

FlashSyn: Flash Loan Attack Synthesis via Counter Example Driven Approximation

ZHIYANG CHEN, University of Toronto, Canada

SIDI MOHAMED BEILLAH, University of Toronto, Canada

FAN LONG, University of Toronto, Canada

In decentralized finance (DeFi) ecosystem, lenders can offer flash loans to borrowers, i.e., loans that are only valid within a blockchain transaction and must be repaid with some fees by the end of that transaction. Unlike normal loans, flash loans allow borrowers to borrow a large amount of assets without upfront collaterals deposits. Malicious adversaries can use flash loans to gather large amount of assets to launch costly exploitations targeting DeFi protocols.

In this paper, we introduce a new framework for automated synthesis of adversarial contracts that exploit DeFi protocols using flash loans. To bypass the complexity of a DeFi protocol, we propose a new technique to approximate the DeFi protocol functional behaviors using numerical methods (polynomial linear regression and nearest-neighbor interpolation). We then construct an optimization query using the approximated functions of the DeFi protocol to find an adversarial attack constituted of a sequence of functions invocations with optimal parameters that gives the maximum profit. To improve the accuracy of the approximation, we propose a new counterexamples-driven approximation refinement technique. We implement our framework in a tool called FlashSyn. We evaluate FlashSyn on 12 DeFi protocols that were victims to flash loan attacks and DeFi protocols from Damn Vulnerable DeFi challenges. FlashSyn automatically synthesizes an adversarial attack for each one of them.

1 INTRODUCTION

Blockchain technology provides decentralized, robust, and programmable ledgers at Internet scale. Developers can deploy smart contracts onto a blockchain to encode arbitrarily complicated transaction rules that operate on the ledger. Blockchains and smart contracts have become the key infrastructure for various decentralized financial services (DeFi). The Total Value Locked (TVL) in 1417 DeFi smart contracts has reached 85.64 billion USD on June 1st 2022 [DefiLlama 2022].

Security attacks are critical threats to smart contracts. Attackers can send malicious transactions to exploit vulnerabilities of a contract to steal millions of dollars from the contract users. Particularly, a new kind of security attacks that use large amount of digital assets to exploit design flaws of DeFi contracts has become prevalent. Because such attacks typically borrow the used assets from flash loan contracts, they are called *flash loan attacks* [McKay 2022; SlowMist 2022]. Among the top 200 costliest attacks recorded in Rekt Database, the financial loss caused by in total 36 flash loan attacks is exceeding 418 million dollars [McKay 2022].

A malicious flash loan attack transaction typically contains a sequence of actions (i.e., function calls to smart contracts). The first action borrows a very large sum of digital assets from a flash loan contract and the last action returns the borrowed assets. The sequence of actions in the middle interacts with multiple DeFi contracts using the borrowed assets to exploit their design flaws. When a DeFi contract fails to consider corner cases caused by the large sum of borrowed assets, the attacker may extract prohibitive profits. For example, many flash loan attacks use borrowed assets to temporarily manipulate asset prices in a DeFi contract to fool the contract to make unfavorable

Authors' addresses: Zhiyang Chen, University of Toronto, Toronto, Ontario, Canada, zhiychen@cs.toronto.edu; Sidi Mohamed Beillahi, University of Toronto, Toronto, Ontario, Canada, sm.beillahi@utoronto.ca; Fan Long, University of Toronto, Toronto, Ontario, Canada, fanl@cs.toronto.edu.

2022. 2475-1421/2022/1-ART1 \$15.00

<https://doi.org/>

trades with the attacker [Cao et al. 2021; Qin et al. 2021]. Although researchers have developed many automated program analysis and verification techniques [Brent et al. 2020; Feng et al. 2019; Grieco et al. 2020; Mossberg et al. 2019; Wang et al. 2019] to detect and eliminate bugs in smart contracts, these techniques cannot handle flash loan attack vulnerabilities. This is because such vulnerabilities are design flaws rather than implementation bugs. Moreover, these techniques typically operate with one contract at a time, but flash loan attacks almost always involve multiple DeFi contracts interacting with each other.

FlashSyn: This paper presents FlashSyn, the first automated end-to-end program synthesis tool for detecting flash loan attack vulnerabilities. Given a set of smart contracts and candidate actions on these contracts, FlashSyn automatically synthesizes an action sequence along with all action parameters to interact with these contracts to exploit potential flash loan vulnerabilities.

The main challenge FlashSyn faces is that the underlying logic of DeFi actions is often too sophisticated for standard solvers or optimizers to handle. Even if the action sequence was already known, a naive approach such as symbolic execution might not be able to find action parameters because it might extract an overly complicated symbolic constraints causing the underlying solvers to time out. Moreover, FlashSyn synthesizes the action sequence and the action parameters together and therefore faces an additional search space explosion challenge.

