing Solidity code. It offers VS Code extensions and is actively maintained by Consensys⁶. The trade-off is the limited expressiveness in comparison with KEVM and Dafny.

Finally, experiments have been conducted to verify SC without any formal annotations. In particular, VeriSmart focuses explicitly on ensuring arithmetic safety and preciseness in SCs [24]. However, VeriSmart fails to detect other types of errors, although an effort has been made to apply the verifier to more areas of SC [25].

Formal verification is a multi-disciplinary field offering multiple ways of reason about the systems. One of the actively researched topics is bounded model verification [26]. Developers are required to reason about the programs as finite state machines (FSM)[27]. This reasoning approach is more apparent in SC development since the state transition is at the core of blockchain execution. Bounded model checking has been realised by only a few experimental projects such as Solidifier [28] and Microsoft [25]. Both projects attempt to translate Solidity code to an intermediate modelling language, Boogie [29]. Boogie then leverages SMT solvers to find any assertion violations.

Overall, we can see that formal verification tools provide a robust way of ensuring the safety of SCs. While significant effort has been made in the field, it is evident that formal verification tools in SC development attempt to compensate for Solidity's implicit state transitions and lack of *implicit* safety.

3.3. Safe Smart Contract Languages

Multiple attempts have been made to address a flawed programming model of Solidity [30]. Alternative SCLs aim to provide built-in safety features in a type system, modelling, and function declaration to minimise the need for external tooling.

Some languages, such as Vyper⁷, strive for simplicity. By stripping off some low-level features, Vyper minimises the developer's chances of misusing the dangerous operations. It also provides overflow checking, signed integers, and other safe arithmetic operations. However, Vyper is still immature, and the recent bug in the

⁴<https://dafny.org/latest/DafnyRef/DafnyRef>

⁵<https://docs.scribble.codes>

⁶<https://consensys.io/diligence/scribble>

compiler caused a massive re-entrance exploit in the *curve.fi* AMM protocol [31]. Furthermore, Vyper still suffers from the same implicit state transition problem as Solidity.

Flint is an experiment language with protected calls and asset types [32]. Protected calls introduce a role-based access system where the SC developer can specify the permitted caller to a message function. Another unique feature is array-bounded loops that partially address the halting problem. Flint also addresses a state-transition problem by allowing developers to specify all possible states in the contract. The message functions need to specify the state transition, which occurs explicitly. The language provides a significant improvement in a modelling approach. However, it still lacks the modelling SC input data in terms of constraints and invariants, and explicit state transition is still an optional feature that the developer can miss in using.

Another promising SCL reasons about SC development through dependent and polymorphic types [33]. It extends Idris⁸ and makes the developer model the SC as part of a state transition function by adopting a functional programming style. Dependent types provide a more fine-grained control over the input and output data that flow through the SC functions. In particular, similar to Haskell, the language offers *side-effects* functionality that resembles *IO* monads in Haskell. The downside of the approach is that the syntax has become too cumbersome for other developers to learn. Thus, it has been stated that the language does not strive for simplicity and sacrifices it for safety.

3.4. Problem Statement

We can identify the positive trend in providing the safety of SCs. Modern formal verification methods offer support to SC developers in ensuring that their code satisfies the requirements of the system, while proposed SCL solutions offer runtime safety, minimising the need for the former.

However, there has been no effort to combine the two approaches into a single development process. Formal verification tools focus on the validation of functional correctness and model consistency of a program at the compile time, whereas SCLs focus on data validation at the runtime. Recent work suggests that the improved optimisation of SMT solvers allows us to turn the formal model specification into the runtime assertions [34]. Furthermore, no effort has been made to minimise false negatives in SC formal modelling, even though the methods have been developed for traditional systems, such as Event-B [35].

⁷<https://docs.vyperlang.org/en/latest/index.html>

⁸<https://www.idris-lang.org>

4. Proposed Solution

4.1. Outline

In light of the above, we believe there is a need for a solution that combines two approaches to allow SC developers to reason about their program in terms of FSM models that can be verified at the compile time for functional correctness and model consistency, and enable an automatic generation of invariants and constraints to validate the data at runtime.

We propose *Folidity*, a safe smart contract language. Folidity offers a model-first approach to the development process while featuring a functional-friendly programming style. The language intends to offer a safe and secure-by-design approach to the programming, ensuring the developer is aware of any state transitions during execution.

The list of feature requirements has been comprised based on the vulnerabilities described in Table 2.1.

1. **Provide abstraction over timestamp** in response to *SCV1*. We are interested in the limited use of timestamps in SCs in favour of block number or another safe primitive.
2. **Provide a safe interface for randomness** in response to *SCV2*. Folidity should also provide a source of randomness through a standardised interface.
3. **Enable model-first approach in development** in response to *SCV3*. Developers should reason about the storage in terms of models and how they are updated by events. This approach is inspired by the Event-B [35] work, which can also be applied to SC development.
4. **Explicit state checks at runtime** in response to *SCV3* and *SCV6*. Similar to *Requirement 3*, SC developers should be aware of any state transitions that update the state of the model. State transitions must happen explicitly and be validated at the runtime to guarantee *liveness*.
5. **Static typing** in response to *SCV3* and *SCV5*.
6. **Polymorphic-dependent types** in response to *SCV3*. Polymorphic-dependent types should be part of a runtime assertion check during state transition and model mutation⁹.

