ConFuzzius: A Data Dependency-Aware Hybrid Fuzzer for Smart Contracts

Christof Ferreira Torres*, Antonio Ken Iannillo*, Arthur Gervais†, and Radu State*
*SnT, University of Luxembourg
†Imperial College London

***Abstract*—Smart contracts are Turing-complete programs that are executed across a blockchain. Unlike traditional programs, once deployed, they cannot be modified. As smart contracts carry more value, they become more of an exciting target for attackers. Over the last years, they suffered from exploits costing millions of dollars due to simple programming mistakes. As a result, a variety of tools for detecting bugs have been proposed. Most of these tools rely on symbolic execution, which may yield false positives due to over-approximation. Recently, many fuzzers have been proposed to detect bugs in smart contracts. However, these tend to be more effective in finding shallow bugs and less effective in finding bugs that lie deep in the execution, therefore achieving low code coverage and many false negatives. An alternative that has proven to achieve good results in traditional programs is hybrid fuzzing, a combination of symbolic execution and fuzzing.**

In this work, we study hybrid fuzzing on smart contracts and present ConFuzzius, the first hybrid fuzzer for smart contracts. ConFuzzius uses evolutionary fuzzing to exercise shallow parts of a smart contract and constraint solving to generate inputs that satisfy complex conditions that prevent evolutionary fuzzing from exploring deeper parts. Moreover, ConFuzzius leverages dynamic data dependency analysis to efficiently generate sequences of transactions that are more likely to result in contract states in which bugs may be hidden. We evaluate the effectiveness of ConFuzzius by comparing it with state-of-the-art symbolic execution tools and fuzzers for smart contracts. Our evaluation on a curated dataset of 128 contracts and a dataset of 21K real-world contracts shows that our hybrid approach detects more bugs than state-of-the-art tools (up to 23%) and that it outperforms existing tools in terms of code coverage (up to 69%). We also demonstrate that data dependency analysis can boost bug detection up to 18%.

***Index Terms*—Ethereum, smart contracts, hybrid fuzzing, data dependency analysis, genetic algorithm, symbolic execution**

I. Introduction

The inception of immutable, blockchain-based smart contracts has shown how to enable multiple mistrusting parties to trade and interact without relying on a centralized, trusted third party. The immutability of a contract is crucial: if at least one of the engaging parties were allowed to modify a digital contract, the contract's trust would vanish. Unlike traditional legal contracts, smart contracts do not allow a dispute resolution with a neutral third party. Most importantly, smart contracts cannot be nullified — parties cannot revoke any deployed smart contract, even if its code figures undeniable software bugs. Therefore, this very immutability comes at a price: smart contracts must be tested extensively before exposing them and their users to significant monetary value. In the past, simple vulnerabilities (e.g. missing access control [36]) and subtle vulnerabilities (e.g. reentrancy [40]) have led to losses exceeding many tens of millions of USD.

We can verify the behavior of a smart contract with four different approaches. (i) *Unit testing* requires manual effort to cover the different sections of the code, but it unveils only a limited number of bugs within the test cases. (ii) *Symbolic execution* analyzes contract behavior abstractly but performs slowly on complex contracts (path explosion problem). (iii) *Static analysis* does not execute code and over-approximates the contract behavior — it can capture the entire contract execution surface, but it exhibits false positives that must be manually inspected. Finally, (iv) *fuzzing* tests a contract reasonably fast by automatically generating test cases, with a generally lower false positive rate than static analysis. Fuzzing, however, can suffer from low code coverage, especially when input is fuzzed at random and hence does not overcome simple input sanity verification.

When fuzzing smart contracts, we face the following three challenges: 1) *input generation*, 2) *stateful exploration*, and 3) *environmental dependencies*. When it comes to input generation, the input space can be significantly broad. However, the solution might be limited to a specific point. For example, if a condition requires an input value of type `uint256` to equate to 42, then the probability of randomly generating 42 as input is tremendously small. Moreover, smart contracts are stateful applications, i.e. the execution may depend on a state that is only achievable following a specific sequence of inputs. Finally, the runtime environment of smart contracts exposes additional inputs related to the underlying blockchain protocol, such as the current block timestamp or other contracts deployed on the blockchain. As a result, the execution flow of smart contracts may depend on environmental information besides transactional information.

We solve these three challenges as follows. In tandem with the fuzzing procedure, we employ symbolic taint analysis to generate path constraints on tainted inputs. Once we detect that the fuzzer is not progressing, we activate a constraint solver to solve the constraint in question. We collect this solution within a mutation pool, from which the fuzzer can draw to move past the challenging contract condition. Existing hybrid fuzzing approaches, e.g. Driller [43], cease the fuzzer when they are stuck and switch to concolic execution to get past the complex condition. Then, they restart the fuzzer once passed the condition. Our approach keeps the fuzzer running and only uses constraint solving to generate inputs on the fly, which will

eventually be picked by the fuzzer via the mutation pools. In addition to constraint solving, we perform a path termination analysis to purge irrelevant inputs from the mutation pools. To deal with the statefulness of smart contracts, we chose to take advantage of the selection and crossover operators of genetic algorithms. Genetic algorithms follow three main steps: selection, crossover, and mutation. The selection operator's task is to choose two individuals from the population, which are afterwards combined by the crossover operator to create two new individuals. The challenge here is to generate meaningful combinations of inputs. Therefore, data dependencies between individuals guide our selection and crossover operators that accept two individuals only if they follow a *read-after-write* (RAW) data dependency. Finally, to solve the third and last challenge, we instrument the execution environment (i.e. the Ethereum Virtual Machine) to fuzz environmental information and model the input to a contract as a tuple consisting of transactional *and* environmental data.

Contributions. Our main contributions are as follows:

- To the best of our knowledge, we propose the first design of a hybrid fuzzer for smart contracts.
- We present a novel method to efficiently create meaningful sequences of inputs at runtime by leveraging dynamic data dependencies between state variables.
- We introduce CONFUZZIUS, the first implementation of a hybrid fuzzer for smart contracts.
- We evaluate CONFUZZIUS on a set of 128 curated smart contracts as well as 21K real-world smart contracts, and demonstrate that our approach not only detects more vulnerabilities (up to 23%) but also achieves more code coverage (up to 69%) than existing symbolic execution tools and fuzzers.

II. BACKGROUND

This section provides the required background on Ethereum smart contracts and fuzzing in order to better understand the approach proposed in this work.

A. Ethereum Smart Contracts

Smart Contracts. Ethereum [49] enables the execution of so-called *smart contracts*. These are fully-fledged programs stored and executed across the Ethereum blockchain, a network of mutually distrusting nodes. Ethereum supports two types of accounts, externally owned accounts (i.e. user accounts) and contract accounts (i.e. smart contracts). Smart contracts are different from traditional programs in many ways. They own a balance and are identifiable via a 160-bit address. They are developed using a dedicated high-level programming language, such as Solidity [48], that compiles into low-level bytecode. This bytecode gets interpreted by the Ethereum Virtual Machine. By default, smart contracts cannot be removed or updated once deployed. It is the task of the developer to implement these capabilities before deployment. The deployment of smart contracts and the execution of smart contract functions occurs via transactions. The data field of a transaction includes both, the name of the function to be executed and its arguments. Transactions are created by user accounts and afterwards broadcast to the network. They contain a sender and a recipient. The latter can be the address of a user account or a contract account. Besides carrying data, transactions may also carry a monetary value in the form of ether (Ethereum's own cryptocurrency).

Ethereum Virtual Machine. The Ethereum Virtual Machine (EVM) is a purely stack-based, register-less virtual machine that supports a Turing-complete set of instructions. Although the instruction set allows for Turing-complete programs, the instructions' capabilities are limited to the sole manipulation of the blockchain's state. The instruction set provides a variety of operations, ranging from generic operations, such as arithmetic operations or control-flow statements, to more specific ones, such as the modification of a contract's storage or the querying of properties related to the transaction (e.g. sender) or the current blockchain state (e.g. block number). Ethereum uses a *gas* mechanism to assure the termination of contracts and prevent denial-of-service attacks. The gas mechanism associates costs to the execution of every single instruction. When issuing a transaction, the sender specifies how much gas they are willing to spend to execute the smart contract. This amount is known as the *gas limit*.

B. Fuzzing

Evolutionary Fuzzing. Fuzzing, or fuzz testing, is an automated software testing technique that finds vulnerabilities in programs by feeding malformed or unexpected data as input to programs, executing them, and monitoring the effects. Evolutionary fuzzing aims at converging towards the discovery of vulnerabilities by using a genetic algorithm (GA). A generation of test cases is defined as a *population*, whereas a single test case is an *individual*. In short, every individual of a generation is evaluated based on a fitness function. At the end of each generation, solely the fittest individuals are allowed to breed, following Darwin's idea of natural selection, or "survival of the fittest". Eventually, the individuals will trigger vulnerabilities while converging towards an optimal solution. We briefly describe the main steps of a GA (see Algorithm 1). We start by creating an initial population of individuals, either generated at random or seeded via heuristics, and compute their fitness values (*line 1*). Based on the fitness value, we select two individuals from the current population, which act as parents for breading (*line 5*). Then, we apply crossover and mutation operators on the parents to generate two new individuals, also denoted as offsprings (*lines 6-7*). The generation of new individuals continues until the new population reaches the same size as the current one (*line 4*). Finally, the new population's fitness values are computed, and we replace the current population with the new population (*lines 9-10*). This entire process is repeated until a termination condition is met (*line 3*), e.g. the maximum number of generations is reached or a maximum amount of time has passed.