FlashSyn addresses these challenge with its novel *counter-example driven approximation refinement* technique. Instead of attempting to extract symbolic expressions to exactly match the logic of actions, FlashSyn collects data points to approximate the effect of actions with polynomials and interpolations. FlashSyn then uses the approximated expressions to drive the synthesis. If the synthesis fails because of a large deviation caused by the approximations, FlashSyn collects the corresponding data points as counter examples to iteratively refine the approximations. This novel technique allows the underlying optimizer of FlashSyn to work with more tractable expressions to avoid the awkward timeout situation. It also decouples the two difficult tasks, finding the action sequence and finding the action parameters. When working with a set of coarse-grained approximated expressions, FlashSyn can filter out unproductive action sequences with a small cost.

Experimental results: We evaluate FlashSyn on 5 DeFi protocols that were victims to flash loan attacks and DeFi protocols from Damn Vulnerable DeFi challenges. FlashSyn automatically synthesizes an adversarial attack for each one of them. On average FlashSyn takes 555 seconds to finish the synthesis task.

Contributions: This paper makes the following contributions:

- **FlashSyn:** This paper presents FlashSyn, the first automated end-to-end program synthesis tool for detecting flash loan attack vulnerabilities.
- **Counter-example Driven Approximation Refinement:** This paper presents the novel counter-example driven approximation refinement technique to handle sophisticated logic of DeFi contract actions.

The rest of the paper is organized as follows. Section 2 presents a motivating example to give a high-level overview of FlashSyn. Section 3 briefly introduces background information. In Section 4, we define the problem we tackle in this paper. Section 5 presents our core synthesis algorithm and the counter-example driven approximation refinement technique. Section 6 describes FlashSyn, an implementation of our techniques. We evaluate FlashSyn on 12 DeFi protocols that were victims to flash loan attacks from three blockchains and DeFi protocols from Damn Vulnerable DeFi challenges. In Section 8 we discuss related work. Section 9 concludes the paper.

2 MOTIVATING EXAMPLE

In this section, we present a motivating example to describe our approach.

2.1 History

Background: On October 26th 2020, an attacker maliciously exploited the Harvest Finance protocol. The exploit stole \$24M of USDT [Limited 2022] and USDC [Consortium 2022] tokens from the USDC vault and the USDT vault of Harvest Finance. After borrowing large amounts of USDC and USDT from flash loan providers, the attacker performed the same attack vector multiple times, draining the USDC vault of Harvest Finance first. The attack vector is shown below:

Preparation flash loan of 18.3M USDT and 50M USDC.¹

Action 1 `exchange_underlying(1, 2, 10554172e6, 0)`²

Action 2 `deposit(49972546e6)`

Action 3 `exchange_underlying(2, 1, 10564726e6, 0)`

Action 4 `withdraw(51543726e6)`

The function `exchange_underlying` requires four parameters, the two first parameters identify the token to swap (1 for USDC and 2 for USDT), the third parameter specifies the quantity to swap, and the last parameter specifies the minimal quantity expected to receive from the swapping.

In the actual attack, Action 1 swapped 10554172 USDC to 10537272 USDT in Curve Y Pool, which raised the USDT/USDC exchange rate. This results in Harvest Finance's evaluation of its invested underlying assets reduced. Action 2 deposited 49972546 USDC into USDC vault using the function `deposit` and due to the reduced evaluation of invested underlying assets, the attacker received 51543825 fUSDC back, which is abnormally larger than when the USDT/USDC rate is 1. Action 3 swapped 10564726 USDT to 10573194 USDC, which normalized the manipulated USDT/USDC rate. It also brings Harvest Finance's evaluation of its invested underlying asset back to normal. Finally, Action 4 withdrew 50319524 USDT (equivalent to 51543825 fUSDC) from the USDC vault. If we assume the market prices of USDC and USDT are both equal to 1 USD, the adjusted profit of the above attack vector is 338448 in USD.

This attack is a typical case of oracle manipulation. The exploiter manipulated the USDT/USDC rate in Curve Y Pool by swapping a large amount between USDC and USDT back and forth, which caused Harvest Finance protocol to wrongly evaluate the value of its underlying asset, leaving large arbitrage space for the exploiter.

2.2 Challenges

We now use the above attack vector as an example to illustrate the challenges of synthesizing such attack vectors and its corresponding parameters.

Sophisticated interactions: DeFi smart contracts often interact with other smart contracts inside or outside the same DeFi protocol. The change of states of one smart contract will possibly influence the effect of another smart contract. For example, `exchange_underlying` only changes the states of Y Pool on Curve Finance. However, since `deposit` and `withdraw` use the states of Y Pool to calculate the amount to deposit and withdraw, their effect is influenced by the states of Curve Finance. This makes the synthesis problem more complicated as the effect of an action is determined by its predecessor actions thus each action cannot be treated separately.