- 7. Role-based access** in response to *SCV4*. All message functions that mutate the model should be annotated with the role-access header specifying which set of accounts is allowed to call it.
- 8. Checked arithmetic operations** in response to *SCV5*. All arithmetic operations should be checked by default, and the developer is responsible for explicitly specifying the behaviour during over/underflow, similar to Rust.
- 9. Enforced checked recursion or bounded loops** in response to *SCV3*.

Infinite loops should not be permitted, and any loops should generally be discouraged in favour of recursion. The recursion base case should be specified explicitly with appropriate invariants. Bounded loops may be used but should be limited to list or mapping iterations.

As part of the language design, the SC building workflow is illustrated in Figure 1.

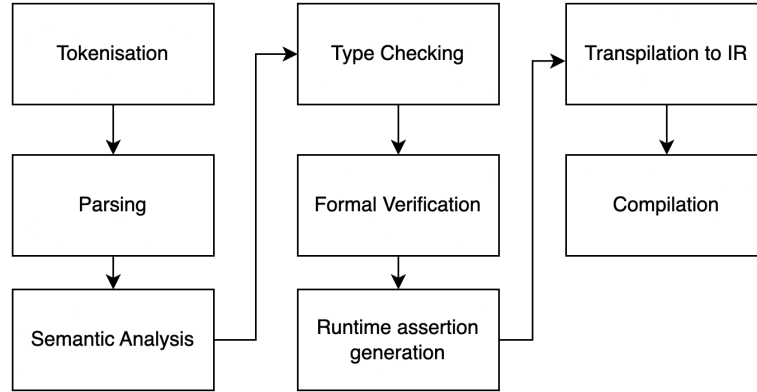


Figure 4.1: Build workflow

As one of the core features of Folidity, formal verification is part of the build process. Having verified the model consistency, invariants, and constraints, the program is considered safe to generate runtime assertions.

Another core feature is a pure computation context of the SC in Folidity. As illustrated in Figure 4.2, state mutations to the contract storage and the global state (e.g. account balances) happen independently of each other. Folidity proposes a new execution model when a portion of a global state is *embedded* into the local state of the SC as shown in Figure 4.3. *Global state* refers to the overall state of the blockchain system (e.g. account balances), whereas *local state* describes the storage of an individual SC.

⁹ *Model mutation* and *state transition* refer to the same process. They are used interchangeably

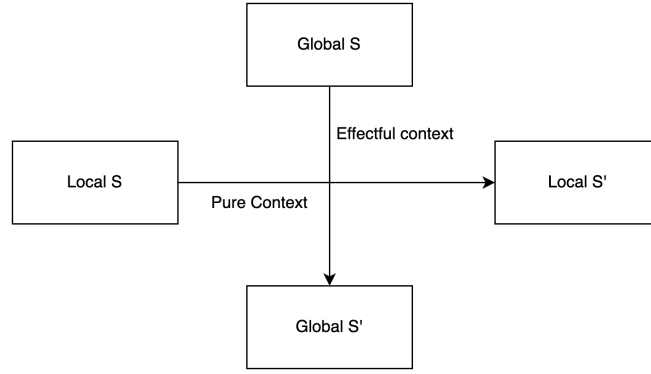


Figure 4.2: Traditional execution context

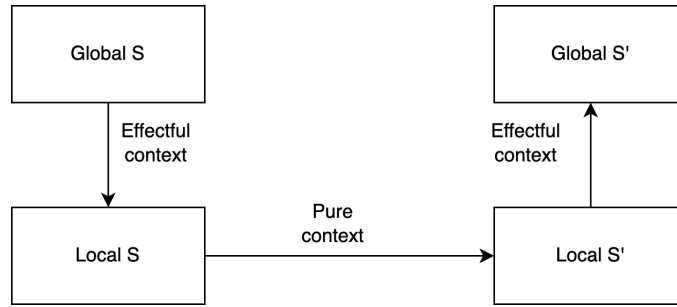


Figure 4.3: Transformed execution context

4.2. Language design

Folidity features a rich grammar that allows one to abstract away from low-level operations while providing a high level of readability and expressivity. Certain considerations have been taken into account to reflect the desired features described in Section 4.1.

Folidity is described using LR(1)¹⁰ grammar as outlined in Appendix A. One of the advantages of using LR(1) grammar is its expressiveness and clarity which allows describing sophisticated data structures. It additionally enables easier implementation of the error-recovery [36] for reporting purposes which lies at the core of the Folidity compiler.

4.2.1. Primitives, Expressions and Statements

Starting from primitives, Folidity provides numerous data types allowing encoding data for the domain of use cases in dApps:

- `int`, `uint`, `float` - signed, unsigned integers and floating-point numbers
- `()` - unit type, similar to rust this means no data.
- `string` - string literals, can be provided as `s"Hello World"`

¹⁰https://en.wikipedia.org/wiki/LR_parser

- `hex` - hexadecimal string literals, provided as `h"AB"`
- `address` - account number literal, provided as `a"<address>"`
- `list<a>`, `set<a>` - lists of elements of type `a`, `set` describes a list of unique elements.
- `mapping<a -> b>` - a mapping from type `a` to type `b` using the relation `->`
 - `->` : total relation
 - `-/>` : partial relation, can be combined with injective and surjective notations.
 - `>->` : (partial) injective relation
 - `->>` : (partial) surjective relation
 - `>->>` : bijective relation
- `char` - character, provided as `'a'`
- `bool` - boolean literals `true` or `false`

By describing the type of relations in mappings, we can combine Event-B approach of proof obligation with symbolic execution to provide strong formal guarantees of member inclusion and member relations.