Algorithm 1 Pseudo-Code of a Genetic Algorithm

```
1: Create initial population and compute its fitness
2: Set initial population as current population
3: while termination condition is not met do
4:     while new population < current population do
5:         Select two parents from current population
6:         Recombine parents to create two new offsprings
7:         Mutate offsprings and add them to new population
8:     end while
9:     Compute fitness of new population
10:    Replace current population with new population
11:    Create a new empty population
12: end while
```

Hybrid Fuzzing. Although fuzzing is one of the most effective approaches to find vulnerabilities, it often has difficulties in getting past complex path conditions, resulting in low code coverage. A popular alternative to fuzzing is symbolic execution. It works by abstractly executing a program and supplying abstract symbols rather than actual (*concrete*) values. The execution will then generate symbolic formulas over the input symbols, which can be solved by a constraint solver to prove satisfiability and produce concrete values. In theory, symbolic execution is capable of discovering and exploring all potential paths in a program. However, in practice, symbolic execution is often not scalable since the number of explorable paths becomes exponential in more extensive programs (*path explosion problem*). Another limitation of symbolic execution is the limited capability to interact with the execution environment. Programs often interact with their environment by performing calls to libraries, for example. Correctly modeling these calls and other environmental information is extremely challenging. The goal of hybrid fuzzing, or hybrid fuzz testing, is to take advantage of both worlds. Hybrid fuzzing starts by performing traditional fuzzing until it saturates, *i.e.*, the fuzzer is not capable of covering any new code after running some predetermined number of steps. Hybrid fuzzing then automatically switches to symbolic execution to perform an exhaustive search for uncovered branching conditions. As soon as the symbolic execution finds an uncovered branching condition, it solves it, and the hybrid fuzzer reverts to fuzzing. The interleaving of fuzzing and symbolic execution counts on shallow program paths' quick execution via fuzzing and the execution of complex program paths via symbolic execution.

III. Overview

This section discusses the three main challenges of fuzzing smart contracts via a motivating example and presents our solution towards solving these challenges.

A. Motivating Example

Suppose a user participated in an initial coin offering (ICO) on the blockchain and now owns a number of tokens. Now assume the user wants to sell a certain amount of their tokens at a variable price that increases 1 ether per day. Fig. 1 shows a possible implementation of an Ethereum smart contract using Solidity. The smart contract allows a user to sell its tokens to an arbitrary user on the Ethereum blockchain. The contract sells the tokens to the first buyer willing to pay 42 ether, plus 1 ether for each day since the start of the sale. Moreover, the token sale should last no longer than 30 days. In this example, the smart contract acts as a simple mediator that automatically settles the trade between the user owning the tokens and the user willing to buy the tokens without both users requiring to know or trust each other. Smart contract based ICOs often follow a standard that is known as ERC-20 [15]. This standard provides an interface that standardizes function names, parameters, and return values. For example, the standard includes a function called `transferFrom`, which allows a user to transfer a limited amount of tokens to an arbitrary user on behalf of the owning user. Another example is the function `allowance`, which returns the number of tokens that a user can spend on behalf of the owning user. The smart contract in Fig. 1 works as follows. An arbitrary user can call

```
1  interface Token {
2    function transferFrom(address sender, address
         recipient, uint256 amount) external
         returns (bool);
3    function allowance(address owner, address
         spender) external view returns (uint256);
4  }
5
6  contract TokenSale {
7    uint256 start = now;
8    uint256 end = now + 30 days;
9    address wallet = 0xcafebabe...;
10   Token token = Token(0x12345678...);
11
12   address owner;
13   bool sold;
14
15   function Tokensale() public {
16     owner = msg.sender;
17   }
18
19   function buy() public payable {
20     require(now < end);
21     require(msg.value == 42 ether + (now - start)
          / 60 / 60 / 24 * 1 ether);
22     require(token.transferFrom(this, msg.sender,
          token.allowance(wallet, this)));
23     sold = true;
24   }
25
26   function withdraw() public {
27     require(msg.sender == owner);
28     require(now >= end);
29     require(sold);
30     owner.transfer(address(this).balance);
31   }
32 }
33
```

Fig. 1. Example of a vulnerable token sale smart contract. Lines highlighted in red represent complex conditions, whereas lines highlighted in gray illustrate read-after-write data dependencies and finally, lines highlighted in blue depict environmental dependencies.

the function buy to purchase the tokens for 42 ether and a fee of 1 ether for the number of days passed since the launch of the token sale. The contract will automatically transfer the tokens by calling the function transferFrom on the ICO's contract. After the purchase, the smart contract owner can call the function withdraw to retrieve the 42 ether and the fee of the purchase.

The contract contains two vulnerabilities, one known as *block dependency* and another known as *leaking ether*. The latter is enabled via a bug in the function Tokensale (see line 15 in Fig. 1). Before Solidity version 0.4.22, the only way of defining a constructor was to create a function with the same name as the contract. The function Tokensale is supposed to be the constructor of the contract TokenSale. Due to a typo, the names do not match, and the compiler does not consider the function as the contract's constructor. As a result, the function Tokensale is considered a public function that any user on the blockchain can call. This type of programming mistake has led to multiple attacks in the past [1]. The first vulnerability, namely block dependency, occurs when the transfer of ether depends on block information, such as the timestamp (see line 28 in Fig. 1). Malicious miners can alter the timestamp of blocks that they mine, especially if they can gain advantages. Although miners cannot set the timestamp smaller than the previous one, nor can they set the timestamp too far ahead in the future, developers should still refrain from writing contracts where the transfer of ether depends on block information. The second vulnerability, namely leaking ether, occurs whenever a contract allows an arbitrary user to transfer ether, despite having never transferred ether to the contract before. The following sequence of transactions triggers both vulnerabilities:

- t_0: A non-malicious user calls the function buy with a value equals to 42 ether + fee;
- t_1: An attacker calls the function Tokensale;
- t_2: The same attacker calls the function withdraw after 30 days.

When running the above example using ILF [18] (a state-of-the-art smart contract fuzzer), it is not capable of finding the two vulnerabilities even after 1 hour. Inspecting the code coverage reveals that ILF achieves only 39%. For comparison, CONFUZZIUS achieves roughly 95% code coverage and correctly identifies the two vulnerabilities in less than 10 seconds.

B. Input Generation

Generating meaningful inputs is crucial for automated software testing. Fuzzers generate inputs in order to execute not-yet-executed code. This generation can be completely random (black-box fuzzers) or driven by runtime information (grey-box fuzzers). In both cases, the primary approach is to mutate previous inputs to generate new inputs to test. Thus, finding the right heuristics is of fundamental importance to efficiently explore the target input space and, eventually, find latent bugs in the code. However, real-world programs tend to contain conditions that are hard to trigger. These complex conditions need to be addressed by fuzzers in order to execute as much

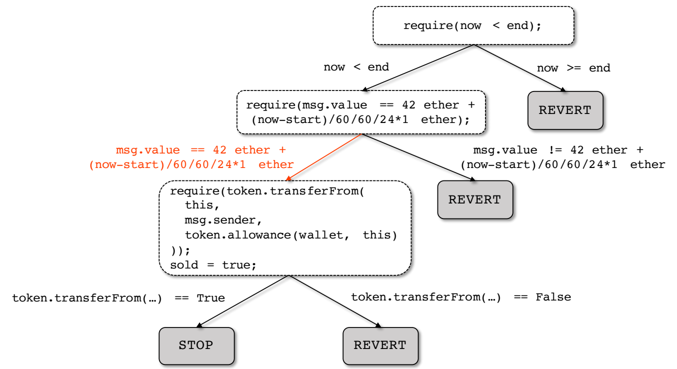

Fig. 2. Control-flow graph of the function buy(), where the complex condition is highlighted in red.

code as possible. Line 21 in Fig. 1 provides an example of a complex condition. Function buy requires the transaction value to be equal to 42 ether along with a variable fee that depends on the number of days that have past since the launch of the token sale. Fig. 2 illustrates the control-flow graph (CFG) of the function buy along with its branching conditions. The complex condition is highlighted in red in the CFG. A fuzzer following a traditional random strategy will fail to get past this condition since it will generate the desired value only once every 2^{256} trials.

Existing smart contract fuzzers such as HARVEY [50] instrument the code and compute cost metrics for every branch to mutate the inputs. Our approach applies constraint solving to generate values for complex conditions on-demand. However, our fuzzer does not directly propagate these values, but instead, it stores them in so-called mutation pools. Mutation pools manage a set of values that the fuzzer can use to get past complex conditions. Every function has its own set of mutation pools, namely a mutation pool per function argument, transaction argument (e.g. transaction value), and environmental argument (e.g. block timestamp). Initially, all the pools are empty, and the fuzzer uses randomly generated values to feed the target functions. Once the fuzzer cannot discover new paths, it activates the constraint solver to generate new values. We use symbolic taint analysis to create the expressions required by the constraint solver to generate new values. We introduce taint in the form of a symbolic value whenever we come across an input during execution. This symbolic value is then propagated throughout the program execution, thereby forming step-by-step a symbolic expression that reflects the constraints on the particular input. Solving these expressions will result in new values that will be added to the mutation pools. The fuzzer will then pick these values from the mutation pools and generate new inputs that execute new paths. In Fig. 1, once CONFUZZIUS realizes that the code coverage is not increasing, it activates the constraint solver, which outputs the value 42 along with the current fee depending on the current block timestamp. It adds it to the mutation pool that manages the transaction value for the function buy. The value will be then picked up from the mutation pool by the fuzzer

in the next round, and the execution of the transaction will evaluate the condition at line 21 to `True`, which results in getting past the missing branch and executing new lines of code.

C. *Stateful Exploration*

Due to the transactional nature of blockchains, smart contract fuzzers must consider that each transaction may have a different output depending on the contract's current state, i.e. all the previously executed transactions. Appropriately combining multiple transactions is necessary to generate states that trigger the execution of new branches. Ethereum smart contracts have, besides a volatile memory model, also a persistent memory model called *storage*, which allows them to keep state across transactions. For example, the global variables `end`, `wallet`, `token`, `owner`, and `sold` in Fig. 1 are storage variables and their values might change across transactions. Let us consider the two vulnerabilities mentioned earlier. An attacker will only be able to extract the funds via the function `withdraw`, if the two variables `owner` and `sold` contain the address of the attacker and `True`, respectively. However, this is only possible if the functions `buy` and `Tokensale` are called before the function `withdraw`. Thus only a particular combination of the three functions will trigger the two vulnerabilities. Although this example may seem straightforward, automatically finding the right combination of function calls within contracts with many functions can become challenging as the number of possible combinations grows exponentially. We base our solution on a simple observation: a transaction influences the output of a subsequent set of transactions if and only if it modifies a storage variable that one of the subsequent transactions will use. This property is a known data dependency called *read-after-write* (RAW) [19]. In the first step, ConFuzzius traces all the storage reads and writes performed by a transaction along with the storage locations. Afterwards, ConFuzzius combines transactions so that transaction a is executed after transaction b, only if a reads from the same storage location where b writes to. The fuzzer always executes the combination of transactions on a clean state of the contract. Thus, a transaction sequence contains only transactions that change the state used by one of the subsequent transactions within the same sequence by construction. In the example of Fig. 1, ConFuzzius will progressively learn that:

- `buy` reads variable `end` and writes to variable `sold`;
- `Tokensale` writes to variable `owner`;
- `withdraw` reads variable `owner` and variable `sold`.