Close source: While DeFi smart contracts often interact with external smart contracts, not all of them are open-sourced. For example, the source code of the external smart contract `PriceConverter`³ of Harvest Finance protocol is not available on Etherscan [Etherscan 2022], but it is called by Action

¹This step is same for all attack vectors in this Section, and omitted later for brevity.

²Solidity supports scientific notations. The literal `MeE` is equivalent to $M * 10^E$.

³<https://etherscan.io/address/0xfca4416d9def20ac5b6da8b8b322b6559770efbf#code>

2 deposit and Action 4 withdraw to query states of Y Pool on Curve Finance⁴. This close source problem impedes reading and understanding the DeFi protocol implementation.

Mathematical complexity: In some smart contracts, the underlying mathematical model is too complicated to reason about even if every smart contract is open-sourced. For instance, the `exchange_underlying`, which is part of the attack vector, swaps an amount of token i to token j , while maintaining the following StableSwap invariant [Egorov 2019]:

$$A \cdot n^n \sum_i x_i + D = A \cdot n^n \cdot D + \frac{D^{n+1}}{n^n \prod_i x_i}$$

where A is a constant, n is the number of different types of tokens inside the pool (= 4 for Y Pool), x_i represents the liquidities of token i , D represents the total amount of coins when they have an equal price. Obviously there does not exist a closed-form solution for D as it requires finding roots of a quintic equation. Instead, in the actual smart contract, when `exchange_underlying(i, j, dx, 0)` is called, D is first calculated iteratively to converge to a solution which maintains the above invariant, using the liquidities before calling the function. Next x_i is increased subject to dx . Then, using the known D and the updated x_i , the new x_j is again calculated iteratively until the above invariant is satisfied again. The related source code is shown in Figures 6 and 7 in Appendix. This kind of complex underlying mathematical models makes it hard to understand the exact execution results.

2.3 Failure of Symbolic Execution

The standard approach to search for such exploits is to use symbolic execution [King 1976] to explore state space of related programs and check whether pre-defined invariants can be violated [Consensys 2022; Feng et al. 2019; Luu et al. 2016; Mossberg et al. 2019]. However, due to the complexity of DeFi protocols, it is difficult for these symbolic execution based tools to reach all program states, which might extract an overly complex symbolic constraints and easily exhaust the underlying solvers to time out. For instance, to test the applicability of symbolic execution, we tried to symbolically execute an internal function `get_D` (see Figure 1), which is called inside the function `exchange_underlying`. We use Manticore [Mossberg et al. 2019] to execute `get_D` with symbolic inputs and explore all possible states it can reach. Manticore fails and throws a solver-related exception together with an out of memory error. We then simplify `get_D` by removing the for loop and bounding the length of `xp` to 2, Manticore still fails and throws the same error.

2.4 Apply our approach

We now use the above example to describe our proposed approach and show how it tackles the problem of synthesizing attack vectors for DeFi protocols.

2.4.1 Identify Candidate Actions. Our approach first requires users to provide a set of candidate actions (i.e. functions inside smart contracts). Users can use placeholders to replace the parameters to be searched. However, the parameters replaced by placeholders are required to be the ones which might (continuously) change the amount of input/output tokens or smart contract states. The parameters which specify the type of input/output tokens and states to be changed should be given by the user.

For instance, a set of candidate actions for the Harvest USDC attack can be as follows: `{exchange_underlying(1, 2, -, 0), exchange_underlying(2, 1, -, 0), deposit(-), withdraw(-)}`. Recall for the function `exchange_underlying`, the first two input arguments specify the token types to be swapped. Thus, `exchange_underlying(1, 2, -, 0), exchange_underlying(2, 1,`

⁴<https://medium.com/haechi-audit/harvest-finance-economic-attack-deep-dive-part-1-420bc5d5ac0c>

```

1 function get_D(uint256[] calldata xp) returns (uint256) {
2     uint N_COINS = xp.length;
3     uint256 S = sum(xp);
4     // ...
5     uint256 D = S;
6     uint256 Ann = A * N_COINS; // A: a constant selected by the owner of the contract
7     for(uint _i = 0; _i < 255; _i = _i + 1){
8         uint256 D_P = D;
9         for(uint _j = 0; _j < xp.length; _j = _j + 1){
10             D_P = D_P * D / (xp[_j] * N_COINS + 1); // +1 to prevent /0
11         }
12         uint256 Dprev = D;
13         D = (Ann * S + D_P * N_COINS) * D / ((Ann - 1) * D + (N_COINS + 1) * D_P);
14         if(abs(D - Dprev) <= 1) { // Equality with the precision of 1
15             break;
16         }
17     }
18     return D;
19 }

```

Fig. 1. Source code of `get_D()`. Original code is in Vyper, rewritten in Solidity.