Specifically, we can define some axiom where we can have a mapping of partial injective relation between addresses (`address`) and asset ids (`uint`) `assets`: `mapping<address >-/> int>`:

$$\text{Assets: Address} \rightarrow \text{Int}$$

Then, for some statement S : `assets = assets :> add(<a>,)`, we can treat as a hypothesis. The compiler can then assert:

$$S, (a', a \in \text{Address}) \vdash \text{Assets}(a) \neq \text{Assets}(a')$$

Looking at the expressions, Folidity provides a standard set of operators to describe mathematical and boolean expressions (e.g. `+`, `/`, `||`, etc.) with few additions.

- `in` - inclusion operator, return `true` if for `a in A`, the `a ∈ A` is true, if used in boolean context. Otherwise, it extracts an element from `A` and assigns it to `a` when used in iterators.
- `:>` - pipeline operator, heavily inspired by `F# |>` operator¹¹. It allows to pipe the result of one function into the other one. This enables easy integration of a functional style of programming and handling of side effects of the mathematical operations such as overflow or division by zero, hence, addressing *SCV5* and *Requirement 8*.

```
let result: int = a + 1_000_000_000 :> handle((_) -> return 0);
```

¹¹<https://learn.microsoft.com/en-gb/dotnet/fsharp/language-reference/functions/#pipelines>

Statements have close resemblances to Rust syntax by following the following syntax:

```
let <var_ident>: <optional_type> = <expr>;
```

The type can be derived at the compile time from the expression. Other simple statements are similar to Rust statements and are defined in Appendix A.

It is worth looking at the unique additions such as struct instantiation and state transition.

Any structure-like type can be instantiated using the `<ident> : { <args>, ..<object> }` syntax, where

- `<ident>` - Identifier of the declaration.
- `<args>` - list of arguments to assign fields
- `<object>` - Object to fill the rest of the fields from if not all arguments are provided.

This expression can be combined with the state transition statement to execute the explicit change in the internal state of the SC.

```
move <state_ident> : { <args>, ..<object> };
```

4.2.2. Declarations

A typical program in Folidity consists of data structures that describe models, states, and functions that can interact with each other. Models are one of the core structures that provide the model consistency guarantee in Folidity. States can encapsulate different of the same models and describe explicit state transition or state mutations as part of program execution, and functions are the driving points in program execution. Functions declare and describe the state transitions.

As mentioned multiple times before, models lie within the core of Folidity design. They resemble regular `struct` structures in “C-like” languages with few differences.

Models describe some representation of the storage layout that is encapsulated within explicit states.

```
model MyModel {
  a: int,
  b: string,
} st [
  a > 10,
  b == s"Hello World"
]
```

Listing 4.1: Simple model with constraints

Folidity provides developers with a syntax to further constraint the data that the model can accept by specifying model bounds in `st`¹² blocks. This syntax can also be used in state and function declarations as illustrated later. To support context transformation, any global state variables (e.g. block number, current caller) are injected into a model as fields and can be accessed in `st` blocks and expressions in functions. Furthermore, Folidity borrows the idea of model refinements from Event-B by allowing a model to inherit another model's fields and refine its constraints as shown in Listing 2.

```
model ParentModel {
    a: int,
} st [
    a > 10,
]

model MyModel: ParentModel {} st [
    a > 100
]
```

Listing 4.2: Model refinement

States facilitate the tracked mutation of the storage. They can encapsulate models, have raw fields in the body, or not contain any data at all. They are essentially *the* data structures that can be encoded and stored as raw bytes in the contract storage.

```
model ParentModel {
    a: int,
} st [
    a > 10,
]

state StateA(ParentModel) st [
    a < 100
]

state StateB {
    b: uint
} st [
    b < 10
]

state NoState
```

Listing 4.3: States

The idea behind model encapsulation is to enable distinct states to represent identical models with their distinct constraints. Additionally, states' bounds can further be restricted by specifying the incoming state, that is the state only from which we can transition to the specified state.

¹²States for "such that"

```
state StateA from (StateB s) st [
  s.a > 10
]
```

Listing 4.4: State transition bounds

As mentioned earlier, functions facilitate the model mutation of the Folidiy SC. Functions provide a controlled interface for the user and other contracts (non-AVM type of contracts) to interact with the business logic and the state of the application. Therefore, it is important to enable developers to control the execution flow of the incoming data and provide them with fine-grained control over output data and storage mutation.

Let's look at the signature of a typical function in Folidity;

```
@init
@(any)
fn (out: int) my_func(input: hex)
  where (InitialState s1) -> (FinalState s2)
  st [
    input != h"ABC",
    out > 10,
    out < 100
    s1.a == s2.a
  ] { <statements> }
```

Listing 4.5: Function signature

Starting from the top: `@init` is an optional attribute that indicates the function is used for instantiation of the contract. Followed by the end attribute, `@(any)`, a developer can specify who can call the contract. `any` is a wildcard variable indicating that anyone can call it. However, it is possible to specify a list of addresses or a specific address using data from the incoming state `@(s1.whitelist | a"<some_address>")`.