Using the information learned above and combining transactions based on RAW dependencies, ConFuzzius will eventually create the following transaction sequence:

```
buy() → Tokensale() → withdraw()
```

The directed graph in Fig. 3 presents the RAW dependencies to generate all the possible combinations. The graph shows that the functions `buy` and `Tokensale` must be executed before the function `withdraw`, but that the order between the two can be arbitrary.

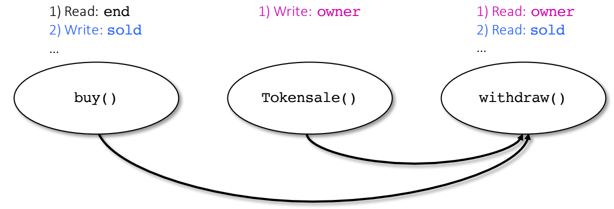

Fig. 3. A dependency graph illustrating the read-after-write (RAW) data dependencies contained in Fig. 1. A node represents a smart contract function and an edge indicates a RAW dependency between the two functions.

D. *Environmental Dependencies*

The execution of a smart contract does not only depend on the transaction arguments or the contract's current state. A smart contract's control-flow can also depend on input originating from the execution environment (e.g. a block's timestamp). Let us consider the contract in Fig. 1. Even though the function `withdraw` has no input argument, the transfer of the balance is bound to some requirements. The requirement at line 28 is only satisfied if the transaction that triggered the function call is part of a block created 30 days after the contract's deployment. Thus, the condition is bound to the mining mechanism of the Ethereum blockchain. While users submit transactions to the blockchain, miners aggregate them into blocks and distribute them to other nodes upon validation. When executing the transactions included in the block, the EVM accesses the block information contained therein. Block information includes the block hash, the miner's address, the block timestamp, the block number, the block difficulty, and the block gas limit. We solve this challenge by modeling this information as a fuzz-able input. These inputs follow the same fuzzing procedure as transaction inputs. We modified the EVM in order to be able to inject the fuzzed block information during the execution of the smart contract. However, modeling block information as fuzz-able inputs is not enough. The EVM also permits to call other contracts deployed on the blockchain. Thus the control-flow of a smart contract may depend on the result of calling other contracts. Consider line 29 in Fig. 1, where the state variable `sold` is required to be set to `True` in order for the attacker to be able to retrieve the funds. The variable `sold` can only be set to `True` if the two contract calls at line 22 (e.g. `token.allowance` and `token.transferFrom`) are successful. We solve this challenge in a similar way by instrumenting calls to contracts and modeling return values as fuzz-able inputs. Our modified EVM then injects the fuzzed return values at runtime.

IV. Design and Implementation

In this section, we provide details on the overall design and implementation of ConFuzzius.

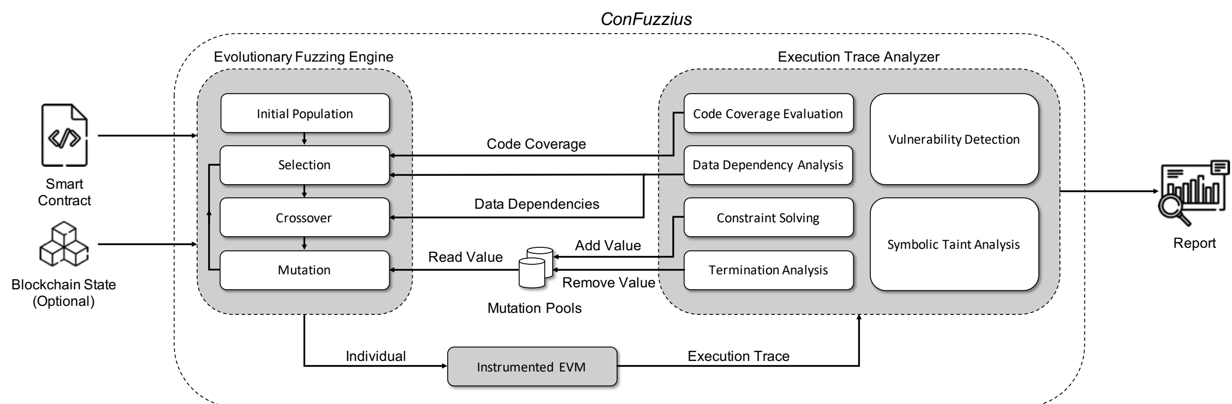

Fig. 4. Overview of ConFuzzius's hybrid fuzzing architecture. The shadowed boxes represent the three main components and form together a feedback loop.

A. Overview

ConFuzzius's architecture consists of three main modules: the evolutionary fuzzing engine, the instrumented EVM, and the execution trace analyzer. Fig. 4 provides a high-level overview of ConFuzzius's architecture and depicts its individual components. ConFuzzius has been implemented in Python with roughly 6,000 lines of code[1]. ConFuzzius takes as input the source code of a smart contract and a blockchain state. The latter is in the form of a list of transactions and is optional. The blockchain state is convenient for fuzzing already deployed smart contracts or contracts that need to be initialized with a specific state. ConFuzzius begins by compiling the smart contract to obtain the Application Binary Interface (ABI) and the EVM runtime bytecode. The evolutionary fuzzing engine then starts by generating individuals for the initial population, based on the smart contract's ABI. After that, the engine follows a standard genetic algorithm (i.e., selection, crossover, and mutation) and propagates the newly generated individuals to the instrumented EVM. The instrumented EVM then executes these individuals and forwards the resulting execution traces to the execution trace analyzer. Next, the execution trace analyzer performs several analyses, e.g. symbolic taint analysis, data dependency analysis, etc. The execution trace analyzer is also responsible for triggering the constraint solver, running the vulnerability detectors, updating the mutation pools, and feeding information related to code coverage and data dependencies to the evolutionary fuzzing engine. This process is repeated until at least one of the two termination conditions is met: a given number of generations has been generated, or a given amount of time has passed. Finally, ConFuzzius outputs a report containing information about the code coverage and the vulnerabilities that it detected.

B. Evolutionary Fuzzing Engine

In the following, we provide details on the encoding, initialization, fitness evaluation, selection, combination, and mutation of individuals.

[1]Source code is available at https://github.com/christoftorres/ConFuzzius.

Encoding Individuals. One of the most important decisions to make while implementing an evolutionary fuzzer is deciding on the representation of individuals. Improper encoding of individuals can lead to poor performance [27]. Fig. 5 illustrates our encoding of individuals. Vulnerabilities are usually triggered either by sending a single transaction or a sequence of transactions to a smart contract. However, transactions alone are not enough to trigger vulnerabilities (see Section III-D). Specific vulnerabilities depend on the execution environment to be in a specific state. Thus, our encoding represents an individual as a sequence of inputs. Every input consists of an environment and a transaction. Both are encoded as key-value mappings. The environment includes block information such as the current timestamp and block number, but it also includes call return values, data sizes, and external code sizes. The latter three are encoded as an array of mappings, where a contract address maps to a mutable value (e.g. a call result or a size). The transaction includes the address of the sending account (*from*), the transaction amount (*value*), the maximum amount of gas for the contract to execute (*gas limit*), and the input data for the contract to execute (*data*). The input data is represented as an array of values where the first element is always the function selector, and the remaining elements represent the function arguments. The function selector is computed using the ABI and extracting the first four bytes of the Keccak (SHA-3) hash of the function signature. As an example, the function `test(string a, uint b)`, has the string `test(string,uint)` as its function signature, which after hashing and extracting the first four bytes, results in `0x7d6cdd25` being its function selector.

Initial Population. The population is initialized with N individuals, each of which initially contains only a single input (i.e. a single transaction). The function selector to be included in the transaction is selected in a round-robin fashion. Function arguments are generated based on their type, which we obtain through the ABI. Depending on the type and size (i.e. fixed or non-fixed) of the argument, we apply different

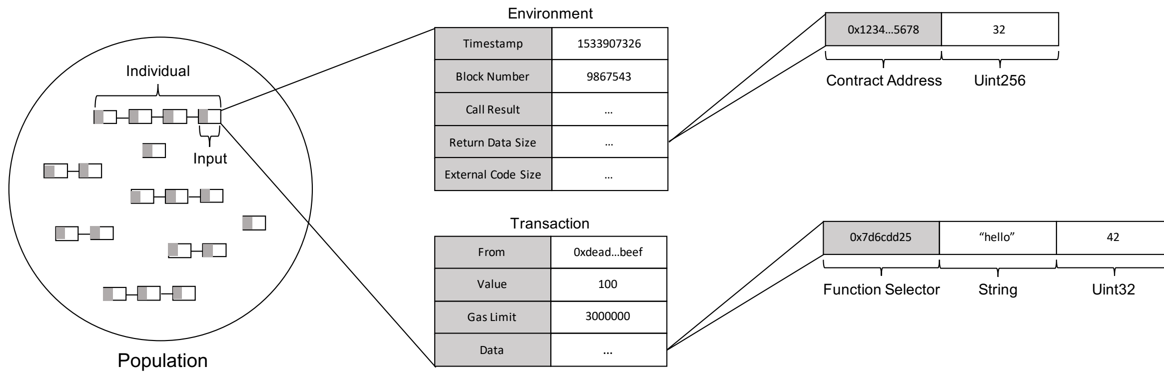

Fig. 5. Encoding of our population and its individuals. The shadowed boxes depict immutable values, whereas the non-shadowed boxes depict mutable ones.

strategies to generate valid arguments for each function. For example, if the argument type is a fixed size `uint32`, then we randomly choose a value either from the valid input domain (e.g. between 0 and $2^{32}-1$) or from a set of inputs that trigger edge cases of the valid input domain (e.g. 0, 1, 4294967295). The population is reinitialized whenever there has been no increase in code coverage for the past k generations. This soft reboot introduces back diversity into the population and procrastinates premature convergence when the population has become homogeneous.