$-$, \emptyset) are treated as two separate action candidates. The fourth parameter specifies the minimal quantity expected to receive, which is not important under our settings and thus assumed to be zero.

2.4.2 Identify Input/Output States and Tokens. Our approach infers approximated state transition functions for each possible action in an attack vector over a predetermined set of states using numerical methods. This allows to abstract internal interactions among DeFi protocols and the associated complexity, and eliminates the requirements of open source smart contracts.

We use s_1 , s_2 , s_3 , s_4 , and s_5 to represent the key states in the Harvest USDC attack and in Table 1 we list the set of prestates and poststates for each action. States s_1 and s_2 represent USDC and USDT liquidities in the Curve.Fi Y Pool. States s_3 , s_4 , and s_5 represent the balance of the vault, the invested balance of the vault, and fUSDT total supply, respectively. Actions `exchange_underlying(1, 2, -, 0)` and `exchange_underlying(2, 1, -, 0)` change the liquidities of USDC(s_1) and USDT(s_2). This also results in Harvest Finance changing its evaluation of the value of its invested balance (s_4). Action `deposit(-)` deposits USDC to the vault(changes s_3), and the vault mints(changes s_5) and sends back fUSDT in return. Action `withdraw(-)` returns fUSDT back to the vault, burns them(changes s_5) and withdraws USDC from the vault(changes s_3). The exchange rate of fUSDT and USDC within `deposit` and `withdraw` are dependent on Harvest Finance protocol's balance in the vault (s_3), the invested Balance (s_4), and fUSDT total supply (s_5), as shown in Figure 2.

Action	Prestates	Token In	Poststates	Token Out
<code>exchange_underlying(1, 2, -, 0)</code>	s_1, s_2, s_4	USDC	s_1, s_2, s_4	USDT
<code>exchange_underlying(2, 1, -, 0)</code>	s_1, s_2, s_4	USDT	s_1, s_2, s_4	USDC
<code>deposit(-)</code>	s_3, s_4, s_5	USDC	s_3, s_4, s_5	fUSDT
<code>withdraw(-)</code>	s_3, s_4, s_5	fUSDT	s_3, s_4, s_5	USDC

Table 1. Prestates and poststates of actions in Harvest USDC attack

```

1 function deposit(uint256 amount) {
2     //...
3     uint256 underlyingBalanceWithInvestment = underlingBalanceInVault() + investedUnderlyingBalance();
4     uint256 toMint = amount.mul(totalSupply()).div(underlyingBalanceWithInvestment);
5     _mint(msg.sender, toMint);
6     USDC.safeTransferFrom(msg.sender, address(this), amount);
7 }
8 function withdraw(uint256 numberOfShares) {
9     //...
10    uint256 underlyingBalanceWithInvestment = underlingBalanceInVault() + investedUnderlyingBalance();
11    uint256 totalSupply = totalSupply();
12    uint256 underlyingAmountToWithdraw = underlyingBalanceWithInvestment
13        .mul(numberOfShares)
14        .div(totalSupply);
15    //...
16    USDC.safeTransfer(msg.sender, underlyingAmountToWithdraw);
17 }

```

Fig. 2. Source code of deposit() and withdraw() in Vault.sol(with irrelevant details omitted)

2.4.3 Capture Initial Approximation. In our proposed approach, for each action in Table 1 we infer a state transition function that accepts the action's input parameters and transforms the prestates to poststates and returns the action's output parameters. For instance, the actions corresponding to exchange_underlying change the states s_1 , s_2 , and s_4 . Our approach infers expressions relating the poststates s'_1 , s'_2 , and s'_4 and the outputted amount of USDC to the prestates s_1 , s_2 , and s_4 and the input parameter of exchange_underlying. Similarly, the actions associated with deposit and withdraw alter the states s_3 , s_4 , and s_5 of Harvest Finance.

We first collect a set of data points where each data point is defined as an *input-output* pair, where the *input* is the action's prestates and its input parameters, and the *output* is poststates and the outputted values. For instance for the Harvest USDC attack, we fork the Ethereum Block 11129499, which is one block before the block of the attack. We execute each action a with different input parameters sampled from a fixed range of possible values (e.g., $(0, u)$ where u is an upper limit of the parameters given by the user) and different prestates. For each successful execution we record the obtained some *input-output* pairs. Then, we use these data points to find the best approximated state transition functions. We consider two methods to solve the above multivariate approximation problem: linear regression based polynomial features method and nearest-neighbor interpolation method.