If no attributes are specified, then it is assumed that the function is private and internal and can only be called within the contract.

Moving on, `(out: int)` is a return type bound by the variable `out` that can be used to specify a post-execution condition to ensure that the function produces results within an acceptable model's range. It is also possible to just specify the return type, `fn int my_func(...)`. The `my_func` is an identifier of the function, followed by the list of typed parameters.

Functions in Folidity feature `where` blocks enabling developers to specify state transition bounds and are also used to inject the current state into the function's execution context. Certain functions can only be executed if the input state matches the current state. After `->` we have a final state that indicates which state we transition to, this can be the same state, meaning that the function only mutates the current state and doesn't logically advance. Both input and output states can be bound by variables in order to specify pre and post mutation con-

straints. You can notice that state's variables are declared in a different fashion from other data types. This is conscious design decision to differentiate the state mutation parts from the traditional manipulation of primitive data.

4.3. Formal Verification

Formal verification is one of the unique features of Folidity. The grammar is structured with first-class support for formal verification in mind. Therefore, the compiler can imply and prove certain mathematical and functional properties of the program directly from the code without the need to perform any context translations like its done in the aforementioned solutions.

This chapter illustrates a couple of examples how model consistency and constraint satisfiability can be directly proven directly from the source code of a typical Folidity program.

4.3.1. Model consistency

As an example of the theory behind model consistency in SCs, let's look at role-based access. Suppose:

- $*$ = {All addresses}
- M = {Moderators of the system}
- A = {Admins of the system}

Then, we can model a role-based access hierarchy as

$$A \subseteq M \subseteq *$$

Subsequently, given some event for the system `add_mod(a: Address)`, we can define the following invariants for the system:

$$\begin{aligned} i_0 &:= \text{card}(A) = 1 \\ i_2 &:= \text{card}(B) = 5 \end{aligned}$$

And the invariant for the event:

$$i_2 := c \in A$$

Where

- c - caller's address
- i_n - enumerated invariant with some boolean statement
- $\text{card}(\dots)$ - cardinality of a set

For the denoted event, suppose we mutate the model by adding an address to a set of admins: $A : A \cup \{a\}$

Then, we can verify the model consistency for some state transition from an initial state S to a new state S' , $S \rightarrow S'$, using propositional logic.

$$\frac{(i_0 \wedge i_1 \wedge i_2) \rightarrow A \cup \{a\}, a \in *, c \in A}{A \cup \{a\}}$$

However, as it can be seen, one of the premises violates the invariant, in particular:

$$\frac{\text{card}(A) = 1 \rightarrow A \cup \{a\}, a \in *}{A \cup \{a\}}$$

In practice, the following error can be picked at the compile time by using symbolic execution of the code. The other invariant, i_2 , can be picked at the runtime by generating an appropriate assertion.

4.3.2. Proving constraint satisfiability

One of the core pieces in the workflow aforementioned is the model bounds that consist of individual boolean constraints as shown in Listing 6. Let's break down how each of the selected techniques can be applied to the program written in Folidity. As a good starting point, we can perform a static analysis and verify that the program statements, declarations and constraints are valid and consistent.

A simple approach is to perform semantic analysis that carries out type checking and verification of correct state transition in the function body. Specifically, if `mutate()` expects to return `StateA`, but instead it performs a state transition to `StateB` we can already detect this inconsistency at a compile time.

The next stage of the analysis involves verification of the consistency of the models described.

```

# Some model and its constraints
model ModelA {
  x: int,
  y: int
} st [
  x > 10,
  y < 5
]
# A state that encapsulates a model and provides its own constraints.
state StateA(ModelA) st [
  x < y
]
# A function that describes mutation.
fn () mutate(value: int) when (StateA s) -> StateA
st [
  value > 100,
  value < 100,
] { ... }

```

Listing 4.6: Simple folidity program

We can generalise the approach using the following mathematical model. Let's describe some verification system **VS** as

$VS = \langle \mathbf{M}, \mathbf{E}, \Upsilon, \Theta, T_M, T_{E, \{E, M\}}, T_{\Upsilon, E} \rangle$ where

- \mathbf{M} - set of models in the system.
- \mathbf{E} - set of states in the system
- Υ - set of functions in the system.
- Θ - set of of constraint blocks in the system, where $\Theta[\mathbf{M}]$ corresponds to the set of constraints for models, $\Theta[\mathbf{E}]$ - state constraints and $\Theta[\Upsilon]$ function constraints.
- T_M - a relation $T : \mathbf{M} \rightarrow \mathbf{M}$ describing a model inheritance.
- $T_{E, \{E, M\}}$ - a relation $T : \mathbf{E} \rightarrow \{\mathbf{E}, \mathbf{M}\}$ describing any state transition bounds and encapsulated models in states, that is some state S' can only be transitioned to from the specified state S , and state some state S can encapsulate some model M
- $T_{\Upsilon, E}$ - a relation $T : \Upsilon \rightarrow \mathbf{E}$ describing any state transition bounds for states \mathbf{E} in functions Υ

In particular, $\forall \mu \in \mathbf{M} \exists \theta \in \Theta[\mu]$ where θ is a set of constraints for μ , and corresponding logic can be applied for elements of \mathbf{E} and Υ .