Fitness Evaluation. The fitness evaluation of individuals plays a crucial role in evolutionary fuzzing. The computation of the fitness function is done repeatedly and must be therefore sufficiently fast. A slow computation can adversely make the fuzzing exceptionally slow. The fitness function is supposed to represent the landscape of the problem. In general, evolutionary fuzzers aim to achieve complete coverage of the code. While obtaining full code coverage does not necessarily mean that all vulnerabilities will be found, it is undoubtedly true that no vulnerabilities will be found in code that has not been explored. Our fitness function is based on branch coverage (a form of code coverage) and data dependencies. We define our fitness function for an individual i as follows:

$$fit(i) = fit_{branch}(i) + fit_{RAW}(i) \tag{1}$$

The fitness fit_{branch} is computed by counting the number of branches that remain unexplored by the individual. We keep track of all the branches that have been executed so far by all the individuals. Then, we iterate through the execution trace of the individual and analyze every conditional jump instruction (i.e. `JUMPI` instruction). A conditional jump always has two destinations, one for the `True` branch and one for the `False` branch. We obtain the jump destination of the `True` branch by extracting it from the stack and the jump destination of the `False` branch by incrementing the program counter by one. We increase the individual's fitness value fit_{branch} by one for every jump destination that is not in our list of executed branches. This approach aims to prioritize individuals that require more exploration since these individuals will allow us to explore new parts of the contract. However, this metric alone is not enough. We are also interested in preserving individuals that allow us to create useful sequences of transactions (e.g. sequences with RAW dependencies), even though these individuals might have been already explored extensively. Therefore, we compute the fitness fit_{RAW}, which takes this into account by using the data dependencies detected by our execution trace analyzer. We start with a fit_{RAW} of zero and increment fit_{RAW} by one for every write to storage that the individual has performed during execution. The final fitness value of an individual is then defined by the the sum of the two fitness values fit_{branch} and fit_{RAW}. The combination of these two values allows us to drive the genetic algorithm to explore unexplored code while preserving individuals that are useful to explore deeper states of the smart contract.

Selection. The process of choosing two individuals for the crossover step is called selection. Literature proposes several selection operators [41]. We choose linear ranking selection as our selection operator since it considers the population as a whole during selection and not just a subset as it is, for example, the case in tournament selection. In linear ranking selection, individuals with a high fitness value are ranked higher than those with a low fitness value. In other words, an individual is selected with a probability that is linearly proportional to the individual's rank in the population. Hence, the worst individual has a rank of 1, the second-worst a rank of 2, the best-performing individual has a rank of N, where N is the size of the population. All individuals have a chance of being selected, although the higher-ranked individuals will be slightly preferred. However, traditional linear ranking selection does not consider data dependencies between individuals. Therefore, we propose a modified version of the linear ranking selection strategy, where the first individual is selected based on linear ranking selection, however, the second individual is selected based on having a RAW dependency with the first individual following a round-robin fashion. In case there is no individual to have a RAW dependency with the first individual, we fallback to traditional linear ranking selection to select the second individual.

Crossover. The crossover operator creates two new individuals by recombining the input sequences of two existing individuals. Instead of randomly combining two individuals, we combine an individual after another only if the first performs a write to a storage location from which the second performs a read (RAW dependency). There are only two possible combinations in our case: individual a appended to individual b, or vice versa, individual b appended to individual a. If a combination yields a RAW dependency, then we combine both individuals by first selecting the individual whose input sequence performs the write and then append the individual whose input sequence performs the read. As opposed to traditional crossover, we are concatenating individuals rather than splitting them apart and swapping their input sequences. This preserves the RAW dependencies within the individuals themselves and creates individuals with new RAW dependencies. If there is no RAW dependency between two individuals, then we simply return one of the two individuals unmodified. However, it should be noted that individuals are not always combined even though they might have a RAW dependency. Individuals are combined based on a given crossover probability p_c. Moreover, to prevent individuals from growing indefinitely large, we check before combining if the sum of their lengths exceeds the maximum size of l, and only combine them if their lengths are lower or equal to l. On the one hand, a small l will produce shorter combinations of individuals, which will result in shorter execution times but also in finding less bugs. On the other hand, a larger l will result in longer combinations of individuals, which will result in finding more bugs that lay deeper in the execution, but also in longer execution times. Therefore, a trade-off between completeness and performance must be taken when selecting an appropriate value for l.

Mutation. The mutation operator randomly modifies parts of a single individual in order to create a new individual. It introduces diversity in the population. Our mutation operator works by iterating through the sequence of inputs of an individual and mutating every environmental and transactional value based on a shared mutation probability p_m. A value can be mutated in two ways: replacing the original value with a random one or replacing the original value with a value from a *mutation pool*. Mutation pools act as a form of short-term memory. They allow the fuzzer to reuse values that have been previously observed or learned during past executions. There are in total nine different mutation pools, one per transactional and environmental value type. Hence, our fuzzer has a mutation pool for senders, amounts, gas limits, function arguments, timestamps, block numbers, call results, call data sizes, and external code sizes. All mutation pools are implemented to map a function selector to a circular buffer, except for the mutation pool on function arguments. The implementation is similar, except that we do not directly map the function selector to a circular buffer but to another mapping that maps to an argument index and then to a circular buffer. Thus, the pool for function arguments first maps to a function selector, then to an argument index, and then to a circular buffer. This is because functions can have more than just one argument, and we want to keep track of interesting values for every argument separately. Circular buffers help us ensure that the values contained therein are rotated in a round-robin like fashion. Old values are overwritten by newer ones (i.e. mimicking short-term memory). Our buffers can hold up to 10 values by default. All mutation pools are initially empty, except for the mutation pool tracking transaction amounts, which gets initialized with the values 0 and 1. When mutating a transactional or environmental value, we first check if the associated mutation pool is empty. If the pool is empty, we inject a randomly generated value based on the type of information extracted from the ABI. Otherwise, we inject the current value at the head of the circular buffer and rotate it.

C. *Instrumented EVM*

The EVM is responsible for executing the transactions generated by the individuals on the runtime bytecode of the contract that is under test. Its efficiency has a significant impact on the overall performance of the fuzzer. Hence, the EVM must achieve a high processing rate of transactions. Every official Ethereum client implementation allows users to deploy smart contracts locally and send transactions to them. However, all of these clients require transactions to be encoded using the Recursive Length Prefix (RLP) format in order to be mined. We realized that the actual EVM execution time is negligible compared to the effort of encoding and decoding a transaction to and from the RLP format. Therefore, we decided to reuse an official Python implementation of the EVM [13] and incorporate it within our fuzzer. This removes the burden of mining blocks as well as encoding transactions and thus significantly speeds up the execution. Moreover, we slightly modified the EVM in order to be able to retrieve the execution trace of a transaction. An execution trace consists of an array, where every element contains the name of the executed instruction, the program counter, the execution stack, the call-stack depth, and a flag stating if an internal error has occurred during execution. The EVM itself is by default stateless and uses the blockchain to preserve states. However, since we are not interested in the internal persistence mechanism of the Ethereum blockchain, we decided to implement a simple storage emulator that is used by our EVM to persist the state changes that are performed during execution. All state changes are kept in memory to further improve the speed of execution. Besides persisting the state of smart contract executions, the storage emulator also allows us to inject custom environmental information such as the block timestamp or modify the result of a call. Finally, the storage emulator also enables us to create snapshots of the current state of the EVM. This allows us to quickly reset the state of the EVM to an initial state without having to redeploy the smart contract every time from scratch when executing the transactions of an individual.

D. *Execution Trace Analyzer*

The execution trace analyzer receives from the instrumented EVM the execution trace and then performs a number of

TABLE I
STORAGE LAYOUT OF STATE VARIABLES IN SOLIDITY

Variable Type	Declaration	Access	Storage Location
Primitive	`T v`	`v`	s(v)
Struct	`struct v { T a }`	`v.a`	s(v) + s(a)
Fixed Array	`T[10] v`	`v[n]`	s(v) + n · |T|
Dynamic Array	`T[] v`	`v[n]`	h(s(v)) + n · |T|
		`v.length`	s(v)
Mapping	`mapping(T1 => T2) v`	`v[k]`	h(k || s(v))

analyses, such as code coverage evaluation, data dependency analysis, symbolic taint analysis, vulnerability detection, constraint solving, and termination analysis. Moreover, the execution trace analyzer also manages the values stored within the mutation pools and is responsible for providing code coverage information and data dependencies to the evolutionary fuzzing engine. Finally, it is also in charge of generating a report containing statistics about code coverage and vulnerabilities.

Code Coverage Evaluation. Code coverage is necessary for computing an individuals' fitness and detecting when the evolutionary fuzzing engine should activate constraint solving or reset the population. The code coverage is computed by counting the number of unique program counter values within the execution trace.

Data Dependency Analysis. Fitness evaluation, selection, and crossover require information about data dependencies. The data dependency analysis tracks all the state variables, read from and written to, throughout the execution of the fuzzer. In contrast to existing approaches [52], which extract data dependencies via static analysis, our fuzzer retrieves the access patterns to state variables at runtime (i.e. dynamically) by iterating through the execution trace and scouting for SLOAD and SSTORE instructions. The advantage of static analysis is that it is fast compared to dynamic analysis. However, the disadvantage is that it requires source code to precisely track data flows across variables. While data flows could be extracted from bytecode through static analysis, it would only work for simple variable types such as primitives and not for complex types such as mappings or arrays. Dynamic analysis on the other hand, allows us to track variables with complex types, even without source code. The disadvantage is the additional runtime overhead and implementational effort. The instruction SLOAD denotes a read from storage, whereas an SSTORE instruction denotes a write to storage. Table I depicts how Solidity computes the storage location for different types of state variables [38]. The function s() determines the so-called *storage slot* of a particular variable v, whereas the function h() computes a Keccak-256 hash. Statically-sized variables such as primitives, structs, and fixed-size arrays, are laid out contiguously in storage starting from position 0. Solidity uses a Keccak-256 hash computation to define the stored data's starting position due to the unpredictable size of dynamically-sized arrays and mappings. However, we are not interested in identifying individual storage locations but rather access to a particular variable. Therefore, our goal is to extract the storage slot s(v) for a variable v, instead of the storage location. As an example, assume we have a variable called balances of type mapping, which maps an address to a uint, and we have two addresses a and b. The storage location for balances[a] and balances[b] will be different, since the computation of the storage locations will be h(a || s(balances) and h(b || s(balances)), respectively, where || means concatenation. However, these two storage locations share the same storage slot s(balances), which enables us to link both storage locations together to the same variable balances. Extracting the storage slot for statically-sized variables is straightforward as it is equivalent to the storage location. It can be achieved by merely popping the first element from the stack for both instructions, SLOAD and SSTORE. Extracting storage slots for mappings and dynamic arrays is more challenging. As it is not possible to invert the result of a hash, we must keep track of the Keccak-256 hash computations by mapping the result of a SHA3 instruction to the memory contents that were involved in the computation. Note that the SHA3 instruction computes the Keccak-256 hash from a memory slice determined via two arguments on the stack, namely the memory offset and the memory size. Thus, for mappings, all we need to do is to obtain the mapped memory contents. To deal with concatenation, we only extract the last 32 bytes of the memory contents as these represent the storage slot. Finally, for dynamic arrays, we must keep track of the arithmetic addition of Keccak-256 hashes. The storage slot is then determined the same way as for mappings, except that there is no concatenation.