In the first method, we generate a feature matrix consisting of all polynomial combinations of the features with degree less than or equal to n . Then, we use linear regression to find the optimal coefficients for the polynomial function f . In the second method, we build a nearest- N -dimensional interpolator based on the input-output pairs.

For the action associated with the function exchange_underlying(1, 2, -, 0), using the collected data points and polynomial approximation, we obtain the following approximated transition function associated with the poststate s'_1 :

$$\begin{aligned}
 s'_1 = & 1.000005226565857 * s_1 + 4.631021268852731 * 10^{-6} * s_2 \\
 & + 0.8591736013991388 * p - 536.4108634740114
 \end{aligned}$$

where p is the third input parameter passed to the function exchange_underlying, i.e., exchange_underlying(1, 2, p , 0).

Table 2. Number of Data Points for Harvest USDC(**Start** represents the moment after initial pass of data collection. **End(X)** represents at the end of synthesis using **X** method)

Action	Start	End(Polynomial)	End(Interpolation)
exchange_underlying(1, 2, -, 0)	8171	8533	8545
exchange_underlying(2, 1, -, 0)	8623	9032	9028
deposit(-)	7413	7676	7477
withdraw(-)	4508	5263	4659

2.4.4 Synthesize Action Sequence. After capturing an initial approximation of state transition functions, FlashSyn leverages an enumeration-based top-down algorithm to synthesize different action sequences, with several pruning heuristics applied to eliminate infeasible ones(details in Section 5.1). For each feasible action sequence, using the approximated state transition functions, FlashSyn automatically constructs an optimization framework(details in Section 5.4), consisting of an objective function(which represents profit) and constraints. Then FlashSyn applies an off-the-shelf optimizer to solve the optimization framework, and get a list of parameters that maximize the profit estimated using approximated transition functions.

2.4.5 Verify Synthesized Results. After getting a list of parameters that maximize the estimated profit of one action sequence, FlashSyn verifies synthesized attack vectors by executing them in a forked environment and checking their actual profit. For an attack vector if its actual profit is positive and close to the estimated profit, FlashSyn considers it as correct and include them in the final return output. However, inevitably there are some cases where the actual profit is different from the estimated profit, which indicates inaccuracy of our approximated transition functions due to the limited state space explored during the initial round of data collection.

2.4.6 Counter-Example Driven Refinement. To overcome the above challenge we use a novel *counterexamples-driven approximation refinement* technique. In particular, for a given attack vector, if its estimated profit is different from the actual profit obtained by executing the attack vector, we report it as a counterexample. Then, we collect new data points using these counterexamples. This allows to correct the approximated functions and increase the accuracy. In the Harvest USDC example, the number of data points collected during different stages is listed in Table 2.

Using our approach, we successfully find the following attack vector with a profit of 125843 in USD:

Action 1 exchange_underlying(1, 2, 9325577e6, 0)
Action 2 deposit(48501177e6)
Action 3 exchange_underlying(2, 1, 10423054e6, 0)
Action 4 withdraw(49745995e6)

The synthesized attack vector has the same action sequence and similar parameters as used by the exploiter in history. Such profitable attack vectors are direct proofs of the vulnerabilities inside Harvest Finance protocols, as they imply financial loss of corresponding DeFi smart contracts.

3 BACKGROUND

This section briefly gives introductory descriptions of blockchain, smart contracts, and decentralized finance.

```

1 pragma solidity ^0.8.0;
2 // ...
3 contract ERC20 {
4     mapping (address => uint) _balances;
5     event Transfer(address from, address to, uint amount);
6     // ...
7     function transfer(address from, address to, uint amount) public {
8         require(from != address(0), "ERC20: transfer from the zero address");
9         require(to != address(0), "ERC20: transfer to the zero address");
10        uint fromBalance = _balances[from];
11        require(fromBalance >= amount, "ERC20: transfer amount exceeds balance");
12        _balances[from] = fromBalance - amount;
13        _balances[to] += amount;
14        emit Transfer(from, to, amount);
15    }
16    // ...
17 }

```

Fig. 3. An excerpt of ERC20 smart contract.

3.1 Blockchain

Blockchain is a distributed ledger that broadcasts and stores information of transactions across different parties. A blockchain consists of a growing number of blocks and a consensus algorithm which decides the order of blocks. Each block is constituted of transactions. While Satoshi Nakamoto invented the first decentralized cryptocurrency Bitcoin in 2008 [Nakamoto 2008], the Ethereum blockchain [Wood et al. 2014] is the first blockchain to support, store, and execute Turing complete programs, known as smart contracts. In particular, the Ethereum blockchain is constituted of a global state and transactions that modify the state. Ethereum supports two types of accounts: user accounts and smart contract accounts, and each account is associated with a unique address. New popular blockchains, e.g., Binance Smart Chain [Chain 2020], Avalanche [Sekniqi et al. 2020], Conflux [Li et al. 2020], and Near [Foundation 2021], Fantom [Choi et al. 2018] work with Ethereum virtual machine (EVM) based execution environment due to the popularity of the EVM stack for developers.