Then, to verify the consistency of the system, we first need to verify the following satisfiability *Sat*:

$$\begin{aligned}
& \forall \mu \in \mathbf{M} \\
& \exists \theta \in \Theta[\mu] \\
& \text{s.t. } \theta = \{c_0, c_1, \dots, c_k\} \\
& \left(\bigwedge_i c_i \right) \Rightarrow \text{Sat}
\end{aligned}$$

We can define the following check by some functions $\rho(\theta) : \Theta \rightarrow \{\text{Sat}, \text{Unsat}\}$ which yields the following proof:

$$\begin{aligned}
& \exists \theta \in \Theta[e] \\
& \text{s.t. } \theta = \{c_0, c_1, \dots, c_k\} \\
& \left(\bigwedge_i c_i \right) \Rightarrow \text{Sat or Unsat}
\end{aligned}$$

This allows to validate the next property of **VS**

$$\begin{aligned}
A &= \{\mathbf{M} \cup \mathbf{E} \cup \Upsilon\} \\
A &= \{e_0, e_1, \dots, e_k\} \\
\left(\bigwedge_i \rho(\Theta[e_i]) \right) &\Rightarrow \text{Sat or Unsat}
\end{aligned}$$

The next stage is to verify co-dependent symbols in the system for satisfiability of their respective constraints.

Let's look at the models **M**, we want to ensure that

$$\begin{aligned}
& \text{if for some } m \in \mathbf{M}, m' \in \mathbf{M} \\
& \exists (m, m') \in \mathbf{T}_{\mathbf{M}} \\
& \text{s.t. } \rho(m) \times \rho(m') = (\text{Sat}, \text{Sat}) \\
& \text{and } \theta = \Theta[m] \cup \Theta[m'] \\
& \rho(\theta) \Rightarrow \text{Sat}
\end{aligned}$$

Very similar verification can applied to $\mathbf{T}_{\Upsilon, \mathbf{E}}$.

For $\mathbf{T}_{\mathbf{E}, \{\mathbf{E}, \mathbf{M}\}}$, the constraints can be extracted in the following way:

$$\begin{aligned}
& \text{if for some } \varepsilon \in \mathbf{E}, \varepsilon' \in \mathbf{E} \\
& \exists (\varepsilon, \varepsilon') \in \mathbf{T}_{\mathbf{E}, \{\mathbf{E}, \mathbf{M}\}} \\
& \text{s.t. } \rho(\varepsilon) \times \rho(\varepsilon') \times \rho(\mu) = (\text{Sat}, \text{Sat}) \\
& \text{and } \theta = \Theta[\varepsilon] \cup \Theta[\varepsilon'] \\
& \rho(\theta) \Rightarrow \text{Sat}
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \text{if for some } \varepsilon \in E, \mu \in M \\
& \quad \exists(\varepsilon, \mu) \in T_{E, \{E, M\}} \\
& \text{s.t. } \rho(\varepsilon) \times \rho(\mu) = (Sat, Sat) \\
& \quad \text{and } \theta = \Theta[\varepsilon] \cup \Theta[\mu] \\
& \quad \rho(\theta) \Rightarrow Sat
\end{aligned}$$

After the completing verification of T relations for consistency, we can provide a mathematical guarantee that **VS** has been modelled consistently.

Having verified the constraints, we can leverage them as the guards during state transitions and can apply proofs from *temporal logic* to verify that the described state transitions will take place under the described constraints.

In the final stage, we can perform the symbolic execution of instructions in the function bodies with the constraints loaded in the global context of the system. Having tracked the states of different symbols, we can verify each function for reachability for described state transitions and provide strong guarantees of functional correctness of the system described in the smart contract.

4.3.3. Other techniques

The above examples leverage the static analysis of the program to derive its mathematical properties. It is worth noting that it is possible to apply other techniques of formal verification such as symbolic execution and interface discovery.

In particular, we can provide even more fine-grained validation of the program by asserting user-defined constraints in the symbolic execution context. This enables to detect unsatisfiability of reachability at the earlier stage of the execution. The traditional methods rely on composing these constraints at the runtime through the statistical discovery of the model bounds whereas Folidity offers this information at the compile time.

In the context of mutli-contract execution, which applies to EVM-compatible blockchains. Instead of carrying out interface discovery through statistical methods, we can potentially encode the function signature with its models bounds and constraints into the metadata and leverage this information at the runtime in order to verify the model consistency and constraint satisfiability as illustrated earlier.

5. Implementation

5.1. Outline

The language is implemented using Rust¹³ due to its memory-safety guarantees and efficiency. The compiler uses Lalrpop¹⁴ parser-generator to streamline the development process. Folidity also requires SMT-solver for formal verification and generation of runtime assertions. In order to facilitate this functionality, Z3¹⁵ will be used since it also provides Rust bindings. It was debated to use Boogie, since it provides a higher-level abstraction, but it was quickly discarded due to lack of documentation and increased development time.

As a target blockchain for the language, Algorand¹⁶ has been selected. Algorand is a decentralized blockchain platform designed for high-performance and low-cost transactions, utilising a unique consensus algorithm called Pure Proof-of-Stake to achieve scalability, security, and decentralisation [37].