Symbolic Taint Analysis. The symbolic taint analysis produces symbolic constraints that are later used by other parts, such as constraint solving and vulnerability detection. We introduce taint in the form of symbolic values and track their flow across instructions. We leverage light dynamic taint analysis by injecting taint only for instructions that can be fuzzed, e.g. CALLDATALOAD, CALLVALUE, or TIMESTAMP. Taint is propagated across stack, memory, and storage of the execution trace. The propagation of taint across storage allows us to do inter-transactional taint analysis. We implemented the stack using an array structure that follows LIFO logic. We used a Python dictionary that maps memory and storage addresses to values to represent memory and storage. Since the EVM is a stack-based register-less virtual machine, the operands of instructions are always passed via the stack. Therefore, our taint propagation method identifies each EVM bytecode instruction's operands and propagates the taint according to each instruction's semantics as defined in the yellow paper [49]. The taint propagation logic follows an over-tainting policy, which tags the instruction's output as tainted if at least one of the instruction's inputs are tainted.

Constraint Solving. There are situations where the evolutionary fuzzing engine converges prematurely because it cannot advance past a complex conditional statement. The constraint

solver's role is then to generate a valid input that allows the evolutionary fuzzing engine to get past the complex condition. The symbolic taint analysis tries to build a logical formula that describes the complex execution path, thereby reducing the problem of reasoning about the execution to the domain of logic. These logical formulas are often called path constraints. We implemented our own lightweight symbolic execution engine, that only executes instructions related to arithmetic operations (e.g. `ADD`, `MOD`, `EXP`), comparison logic (e.g. `LT`, `EQ`) and bitwise logic (e.g. `AND`, `NOT`). The engine consists of an interpreter loop that gets instructions from the execution trace and symbolically executes them. The loop continues until all the instructions contained in the execution trace have been executed. We obtain the formulas only for conditional statements with open branches, i.e. never executed branches. Each formula contains the path constraints to reach the conditional statement. We negate the last constraint, substitute the symbolic variables in the rest of the logical formula with concrete values that have been used as inputs to trigger the execution trace, and use the Z3 SMT solver [11] to produce inputs to reach the open branch. Concretization helps us reduce the complexity of the formula and therefore avoid the path explosion problem. We then add the produced inputs to the mutation pools. Eventually, in one of the following generations, the mutation operator will pick up the solution, and our evolutionary fuzzing engine will now be able to get past the complex condition. We also add the previously used inputs to the mutation pools, allowing the fuzzer to execute both branches.

Termination Analysis. The execution traces may contain valuable feedback on the validity of the inputs. Our fuzzer uses the execution traces to obtain feedback and learn whether an input is meaningful or not. The termination analysis inspects the execution traces for opcodes that indicate either correct or incorrect termination of execution. Invalid inputs will result in the execution trace terminating with a `REVERT`, `INVALID`, or `ASSERTFAIL` instruction, whereas valid inputs will result in the execution terminating with either a `SELFDESTRUCT`, `SUICIDE`, `STOP`, or `RETURN` instruction. Once we detect an incorrect termination, we analyze the last path constraint before the termination and retrieve the input values responsible for the incorrect termination. We then remove these values from the mutation pools. A transaction that results in an incorrect termination reverts all state changes made during execution and is therefore not relevant for creating meaningful RAW dependencies. This helps the fuzzer focus on more relevant parts of the code.

Vulnerability Detection. We detect vulnerabilities by analyzing the execution traces and the information returned by the symbolic taint analysis and the data dependency analysis. We define a detector per vulnerability. We implemented detectors for the following types of vulnerabilities: assertion failure (AF), integer overflow (IO), reentrancy (RE), transaction order dependency (TD), block dependency (BD), unhandled exception (UE), unsafe delegate call (UD), leaking ether (LE), locking ether (LO), and unprotected self-destruct (US). A detailed explanation of each vulnerability and its detector is described in the Appendix A. More detectors can be easily added to extend the detection capabilities of our fuzzer.

V. Evaluation

In this section, we evaluate the effectiveness and performance of ConFuzzius by answering the following three research questions:

- **RQ1** Does ConFuzzius achieve higher code coverage than current state-of-the-art symbolic execution and fuzzing tools for smart contracts?
- **RQ2** Does ConFuzzius discover more vulnerabilities than current state-of-the-art symbolic execution and fuzzing tools for smart contracts?
- **RQ3** How relevant are ConFuzzius's individual components in terms of code coverage and vulnerability detection?

Datasets. We run our experiments using two different datasets. The purpose of the first dataset is to measure code coverage, whereas the second dataset aims to measure the detection of vulnerabilities[2]. The first dataset was obtained by crawling Etherscan's list of verified smart contracts [14]. These are real-world smart contracts where the source code is publicly available and verified to match the bytecode deployed on the Ethereum blockchain. We filtered out contracts that failed to compile using Solidity version 0.4.26, resulting in a dataset of 21,147 contracts. Moreover, we split our dataset into different clusters based on each contract's number of EVM bytecode instructions. The idea is to examine code coverage on smart contracts with different sizes. We used the standard k-means clustering algorithm to create the clusters. The number of clusters has been determined using the Elbow and the Silhouette method. Both methods yield 2 to be the optimal value for k. Table III lists the number of lines of source code (LoSC), number of public functions, and the number of EVM bytecode instructions for each cluster and the overall dataset. The first cluster represents small contracts ($\leq$ 3,632 instructions) and contains 17,803 contracts, whereas the second cluster represents large contracts ($>$ 3,632 instructions) and contains 3,344 contracts. The second dataset is based on Durieux et al.'s curated dataset [12], a collection of annotated smart contracts, which the authors used to evaluate the effectiveness of smart contract analysis tools. Unfortunately, their curated dataset is missing certain types of vulnerabilities (e.g. assertion failures). We decided to reuse their dataset and extend it with annotated vulnerabilities from the Smart Contract Weakness Classification (SWC) registry [42]. We added contracts related to assertion failures, unsafe delegatecalls, leaking ether, locking ether, and unprotected selfdestructs. The extended dataset consists of 128 contracts with 148 annotated vulnerabilities.

[2]Both datasets are available at https://github.com/christoftorres/ConFuzzius.

TABLE II
SECURITY TOOLS EVALUATED IN THIS WORK. TOOLS MARKED WITH ● SUPPORT THE DETECTION OF THE VULNERABILITY, WHILE TOOLS MARKED WITH ○ DO NOT SUPPORT THE DETECTION OF THE VULNERABILITY.

Toolname	Type	Requires Source Code	Requires ABI	Vulnerability Detectors									
				AF	IO	RE	TD	BD	UE	UD	LE	LO	US
OYENTE [28]	Symbolic	✗	✗	●	●	●	●	●	○	○	○	○	●
MYTHRIL [31]	Symbolic	✗	✗	●	●	●	●	●	●	●	●	○	●
M-PRO [31]	Symbolic	✓	✗	●	●	●	●	●	●	●	●	○	●
ILF [18]	Fuzzer	✓	✓	○	○	○	○	●	●	●	●	●	●
SFUZZ [32]	Fuzzer	✓	✓	○	●	●	○	●	●	●	○	●	○
CONFUZZIUS	**Hybrid**	✗	✓	●	●	●	●	●	●	●	●	●	●

TABLE III
STATISTICS OF OUR CLUSTERS AND OVERALL DATASET.

Dataset	LoSC			Functions			Instructions		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Small	1	3,584	119	1	81	14	1	3,632	1,763
Large	44	3,148	429	1	190	32	3,633	16,889	5,495
Overall	**1**	**3,584**	**168**	**1**	**190**	**17**	**1**	**16,889**	**2,353**

Baselines. We compare CONFUZZIUS to the tools listed in Table II. We limit our comparison to symbolic execution tools and fuzzers as we want to know if our hybrid approach performs better than these methods on their own. We chose OYENTE [28] because of its popularity among the community and continuous development. We chose MYTHRIL since a recent study on smart contract analysis tools [12] revealed that it performs better than a variety of existing tools (e.g. Manticore [34], SmartCheck [44], Securify [47], Maian [33], etc.). We chose M-PRO because it employs a similar transaction sequence combining strategy than ours, with the difference being that ours is dynamic, and theirs is static. We chose ILF [18] since it has proven to outperform existing smart contract fuzzers (i.e. CONTRACTFUZZER [22] and ECHIDNA [10]). We chose SFUZZ because it is a recent work that has not been compared yet by previous works and because it is based on the popular fuzzing tool AFL. Table II compares the different types of vulnerabilities detected by each tool. We see that none of the tools are currently able to detect all of the ten vulnerabilities that are currently detectable by CONFUZZIUS. We also see that CONFUZZIUS does not requires source code as compared to the other fuzzers.

Experimental Setup. We followed the guidelines by Klees et al. [24] on evaluating fuzz testing. For each experiment, we performed 10 runs, each with independent seeds. For the first dataset we run experiments on the small contracts for 10 minutes each, whereas on the large contracts for 1 hour each. Preliminary tests showed that most tools did not yield more coverage past these times. Moreover, before every run, we initialized CONFUZZIUS's and ILF's blockchain state with the same values that were used to deploy the contracts on the Ethereum mainnet. For the second dataset, we run the experiments for each contract for 10 minutes. We run our experiments on a cluster of 10 nodes, each with 128 GB of memory. Every node runs CentOS release 7.6.1810 and has 2 Intel® Gold 6132 CPUs with 14 cores, each clocked at 2.60 GHz. We run CONFUZZIUS with a variable population size, that was computed as two times the number of functions contained in the ABI of the contract under test. We set the crossover probability and the probability of mutation to 0.9 and 0.1, respectively. The population is reinitialized whenever the code coverage does not increase for $k = 10$ generations. We set the maximum length for individuals to $l = 5$. Finally, we used Z3 version 4.8.5 as our constraint solver with a timeout of 100 milliseconds per Z3 request.