3.2 Smart Contracts

Users interact with the blockchain by issuing transactions using their user accounts. Smart contracts are programs created to automatically execute those transactions. Each smart contract is associated with a unique address, a persistent account's storage trie, a balance of native tokens, e.g., Ether in Ethereum and BNB in Binance Smart Chain, and bytecode (e.g., EVM bytecode [Wood et al. 2014]) that executes incoming transactions to change the storage and balance. Smart contracts become immutable once deployed to the blockchain. Smart contracts can create internal transactions to interact with other smart contracts. The internal transactions are nested from a top-level transaction, an external transaction, initiated by a user account. If a transaction is aborted then the effects of all nested internal transactions will be reversed.

Currently, there exist several human-readable high-level programming languages, e.g., Solidity [Foundation 2022b], Vyper [Foundation 2022c], Serpent [Foundation 2022a], and Yul [Foundation 2022d], to write smart contracts that compile to the low-level EVM bytecode. In Figure 3, we show an excerpt of the ERC20 token smart contract written in Solidity. The listed function transfer allows an address, from, to transfer ERC20 tokens to another address, to.

3.3 Decentralized Finance (DeFi)

Decentralized Finance (DeFi) is a peer-to-peer financial system built on top of blockchains [Wüst and Gervais 2018]. The building blocks of DeFi are smart contracts that constitute the DeFi protocols [Cao et al. 2021; Popescu et al. 2020]. A few DeFi protocols dominate the DeFi market and serve as references and callees for other decentralized applications: stable coins, price oracles, decentralized exchanges (DEXes), aggregators, and lending and borrowing providers.

3.4 Transactions Execution

A transaction is constituted of the sender address, the recipient address, the transferred value of native token (can be zero), transaction data (can be empty), and a gas value to pay the transaction fees. If a transaction recipient address is associated with user account then the transaction is simple payment transaction that transfers the value of native token from the sender account to the recipient account and the transaction's data is empty. Otherwise, if the transaction recipient address is associated with a smart contract account then the transaction's data identifies a function of the recipient smart contract's code together with arguments passed to execute the function. When the transaction is received by miners, the corresponding smart contract's function is executed by the EVM, modifying the smart contract's balance and storage trie accordingly. Note that the execution of each EVM command, such as read/write operations on the underlying storage trie, is associated with a gas fee. The transaction's gas value must exceed the accumulated gas fees at the end of execution, otherwise the transaction is aborted.

4 PRELIMINARIES AND PROBLEM STATEMENT

We use labeled transition systems to model the behaviors of DeFi smart contracts. A *labeled transition system* (LTS) $A = (Q, \Sigma, q_0, \delta)$ over the possibly-infinite alphabet Σ is a possibly-infinite set Q of states with initial state $q_0 \in Q$, and a transition relation $\delta \subseteq Q \times \Sigma \times Q$. The i th symbol of a sequence $\tau \in \Sigma^*$ is denoted τ_i , and ϵ is the empty sequence. An *execution* of A is an alternating sequence of states and transition labels (also called *actions*) $\rho = q_0, a_0, q_1 \dots a_{k-1}, q_k$ for some $k > 0$ such that $\delta(q_i, a_i, q_{i+1})$ for each $0 \leq i < k$. We write $q_i \xrightarrow{a_i \dots a_{j-1}}_A q_j$ as shorthand for the subsequence $q_i, a_i, \dots, q_{j-1}, a_{j-1}, q_j$ of ρ . (in particular $q_i \xrightarrow{\epsilon} q_i$). The projection $\tau|_\Gamma$ of a sequence τ is the maximum subsequence of τ over the alphabet Γ . A *trace* of A is the projection $\rho|\Sigma$ of an execution ρ of A . The set of traces of an LTS A is denoted by $T(A)$.

In this work, a smart contract is interpreted as an LTS whose traces represent sequences of invocations to the contract's methods together with their inputs parameters and outcomes.