One of the potential drawbacks of Folidity is a computational overhead due to complex abstractions and additional assertions. EVM-based blockchains have varying costs for the execution, i.e. fees, that depend on the complexity of a SC. On the contrary, although Algorand has a limited execution stack, it offers fixed, low transaction fees. Additionally, Algorand execution context explicitly operates in terms of state transition, which perfectly suits the paradigm of Folidity. Finally, Algorand offers opt-in functionality and local wallet storage, allowing users to explicitly opt-in to use the SC. This provides additional support in the role-based access control in Folidity.

The Folidity compiler emits Algoran AVM Teal¹⁷ bytecode. It was originally planned to emit an intermediate representation in Tealish¹⁸. However, this option was soon invalidated due to reduced developer activity in the project and the absence of audits that may compromise the intrinsic security of the Folidity compiler.

Overall, the compilation workflow can be summarised in Figure 4

¹³<https://www.rust-lang.org>

¹⁴<https://github.com/lalrpop/lalrpop>

¹⁵<https://microsoft.github.io/z3guide>

¹⁶<https://developer.algorand.org>

¹⁷<https://developer.algorand.org/docs/get-details/dapps/avm/teal/>

¹⁸<https://tealish.tinyman.org/en/latest>

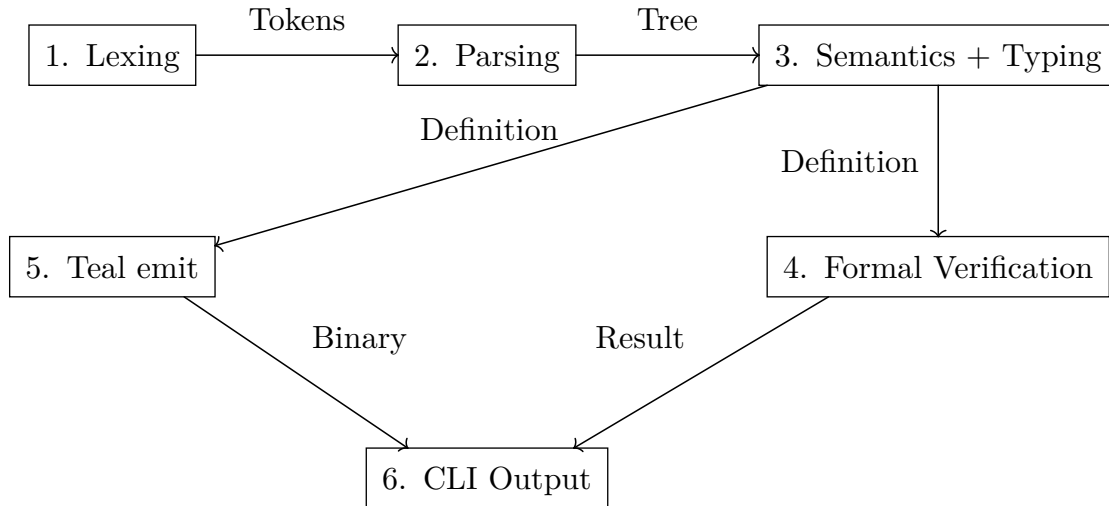


Figure 5.4: Compilation process

Step logics are composed into Rust crates (i.e. modules) for modularity and testability.

Steps 1 and 2 are processed by the `folidity-parser` crate. They produce a syntax AST, which is fed to the `folidity-semantics` for semantic analysis and type checking (step 3). The resulting contract definition is then independently piped into the `folidity-verifier` crate for formal verification (step 4) and the `folidity-emitter` for the final build compilation of binary Teal code (step 5). The artifacts of the compilation and the result of verification are then supplied back to the calling the CLI crate (`folidity`) to display the result to the user and write artifacts into the file (step 6).

5.2. Scope

As part of the development process, it has been decided to limit the scope to supporting only a single SC execution. Cross-contract calls require extra consideration in design and development. Therefore, *SCV6* is only addressed in the theoretical context of this paper. Additionally, optimisation of the execution is not considered relevant at this stage in favour of safety and simplicity. Finally, Algorand offers smart signatures, a program that is delegated a signing authority¹⁹. As they operate in a different way from SCs, they are also outside the scope of this project.

¹⁹<https://developer.algorand.org/docs/get-details/dapps/smart-contracts/smartsigs>