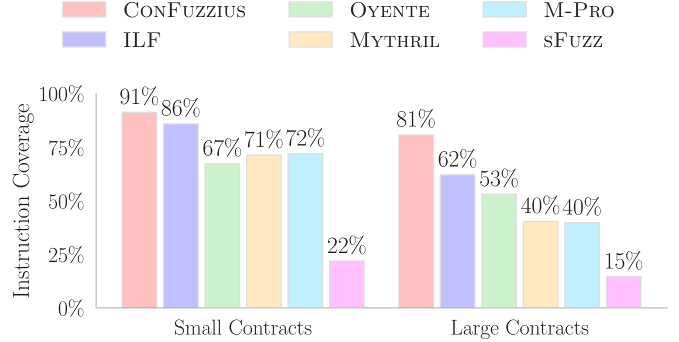

Fig. 6. Overall instruction coverage of CONFUZZIUS and other tools.

A. Code Coverage

Fig. 6 depicts the overall instruction coverage (e.g. the average of all contract runs) of CONFUZZIUS and other security tools on the clusters of small and large contracts. CONFUZZIUS achieves the highest coverage on the small and large contracts, with 91% and 81%, respectively. As expected, every tool struggles with the larger contracts. We see that the overall coverage is less for the larger contracts than for the smaller contracts. However, while the difference is roughly 10% for CONFUZZIUS, symbolic execution tools such as MYTHRIL have a difference of 31%. Fig. 7 compares the overall instruction coverage of CONFUZZIUS and the two other fuzzers, ILF and SFUZZ, over time. We solely plotted the instruction coverage for these three tools as we do not have coverage information over time for symbolic execution tools. CONFUZZIUS not only consistently outperforms ILF and SFUZZ, but it also achieves more code coverage in a shorter time. On the small contracts, CONFUZZIUS achieves

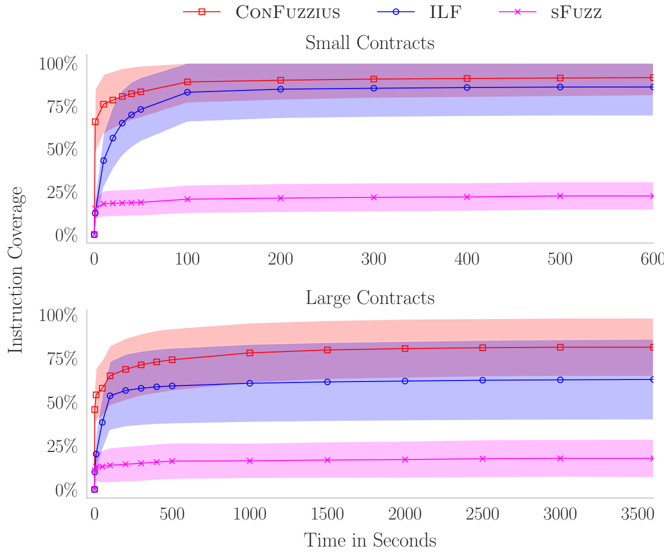

Fig. 7. Overall instruction coverage of CONFUZZIUS, ILF and SFUZZ over time.

TABLE IV
TRUE POSITIVES AND FALSE POSITIVES DETECTED BY EACH TOOL PER VULNERABILITY TYPE.

Toolname	AF	IO	RE	TD	BD	UE	UD	LE	LO	US	Total
	Vulnerabilities										
OYENTE	6/6	12/4	8/0	2/0	0/0	-	-	-	-	0/0	28
MYTHRIL	7/3	18/5	10/0	0/0	3/0	24/0	0/0	4/0	-	2/0	68
M-PRO	7/3	18/5	10/0	0/0	3/0	24/0	0/0	4/0	-	2/0	68
ILF	-	-	-	-	0/0	10/0	1/2	4/0	5/0	3/0	23
SFUZZ	-	12/0	7/0	-	1/0	21/0	1/2	-	0/0	-	42
CONFUZZIUS	**10/0**	**18/0**	**10/0**	**2/0**	**7/0**	**46/0**	**1/0**	**4/0**	**5/0**	**3/0**	**106**
Total Unique	14	19	11	4	7	75	1	9	5	3	148

after 1 second 66% instruction coverage, whereas ILF and SFUZZ achieve solely 12% and 15%, respectively. On the large contracts, CONFUZZIUS achieves after 1 second 46% instruction coverage, whereas ILF and SFUZZ achieve only 10% and 11%, respectively.

B. Vulnerability Detection

Table IV summarizes the vulnerabilities detected by each tool for each of the 10 categories on the extended curated dataset. Each entry shows the number of true positives (left-hand side) and false positives (right-hand side). For example, OYENTE reported for assertion failure (AF), 6 true positives and 6 false positives (i.e. 6/6), whereas MYTHRIL reported for assertion failure (AF), 7 true positives and 3 false positives (i.e. 7/3). Overall, we see that CONFUZZIUS detected the most number of vulnerabilities, namely 106 out of 148 vulnerabilities, that is roughly 71% of all the vulnerabilities. ILF detected the least number of vulnerabilities, with 23 out of 148. In the following, we discuss the results obtained for each category.

Assertion Failure (AF). CONFUZZIUS detects more assertion failures than the other tools, namely 10 out of 14, and does not report any false positives. Both, MYTHRIL and M-PRO, report 7 assertion failures and 3 false positives. OYENTE reports 6 true positives and 6 false positives. Our manual investigation reveals that they over-approximate the satisfiability of execution paths due to incorrect modeling. For example, OYENTE reports in Fig. 8 an assertion failure at line 8. However, this is not possible because `param` is set at the constructor and is checked to be always larger than zero.

```
1  contract AssertMultiTx1 {
2      uint256 private param;
3      constructor(uint256 _param) {
4          require(_param > 0);
5          param = _param;
6      }
7      function run() {
8          assert(param > 0);
9      }
10 }
```

Fig. 8. False positive reported by OYENTE on an assertion failure.

Integer Overflows (IO). CONFUZZIUS reports the same number of integer overflows as MYTHRIL and M-PRO, namely 18 out of 19. However, MYTHRIL and M-PRO also report 5 false positives, whereas CONFUZZIUS does not report any. For example, MYTHRIL reports in Fig. 9 an integer overflow at line 6. However, there is no possibility to initialize `balanceOf`, therefore an overflow can never occur because the **`require`** statement at line 4 will never be satisfied.

```
1  contract IntegerOverflowAdd {
2    mapping (address => uint256) balanceOf;
3    function transfer(address _to, uint256 _value)
         {
4      require(balanceOf[msg.sender] >= _value);
5      balanceOf[msg.sender] -= _value;
6      balanceOf[_to] += _value;
7    }
8  }
```

Fig. 9. False positive reported by MYTHRIL on an integer overflow.

Reentrancy (RE). CONFUZZIUS, MYTHRIL and M-PRO detect the same number of reentrancy vulnerabilities, namely 10 out of 11. OYENTE and SFUZZ detect 8 and 7, respectively.

Transaction Order Dependency (TD). CONFUZZIUS and OYENTE detect 2 contracts vulnerable to transaction order dependency. Both, MYTHRIL and M-PRO do not detect any of the 4 contracts to be vulnerable to transaction order dependency.

Block Dependency (BD). CONFUZZIUS is the only tool capable of detecting all the 7 block dependencies. Our manual investigation reveals that CONFUZZIUS is capable of detecting more block dependencies because of its environmental modeling, which allows CONFUZZIUS to fuzz block information and therefore, CONFUZZIUS can detect more calls that are dependent on block information.

Unhandled Exception (UE). CONFUZZIUS reports the largest number of unhandled exceptions, namely 46 out of 75. MYTHRIL and M-PRO report 24 unhandled exceptions. ILF and SFUZZ report 10 and 21 unhandled exceptions, respectively. Similar to block dependency, CONFUZZIUS detects more unhandled exceptions because it models call return values as environmental information that can be fuzzed. This allows CONFUZZIUS to simulate exceptions and check if they are handled.

Unsafe Delegatecall (UD). CONFUZZIUS is the only tool that detects unsafe delegatecall without false positives. Both, MYTHRIL and M-PRO were not able to detect any unsafe delegatecalls. ILF and SFUZZ detect an unsafe delegatecall, but also report 2 false positives. For example, in Fig. 10 ILF reports an unsafe delegatecall at line 13. However, the variable `callee` can only be changed by the owner and the delegatecall is therefore safe.

```
1  contract Proxy {
2    address callee;
3    address owner;
4    constructor() {
5      callee = address(0x0);
6      owner = msg.sender;
7    }
8    function setCallee(address newCallee) {
9      require(msg.sender == owner);
10     callee = newCallee;
11   }
12   function forward(bytes _data) {
13     require(callee.delegatecall(_data));
14   }
15 }
```

Fig. 10. False positive reported by ILF on an usafe delegatecall.

Leaking Ether (LE). CONFUZZIUS, MYTHRIL, M-PRO and ILF detect the same number of ether leaking vulnerabilities, namely 4 out of 9. None of the tools report false positives.

Locking Ether (LO). Both, CONFUZZIUS and ILF detect all ether locking vulnerabilities. SFUZZ, on the other hand, does not detect any of the vulnerabilities.

Unprotected Selfdestruct (US). Both, CONFUZZIUS and ILF detect all the unprotected selfdestruct vulnerabilities. OYENTE does not detect any of the vulnerabilities and MYTHRIL as well as M-PRO, detect 2 of the 3 unprotected selfdestruct vulnerabilities.