Formally, an *invocation label* $\text{adr}.m(\vec{u})$ is a method name m along with a vector \vec{u} of argument values and the address $\text{adr} \in \mathbf{A}$ of the contract m belongs to. An *operation label* $\ell = \text{adr}.m(\vec{u}) \Rightarrow I, v \sqcup \perp$ is an invocation label $\text{adr}.m(\vec{u})$ along with a return value v , and I is a sequence of operation labels corresponding to the “internal” calls made during the invocation of m (e.g., a call to send with its arguments and return value). We assume the preconditions are satisfied for all the operations in I , otherwise, the external invocation $\text{adr}.m(\vec{u})$ reverts. We assume a fixed, but unspecified, domain Vals of argument or return values. The distinguished invocation outcome \perp is associated to invocations that revert. We assume w.l.g. that the preconditions for all operation labels in I to not revert are satisfied, otherwise, the outcome of $\text{adr}.m(\vec{u})$ is \perp . We use $\text{inv}(\ell)$ to refer to the invocation label in an operation label ℓ . This notation is extended to sequences or sets of operation labels as expected. An *interface* Σ_{adr} is a set of operation labels over a finite set of method names. We use $\Sigma_{\text{adr}}^\checkmark$ to denote the subset of Σ_{adr} that excludes operation labels with \perp as a return value, and $\text{Meths}(\Sigma_{\text{adr}})$ to denote the method names in Σ_{adr} .

Definition 4.1. A *smart contract* is an LTS $C_{\text{adr}} = (Q_{\text{adr}}, \Sigma_{\text{adr}}, q_0, \delta_{\text{adr}})$ over an interface Σ_{adr} .

For uniformity, we use the LTS $C_{\text{adr}} = (Q_{\text{adr}}, \Sigma_{\text{adr}} = \{\text{adr.send}(\text{adr}', u) \Rightarrow \{\text{adr'.send}(u) \Rightarrow \text{True}\}, \text{True} \sqcup \perp; \text{adr.send}(u) \Rightarrow \text{True}; \text{adr.balance} \Rightarrow v\}, q_0, \delta_{\text{adr}})$ to model accounts that are not associated with smart contracts. The above LTS only contains the native token transfer operation labels and a read-only method to fetch the account balance. Note that invocations to read-only methods do not change the state. The invocation label $\text{adr.send}(\text{adr}', u)$ transfer an amount u of native tokens from the account adr to the account adr' . The outcomes is either revert or successful transfer consist of adr local state modification, removing the u amount from its balance, and the operation $\text{adr'.send}(u) \Rightarrow \text{True}$ modifies the local state of adr' , adding the u amount to its balance.

We use $Q = \bigsqcup_{\text{adr} \in A} Q_{\text{adr}}$, $\Sigma = \bigsqcup_{\text{adr} \in A} \Sigma_{\text{adr}}$, $\delta = \bigsqcup_{\text{adr} \in A} \delta_{\text{adr}}$.

Definition 4.2. We define the LTS $B = (Q, \Sigma, q_b, \delta)$ to represent the evolution of the whole blockchain state where $q_b \in Q$ is the initial state of the blockchain.

Given an execution of B is an alternating sequence of states and transition labels $\rho = q_1, \ell_{\text{adr}1}, q_1 \dots \ell_{\text{adr}n}, q_n$ for some $n > 1$ such that $\delta(q_i, \ell_{\text{adr}i}, q_{i+1})$ for each $1 \leq i < n$, we define the notion of symbolic actions vector $S = \ell_{\text{adr}1} \dots \ell_{\text{adr}n}$ such that $\ell_{\text{adr}i} \in \Sigma$ for each $1 \leq i < n$ as the sequence of operation labels possibly from different smart contracts.

Definition 4.3. A *symbolic actions vector* is a sequence of operation labels $S = \ell_{\text{adr}1} \dots \ell_{\text{adr}n}$ such that $\ell_{\text{adr}i} \in \Sigma$ for each $1 \leq i < n$ and $\rho = q_1, \ell_{\text{adr}1}, q_1 \dots \ell_{\text{adr}n}, q_n$ where $\delta(q_i, \ell_{\text{adr}i}, q_{i+1})$ for each $1 \leq i < n$ is an execution of B .

We use T to denote the domain of tokens identifiers (including native token). We define the mapping $M : Q \times A \times T \Rightarrow V$ that maps the tuple $(q, \text{adr}, t) \in Q \times A \times T$ consisting of a blockchain state, an address, a token identifier, to the amount of token t holds by the address adr at the blockchain state q . We define the mapping $P : T \Rightarrow V$ that maps each token t to its price.

Definition 4.4. We define the mapping $B : Q \times A \Rightarrow V$ that maps the pair $(q, \text{adr}) \in Q \times A$ to the weighted sum of tokens the address adr holds at the blockchain state q , i.e., $B(q, \text{adr}) = \sum_{t \in T} M(q, \text{adr}, t) \cdot P(t)$.

Next we define an attack vector by an address adr as a symbolic actions vector S where the symbolic arguments of a method invocation are replaced with a concrete values (integer values) and S transforms a blockchain state q to another state q' such that $B(q', \text{adr}) - B(q, \text{adr}) > 0$, i.e., the adversary adr is able to generate profit when the sequence of actions S is executed with the concrete values.