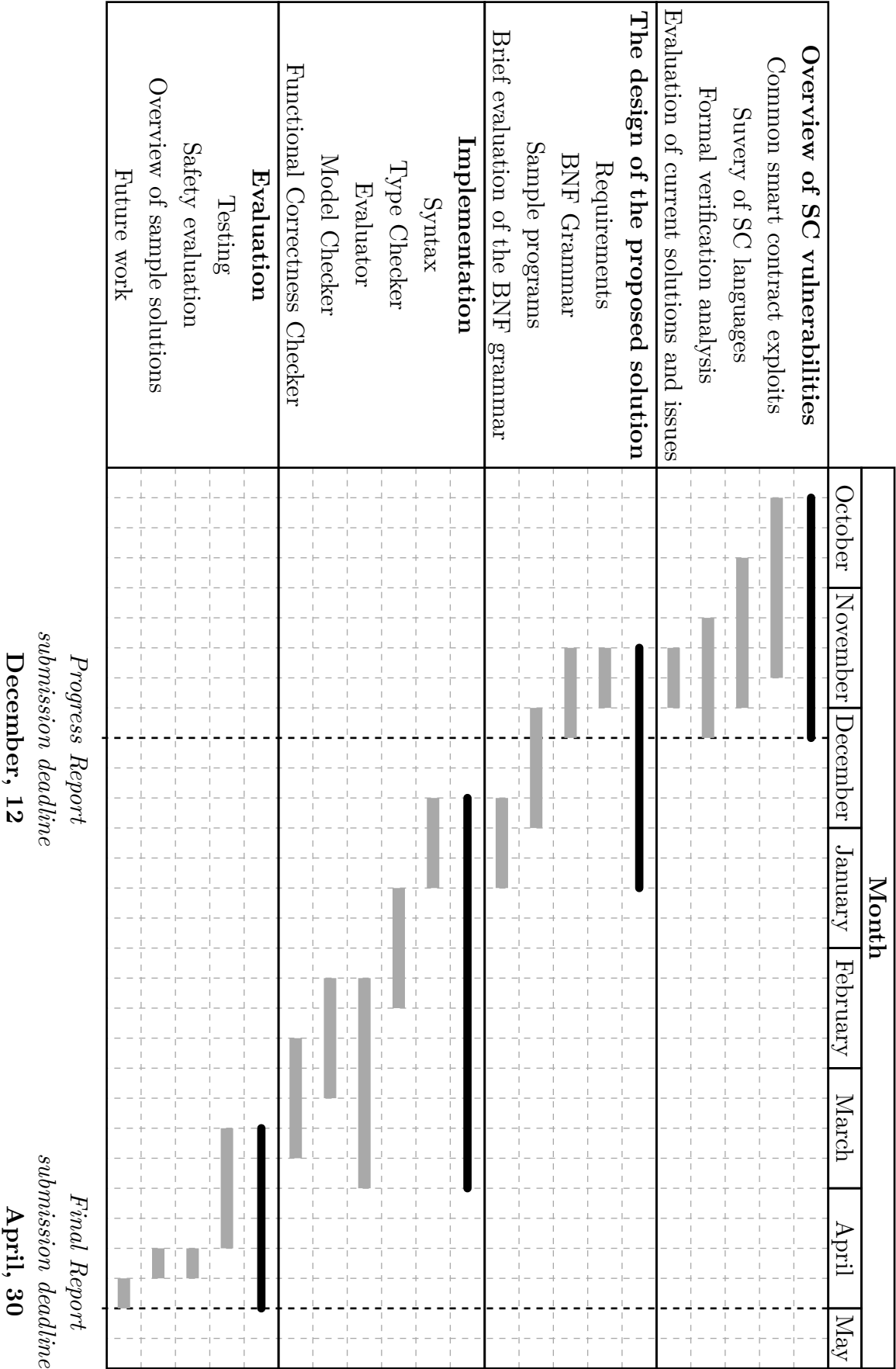
6. Project Planning

A significant groundwork in research of current solutions and their limitations has been done as illustrated by Gantt chart in Appendix B. Since the requirements have been collected, some progress has been made in the design of BNF grammar that will later pave the way for the development of the parser. It is still possible to research more formal verification methods during the grammar design.

From the beginning of January, the first iteration of grammar should be completed, and the active development of the type checker and formal verifier should begin.

A. Folidity Grammar

B. Gannt Chart



Bibliography

- [1] N. Szabo, “Smart Contracts,” in *Nick Szabo’s Papers and Concise Tutorials*, 1994. [Online]. Available: <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html>
- [2] N. Szabo, “Smart Contracts: Building Blocks for Digital Markets,” in *Nick Szabo’s Papers and Concise Tutorials*, 1996. [Online]. Available: <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html>
- [3] D. G. Wood, “ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER.” [Online]. Available: <https://gavwood.com/paper.pdf>
- [4] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System.” 2008.
- [5] E. Foundation, “Solidity.” [Online]. Available: <https://docs.soliditylang.org/en/v0.8.23>
- [6] S. D. Levi, A. B. Lipton, Skadden, Arps, Slate, and M. & Flom LLP, “An Introduction to Smart Contracts and Their Potential and Inherent Limitations.” [Online]. Available: <https://corpgov.law.harvard.edu/2018/05/26/an-introduction-to-smart-contracts-and-their-potential-and-inherent-limitations>
- [7] T. Dounas and D. Lombardi, “A CAD-Blockchain Integration Strategy for Distributed Validated Digital Design Connecting the Blockchain,” 2018, p. . doi: 10.52842/conf.ecaade.2018.1.223.
- [8] X. Zhao, Z. Chen, X. Chen, Y. Wang, and C. Tang, “The DAO attack paradoxes in propositional logic,” in *2017 4th International Conference on Systems and Informatics (ICSAI)*, 2017, pp. 1743–1746. doi: 10.1109/ICSAI.2017.8248566.
- [9] “King of the Ether: Post-Mortem Investigation.” [Online]. Available: <https://www.kingoftheether.com/postmortem.html>
- [10] H. Gall, “Functional safety IEC 61508 / IEC 61511 the impact to certification and the user,” 2008, pp. 1027–1031. doi: 10.1109/AICCSA.2008.4493673.