C. Component Evaluation

In the following, we evaluate the importance of CONFUZZIUS's three main components: 1) *constraint solving*, 2) *read-after-write dependency analysis* and 3) *environmental instrumentation*. We randomly selected 100 contracts from each cluster. We then performed three experiments for each contract, where we deactivated a different component for each experiment. For example, in the "Without Constraint Solving" experiment, we disabled constraint solving but kept RAW dependency analysis, and environmental instrumentation enabled. This means that inputs will be generated following a random uniform distribution, and the constraint solver will not produce inputs. In the "Without Read-After-Write Dependency" experiment, we disabled RAW dependency analysis but kept constraint solving and environmental instrumentation enabled. This means that transaction sequences will not be combined following RAW dependencies, but that they will be combined at random, following a uniform distribution. Finally, in the "Without Environmental Instrumentation" experiment, we disabled environmental instrumentation but kept constraint solving and RAW dependency analysis enabled. This means that environmental information such as block timestamps or call return values will not be fuzzed. We repeated each experiment 10 times and run the experiments on the small contracts for 10 minutes and the large contracts for 1 hour. In Fig. 11, each bar depicts the percentage of the achieved results compared to the results when all three components were enabled (i.e. the gray bar in the back). We see that each component is an added value for CONFUZZIUS, which means that they are all essential to CONFUZZIUS's performance. However, we can see that generating meaningful inputs via constraint solving plays an essential part in achieving broad code coverage and detecting more bugs. Also, environmental instrumentation seems to help achieve code coverage and detect bugs. Nevertheless, our novel method that leverages RAW dependency analysis to create meaningful sequences of inputs at runtime can provide 2% more code coverage on small contracts and 5% on large contracts. Further, it allows the detectors to find 10% and 18% more vulnerabilities in small and large contracts, respectively.

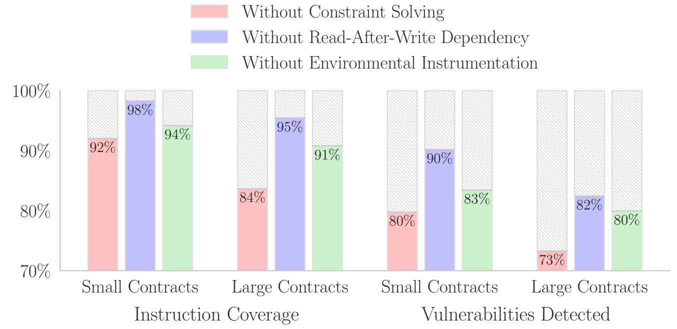

Fig. 11. Comparison of overall instruction coverage and vulnerabilities detected between CONFUZZIUS's individual components.

VI. RELATED WORK

Since its introduction by Miller et al. [30], fuzzing has been applied to many different domains and targets.

Software Fuzzing. American Fuzzy Loop (AFL) [29] is one of the most widespread fuzzers and it is based on evolutionary fuzzing and exploits execution data to guide the generation/-mutation of fuzzed inputs. Besides AFL and its offsprings [3], [4], other fuzzers also use evolutionary approaches to

generate test inputs automatically [20], [37]. On the other hand, KLEE [6] and SAGE [17] are white-box fuzzers, that execute code in a controlled environment. Driller [43] is a hybrid fuzzer that leverages selective concolic execution in a complementary manner. Symbolic execution based fuzzers produce meaningful inputs but tend to be slow [7]–[9], [35]. Fuzzers such as LibFuzzer [39], FuzzGen [21] and FUDGE [2] focus on fuzzing libraries, which cannot run as standalone programs, but instead are invoked by other programs.

Smart Contract Fuzzing. CONTRACTFUZZER [22] generates inputs based on a list of input seeds. While CONTRACTFUZZER deploys an entire custom testnet to fuzz transactions, CONFUZZIUS is more efficient and solely emulates the EVM. Moreover, CONFUZZIUS does not rely on user-provided input seeds but instead analyzes the execution traces and feeds constraints related to the execution to a constraint solver in order to generate new values specific to the contract under test. ECHIDNA [10] is a property-based testing tool for smart contracts that leverages grammar-based fuzzing. ECHIDNA requires user-defined predicates in the form of Solidity assertions and does not automatically check for vulnerabilities. HARVEY [50] predicts new inputs based on instruction-granularity cost metrics. In contrast, CONFUZZIUS exploits lightweight symbolic execution when the population fitness does not increase (see Section IV-D). Further, HARVEY fuzzes transaction sequences in a targeted and demand-driven way, assisted by an aggressive mode that directly fuzzes the persistent state of a smart contract. Instead, CONFUZZIUS relies on the read-after-write dependencies to guide the selection and crossover operators to create meaningful transaction sequences efficiently (see Section IV-B). ILF [18] is based on imitation learning, which requires a learning phase prior to fuzzing. ILF consists of a neural network that is trained on transactions obtained by running a symbolic execution expert over a broad set of contracts. CONFUZZIUS does not have the overhead of the learning phase and uses on-demand constraint solving while actively fuzzing the target. Moreover, ILF is limited to the knowledge that it learned during the learning phase, meaning that ILF has issues in getting past program conditions that require inputs that were not part of the learning dataset. CONFUZZIUS does not have this issue as it learns tailored inputs per target while fuzzing. SFUZZ [32] is an AFL based smart contract fuzzer, whereas ETHPLOIT [51] is a fuzzing based smart contract exploit generator. Both SFUZZ and ETHPLOIT have been developed concurrently and independently of CONFUZZIUS. While SFUZZ follows a random strategy to create transaction sequences, ETHPLOIT uses static taint analysis on state variables to create meaningful transaction sequences. However, static taint analysis has the disadvantage of being imprecise and analyzing parts that are not executable. Despite SFUZZ using a genetic algorithm as CONFUZZIUS, it employs a different encoding of individuals. SFUZZ only models block number and timestamp as environmental information. ETHPLOIT, on the other hand, instruments the EVM in a similar way to CONFUZZIUS. However, ETHPLOIT does not fuzz the size of external code or contract call return values.

Smart Contract Symbolic Execution. Apart from fuzzing, several other tools based on symbolic execution were proposed to assess the security of smart contracts [28] [33] [31] [45] [46] [46] [26] [16]. MPRO [52] combines symbolic execution and data dependency analysis to deal with the scalability issues that symbolic execution tools face when trying to handle the statefulness of smart contracts. MPRO has been developed concurrently and independently from CONFUZZIUS. There are two significant differences in our approach. First, MPRO retrieves data dependencies using static analysis and therefore requires source code, whereas CONFUZZIUS tries to infer data dependencies from bytecode. Second, MPRO works in two separate steps, first, it infers data dependencies via static analysis, and then it applies symbolic execution. CONFUZZIUS, on the other hand, applies a dynamic approach and infers data dependencies while fuzzing. ETHRACER [25] uses a hybrid approach with a converse strategy by primarily using symbolic execution to test a smart contract and using fuzzing only for producing combinations of transactions to detect vulnerabilities such as transaction order dependency. CONFUZZIUS's fuzzing strategy, compared to ETHRACER, is not entirely random but based on read-after-write dependencies, yielding faster and more efficient combinations.

Smart Contract Static Analysis. Besides symbolic execution and fuzzing, other works based on static analysis were proposed to detect smart contract vulnerabilities. ZEUS [23] is a framework for automated verification of smart contracts using abstract interpretation and model checking. SECURIFY [47] uses static analysis based on a contract's dependency graph to extract semantic information about the program bytecode and then checks for violations of safety patterns. Similarly, VANDAL [5] is a framework designed to convert EVM bytecode into semantic logic relations in Datalog, which can then be queried for vulnerabilities.

VII. CONCLUSION

We presented CONFUZZIUS, the first hybrid fuzzer for smart contracts. It tackles the three main challenges of smart contract testing: *input generation*, *stateful exploration*, and *environmental dependencies*. We solved the first challenge by combining evolutionary fuzzing with constraint solving to generate meaningful inputs. The second challenge is solved by leveraging data dependency analysis across state variables to generate purposeful transaction sequences. Finally, we solved the third challenge by modeling block related information (e.g. block number) and contract related information (e.g. call return values) as fuzzable inputs. We run CONFUZZIUS and other state-of-the-art fuzzers and symbolic execution tools for smart contracts against a curated dataset of 128 contracts and a dataset of 21K real-world smart contracts. Our results show that our hybrid approach not only detects more bugs than existing state-of-the-art tools (up to 23%), but also that CONFUZZIUS outperforms these tools in terms of code

coverage (up to 69%) and that data dependency analysis can boost the detection of bugs (up to 18%).

Acknowledgment

We would like to thank the anonymous reviewers for their valuable comments and feedback. This work was partly supported by the Luxembourg National Research Fund (FNR) under grant 13192291 and is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830927.

References

[1] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on ethereum smart contracts (sok)," in *International Conference on Principles of Security and Trust*. Springer, 2017, pp. 164–186.

[2] D. Babić, S. Bucur, Y. Chen, F. Ivančić, T. King, M. Kusano, C. Lemieux, L. Szekeres, and W. Wang, "Fudge: fuzz driver generation at scale," in *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2019, pp. 975–985.

[3] M. Böhme, V.-T. Pham, M.-D. Nguyen, and A. Roychoudhury, "Directed greybox fuzzing," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 2329–2344.

[4] M. Böhme, V.-T. Pham, and A. Roychoudhury, "Coverage-based greybox fuzzing as markov chain," *IEEE Transactions on Software Engineering*, vol. 45, no. 5, pp. 489–506, 2017.

[5] L. Brent, A. Jurisevic, M. Kong, E. Liu, F. Gauthier, V. Gramoli, R. Holz, and B. Scholz, "Vandal: A scalable security analysis framework for smart contracts," *arXiv preprint arXiv:1809.03981*, 2018.

[6] C. Cadar, D. Dunbar, D. R. Engler *et al.*, "KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs." in *OSDI*, vol. 8, 2008, pp. 209–224.

[7] P. Chen and H. Chen, "Angora: Efficient fuzzing by principled search," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 711–725.

[8] P. Chen, J. Liu, and H. Chen, "Matryoshka: fuzzing deeply nested branches," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 499–513.

[9] M. Cho, S. Kim, and T. Kwon, "Intriguer: Field-level constraint solving for hybrid fuzzing," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 515–530.

[10] Crytic, "Echidna: Ethereum fuzz testing framework," February 2020. [Online]. Available: https://github.com/crytic/echidna

[11] L. De Moura and N. Bjørner, "Z3: An efficient smt solver," in *International conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2008, pp. 337–340.

[12] T. Durieux, J. F. Ferreira, R. Abreu, and P. Cruz, "Empirical review of automated analysis tools on 47,587 ethereum smart contracts," in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 2020, pp. 530–541.

[13] Ethereum Foundation, "Py-EVM - A Python implementation of the Ethereum Virtual Machine," August 2019. [Online]. Available: https://github.com/ethereum/py-evm

[14] Etherscan, "Contracts With verified source codes only," November 2020, https://etherscan.io/contractsVerified.

[15] V. B. Fabian Vogelsteller, "Erc-20 token standard," February 2015, https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md.

[16] J. Frank, C. Aschermann, and T. Holz, "ETHBMC: A bounded model checker for smart contracts," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[17] P. Godefroid, M. Y. Levin, D. A. Molnar *et al.*, "Automated whitebox fuzz testing." in *NDSS*, vol. 8, 2008.

[18] J. He, M. Balunović, N. Ambroladze, P. Tsankov, and M. Vechev, "Learning to fuzz from symbolic execution with application to smart contracts," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '19. New York, NY, USA: ACM, 2019, pp. 531–548. [Online]. Available: http://doi.acm.org/10.1145/3319535.3363230

[19] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2011.

[20] A. K. Iannillo, R. Natella, D. Cotroneo, and C. Nita-Rotaru, "Chizpurfle: A gray-box android fuzzer for vendor service customizations," in *2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2017, pp. 1–11.

[21] K. Ispoglou, D. Austin, V. Mohan, and M. Payer, "Fuzzgen: Automatic fuzzer generation," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020.

[22] B. Jiang, Y. Liu, and W. Chan, "Contractfuzzer: Fuzzing smart contracts for vulnerability detection," in *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. ACM, 2018, pp. 259–269.

[23] S. Kalra, S. Goel, M. Dhawan, and S. Sharma, "Zeus: Analyzing safety of smart contracts," in *NDSS*, 2018.

[24] G. Klees, A. Ruef, B. Cooper, S. Wei, and M. Hicks, "Evaluating fuzz testing," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018, pp. 2123–2138.

[25] A. Kolluri, I. Nikolic, I. Sergey, A. Hobor, and P. Saxena, "Exploiting the laws of order in smart contracts," *arXiv preprint arXiv:1810.11605*, 2018.

[26] J. Krupp and C. Rossow, "teether: Gnawing at ethereum to automatically exploit smart contracts," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 1317–1333.

[27] G. E. Liepins and M. D. Vose, "Representational issues in genetic optimization," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 2, no. 2, pp. 101–115, 1990.

[28] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: ACM, 2016, pp. 254–269. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978309

[29] Michal Zalewski, "American Fuzzy Lop (AFL)," December 2016. [Online]. Available: http://lcamtuf.coredump.cx/afl/

[30] B. P. Miller, L. Fredriksen, and B. So, "An Empirical Study of the Reliability of UNIX Utilities," *Communications of the ACM*, vol. 33, no. 12, pp. 32–44, 1990.

[31] B. Mueller, "Smashing ethereum smart contracts for fun and real profit," in *9th annual HITB Security Conference*, 2018.

[32] T. D. Nguyen, L. H. Pham, J. Sun, Y. Lin, and Q. T. Minh, "sfuzz: An efficient adaptive fuzzer for solidity smart contracts," *arXiv preprint arXiv:2004.08563*, 2020.

[33] I. Nikolic, A. Kolluri, I. Sergey, P. Saxena, and A. Hobor, "Finding the greedy, prodigal, and suicidal contracts at scale," *arXiv preprint arXiv:1802.06038*, 2018.

[34] T. of Bits, "Manticore - symbolic execution tool," jun 2018, https://github.com/trailofbits/manticore.

[35] H. Peng, Y. Shoshitaishvili, and M. Payer, "T-fuzz: fuzzing by program transformation," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 697–710.

[36] S. Petrov, "Another parity wallet hack explained," nov 2017, https://medium.com/@Pr0Ger/another-parity-wallet-hack-explained-847ca46a2e1c.

[37] S. Rawat, V. Jain, A. Kumar, L. Cojocar, C. Giuffrida, and H. Bos, "Vuzzer: Application-aware evolutionary fuzzing." in *NDSS*, vol. 17, 2017, pp. 1–14.

[38] C. Reitwiessner, "Solidity 0.6.6 Documentation: Layout of State Variables in Storage," July 2020, https://solidity.readthedocs.io/en/v0.6.6/miscellaneous.html#layout-of-state-variables-in-storage.

[39] K. Serebryany, "Continuous fuzzing with libfuzzer and addresssanitizer," in *2016 IEEE Cybersecurity Development (SecDev)*. IEEE, 2016, pp. 157–157.

[40] D. Siegel, "Understanding the dao attack," jun 2016, https://www.coindesk.com/understanding-dao-hack-journalists/.

[41] R. Sivaraj and T. Ravichandran, "A review of selection methods in genetic algorithm," *International journal of engineering science and technology*, vol. 3, no. 5, pp. 3792–3797, 2011.

[42] SmartContractSecurity, "SWC Registry," November 2020, https://swcregistry.io/.

[43] N. Stephens, J. Grosen, C. Salls, A. Dutcher, R. Wang, J. Corbetta, Y. Shoshitaishvili, C. Kruegel, and G. Vigna, "Driller: Augmenting fuzzing through selective symbolic execution." in *NDSS*, vol. 16, no. 2016, 2016, pp. 1–16.

[44] S. Tikhomirov, E. Voskresenskaya, I. Ivanitskiy, R. Takhaviev, E. Marchenko, and Y. Alexandrov, "Smartcheck: Static analysis of ethereum smart contracts," in *Proceedings of the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain*, 2018, pp. 9–16.
[45] C. F. Torres, J. Schütte, and R. State, "Osiris: Hunting for integer bugs in ethereum smart contracts," in *Proceedings of the 34th Annual Computer Security Applications Conference*, ser. ACSAC '18. New York, NY, USA: ACM, 2018, pp. 664–676. [Online]. Available: http://doi.acm.org/10.1145/3274694.3274737
[46] C. F. Torres, M. Steichen *et al.*, "The art of the scam: Demystifying honeypots in ethereum smart contracts," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 1591–1607.
[47] P. Tsankov, A. Dan, D. Drachsler-Cohen, A. Gervais, F. Buenzli, and M. Vechev, "Securify: Practical security analysis of smart contracts," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2018, pp. 67–82.
[48] G. Wood, "Solidity 0.6.11 Documentation," July 2020, https://solidity.readthedocs.io/en/v0.6.11/.
[49] G. Wood *et al.*, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1–32, 2014.
[50] V. Wüstholz and M. Christakis, "Harvey: A greybox fuzzer for smart contracts," *arXiv preprint arXiv:1905.06944*, 2019.
[51] Q. Zhang, Y. Wang, J. Li, and S. Ma, "Ethploit: From fuzzing to efficient exploit generation against smart contracts," in *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2020, pp. 116–126.
[52] W. Zhang, S. Banescu, L. Pasos, S. Stewart, and V. Ganesh, "Mpro: Combining static and symbolic analysis for scalable testing of smart contract," in *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2019, pp. 456–462.

APPENDIX A
VULNERABILITY DETECTORS

We elaborate on the implementation details of our 10 vulnerability detectors below.

Assertion Failure. We detect an assertion failure by checking if the execution trace contains an `ASSERTFAIL` or `INVALID` instruction.

Integer Overflow. Detecting integer overflows is not trivial, since not every overflow is considered harmful. Integer overflows may also be introduced by the compiler for optimization purposes. Therefore, we only consider an overflow as harmful, if it modifies the state of the smart contract, i.e. if the result of the computation is written to storage or is used to send funds. We follow the approach by Torres et al. [45] and start by analyzing if the execution trace contains an `ADD`, `MUL` or `SUB` instruction. We then extract the operands from the stack and use these to compute the result of the arithmetic operation. Afterwards, we check if our result is equivalent to the result that has been pushed onto the stack. If they are not the same, we know that an integer overflow has occurred and we keep track of the overflow by tainting the result of the computation. We report an integer overflow if the tainted result flows into an `SSTORE` instruction or a `CALL` instruction, as these instructions will result in updating the blockchain state.

Reentrancy. A reentrancy occurs whenever a contract calls another contract, and that contract calls back the original contract. We detect reentrancy by first checking if the execution trace contains a `CALL` instruction whose gas value is larger than 2,300 units and where the amount of funds to be transferred is a symbolic value or a concrete value that is larger than zero. Finally, we report a reentrancy if we find an `SLOAD` instruction that occurs before the `CALL` instruction and an `SSTORE` instruction that occurs after the `CALL` instruction, and which shares the same storage location as the `SLOAD` instruction.

Transaction Order Dependency. We detect transaction order dependency by checking if there are two execution traces with different senders, where the first execution trace writes to the same storage location from which the second execution trace reads.

Block Dependency. We detect a block dependency by checking if the execution trace contains either a `CREATE`, `CALL`, `DELEGATECALL`, or `SELFDESTRUCT` instruction, that is either control-flow or data dependent on a `BLOCKHASH`, `COINBASE`, `TIMESTAMP`, `NUMBER`, `DIFFICULTY`, or `GASLIMIT` instruction.

Unhandled Exception. We detect unhandled exceptions by first checking if the execution trace contains a `CALL` instruction that pushes to the stack the value 1 as a result of the call. A value of 1 means that an error occurred during the call (i.e. an exception). Afterwards, we check if the result of the call flows into a `JUMPI` instruction. If the result does not flow into a `JUMPI` instruction until the end of the execution trace, then this means that the exception of the call was not handled and we report an unhandled exception.

Unsafe Delegatecall. We detect an unsafe delegate call by checking if there is an execution trace that contains a `DELEGATECALL` instruction and terminates with a `STOP` instruction, but whose sender is an attacker address. Attacker and benign user addresses are generated at the start by the fuzzer.

Leaking Ether. We detect the leaking of ether by checking if the execution trace contains a `CALL` instruction, whose recipient is an attacker address that has never sent ether to the contract in a previous transaction and has never been passed as a parameter in a function by an address that is not an attacker.

Locking Ether. We detect the locking of ether by checking if a contract can receive ether but cannot send out ether. To check if a contract cannot send ether, we check if the runtime bytecode of the contract does not contain any `CREATE`, `CALL`, `DELEGATECALL`, or `SELFDESTRUCT` instruction. To check if a contact can receive ether, we check if the execution trace has a transaction value larger than 0 and terminates with a `STOP` instruction.

Unprotected Selfdestruct. Similar to the leaking ether or unsafe delegatecall vulnerability detectors, this detector relies on attacker accounts. We detect an unprotected selfdestruct by

checking if the execution trace contains a `SELFDESTRUCT` instruction where the sender of the transaction is an attacker and its address has not been previously passed as an argument by a benign user.