- [11] B. C. Pierce, *Types and Programming Languages*. The MIT Press, 2002.
- [12] R. van Glabbeek, “Ensuring liveness properties of distributed systems: Open problems,” *Journal of Logical and Algebraic Methods in Programming*, vol. 109, p. 100480–100481, 2019, doi: <https://doi.org/10.1016/j.jlamp.2019.100480>.
- [13] C. C4, “Ondo Finance. Findings and Analysis Report.” [Online]. Available: <https://code4rena.com/reports/2023-09-ondo>
- [14] J. V. Behanan, “OWASP Smart Contract Top 10.” [Online]. Available: <https://owasp.org/www-project-smart-contract-top-10/>
- [15] S. D. Angelis, F. Lombardi, G. Zanfino, L. Aniello, and V. Sassone, “Security and dependability analysis of blockchain systems in partially synchronous networks with Byzantine faults.” [Online]. Available: <https://doi.org/10.1080/17445760.2023.2272777>
- [16] N. Atzei, M. Bartoletti, and T. Cimoli, “A Survey of Attacks on Ethereum Smart Contracts SoK,” in *Proceedings of the 6th International Conference on Principles of Security and Trust - Volume 10204*, Berlin, Heidelberg: Springer-Verlag, 2017, pp. 164–186. doi: 10.1007/978-3-662-54455-6_8.
- [17] C. C4, “Arcade.xyz. Findings and Analysis Report.” [Online]. Available: <https://code4rena.com/reports/2023-07-arcade>
- [18] C. C4, “Centrifuge. Findings and Analysis Report.” [Online]. Available: <https://code4rena.com/reports/2023-09-centrifuge>
- [19] E. Foundation, “Formal Verification of Smart Contracts.” [Online]. Available: <https://ethereum.org/en/developers/docs/smart-contracts/formal-verification/>
- [20] A. Souri, A. M. Rahmani, N. J. Navimipour, and R. Rezaei, “A symbolic model checking approach in formal verification of distributed systems,” *Human-centric Computing and Information Sciences*, vol. 9, no. 1, p. 4–5, doi: 10.1186/s13673-019-0165-x.
- [21] J. Harrison, J. Urban, and F. Wiedijk, “History of Interactive Theorem Proving,” in *Computational Logic*, 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:30345151>
- [22] Hildenbrandt *et al.*, “KEVM: A Complete Formal Semantics of the Ethereum Virtual Machine,” in *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, 2018, pp. 204–217. doi: 10.1109/CSF.2018.00022.
- [23] F. Cassez, J. Fuller, and H. M. A. Quiles, “Deductive Verification of Smart Contracts with Dafny.” 2022.

Total words: 4964

- [24] S. So, M. Lee, J. Park, H. Lee, and H. Oh, “VeriSmart: A Highly Precise Safety Verifier for Ethereum Smart Contracts.” 2019.
- [25] Y. Wang *et al.*, “Formal Specification and Verification of Smart Contracts for Azure Blockchain.” 2019.
- [26] B. Beckert, T. Bormer, F. Merz, and C. Sinz, “Integration of Bounded Model Checking and Deductive Verification,” in *Formal Verification of Object-Oriented Software*, B. Beckert, F. Damiani, and D. Gurov, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 86–104.
- [27] M. Nguyen, D. Stoffel, M. Wedler, and W. Kunz, “Transition-by-transition FSM traversal for reachability analysis in bounded model checking,” in *ICCAD-2005. IEEE/ACM International Conference on Computer-Aided Design, 2005.*, 2005, pp. 1068–1075. doi: 10.1109/ICCAD.2005.1560219.
- [28] P. Antonino and A. W. Roscoe, “Solidifier: Bounded Model Checking Solidity Using Lazy Contract Deployment and Precise Memory Modelling,” in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, in SAC '21. Virtual Event, Republic of Korea: Association for Computing Machinery, 2021, pp. 1788–1797. doi: 10.1145/3412841.3442051.
- [29] K. R. M. Leino, “This is Boogie 2,” Jun. 2008.
- [30] M. H. Majd Soud Gísli Hjálmtýsson, “Dissecting Smart Contract Languages: A Survey.” [Online]. Available: arxiv:2310.02799v2
- [31] Chainsecurity, “Curve LP Oracle Manipulation: Post Mortem.” [Online]. Available: <https://chainsecurity.com/curve-lp-oracle-manipulation-post-mortem/>
- [32] F. Schrans, D. Hails, A. Harkness, S. Drossopoulou, and S. Eisenbach, “Flint for Safer Smart Contracts.” 2019.
- [33] J. Pettersson and R. Edström, “Safer smart contracts through type-driven development.” 2016.
- [34] F. Maurica, D. R. Cok, and J. Signoles, “Runtime Assertion Checking and Static Verification: Collaborative Partners,” in *Leveraging Applications of Formal Methods, Verification and Validation. Verification*, T. Margaria and B. Steffen, Eds., Cham: Springer International Publishing, 2018, pp. 75–91.
- [35] T. S. Hoang, “An Introduction to the Event-B Modelling Method,” 2013, pp. 211–236.
- [36] A. Sorkin and P. Donovan, “LR(1) Parser Generation System: LR(1) Error Recovery, Oracles, and Generic Tokens.” 2012.

Total words: 4964

- [37] J. Chen and S. Micali, “Algorand.” 2017.