Definition 4.5. An *attack vector* by an adversary adr consists of a symbolic actions vector S where the symbolic arguments are replaced by concrete values and S transforms a blockchain state q to another state q' such that $B(q', \text{adr}) - B(q, \text{adr}) > 0$.

Example 4.1. The following is the attack vector of the Eminence exploit:

Eminence.buy(15000000e18), eAAVE.buy(810280759e18), Eminence.sell(573231582e18),
eAAVE.sell(631409e18), Eminence.sell(800342682e18)

The corresponding symbolic actions vector:

Eminence.buy(a), eAAVE.buy(b), Eminence.sell(c), eAAVE.sell(d), Eminence.sell(e)

The above actions vector consists of five methods invocations from two contracts: Eminence and eAAVE.

4.1 Problem Statement

Our goal in this paper is to synthesize a concrete attack vector C from a given specification φ . We assume that φ contains possible actions that will form the attack vector, e.g., methods that constitute a DeFi protocol.

Definition 4.6 (Problem Statement). Given a specification φ , an adversary adr , and a blockchain state q , our objective is to find an attack vector consisting of a concretization of the symbolic actions vector $S = \ell_{\text{adr}1} \dots \ell_{\text{adr}n}$ such that $\ell_{\text{adr}i} \in \Sigma \cap \varphi$ for each $1 \leq i < n$, transforming the state q to a state q' , and that maximizes the profit $\mathcal{B}(q', \text{adr}) - \mathcal{B}(q, \text{adr})$.

5 OUR APPROACH

The most naive approach to the above synthesis problem is, to enumerate all possible symbolic attack vectors, fill in the symbolic values with all possible parameters, and execute all these concrete attack vectors in a forked simulation environment to check if they actually yield positive profit.

However, this solution is infeasible in many aspects. (1) Large range of parameters. In Solidity, the range of an integer is 0 to $2^{256} - 1$, which indicates it is impossible to try all possible parameters. (2) Slow execution. It is notoriously slow to fork a block and execute our transactions in a locally simulated environment. [Kim et al. 2021]

Another approach is to manually extract closed-form expressions of state transition functions, construct the optimization framework for an attack vector and solve the optimization problem mathematically, as shown in prior work [Cao et al. 2021; Qin et al. 2021]. However, this approach still require (1) an expert knowledge about DeFi protocols and smart contracts (2) is not applicable to DeFi attacks involving smart contracts that are not open-sourced, such as the case for the Harvest USDC attack described in Section 2 (3) is not applicable which it is not possible to extract the closed-form expression of a DeFi protocol endpoint. For example, the decentralized exchange Curve [Egorov 2019] uses an iterative method to calculate stablecoins' prices while maintaining its StableSwap invariant. It is impossible to extract a closed form expression of stablecoins' prices for a pool with ≥ 3 tokens as it requires solving a three-variable cubic equation.

To address the above concerns, we present a framework for attack vectors synthesis which first searches all symbolic actions vector that can lead to feasible attacks, approximates state transition functions, then constructs the optimization framework to automatically find for all possible attack vectors.

5.1 Symbolic Attack Vector Synthesis

We now explain the synthesis procedure in Algorithm 1. The procedure first collects initial datapoints to use approximate the actions in A (line 2) where we use a starting blockchain state the state q . Then, using the sub-procedure *approximate* we generate the approximation of the actions in A using the collected datapoints (line 6). We use the sub-procedure *actionsVectors* to generate all possible actions vectors of length less than len (line 7). We iterate over the generated actions vectors where use some refinement heuristics implemented in the sub-procedure *isFeasible* to prune actions vectors (line 9), for instance an actions vector containing two adjacent actions invoking the same method that swaps token X to token Y can be pruned to an actions vector where the two adjacent actions are merged. For a pruned actions vector, we use the optimization sub-procedure *optimize* (line 10) to find the suitable concrete values to pass as input parameters to the methods in the actions vector that satisfy the constraints encoded in the objective function \mathcal{P} . In the optimization procedure we use for each method its approximated version. We then validate whether the attack vector generated by the optimizer does indeed generate the profit. To do this we use the sub-procedure *queryOracle* to execute the actions vector with the optimizer generated

Algorithm 1: Attack vectors synthesis procedure.**Procedure:** Synthesize($A, len, \mathcal{P}, q, it$)**Input** : a set of actions A , a maximum length of an actor vector len , a target profit function \mathcal{P} , a blockchain state q , and a maximum number of iterations for the counterexample-guided for loop it .**Output** : sequences of actions with corresponding parameters which yield the best target profit

```

1 for  $a$  in  $A$  do
2   | datapoints[ $a$ ] := dataCollect( $q, a$ )
3 end
4  $it := 0$ 
5 for  $it < It$  do
6   |  $A' := \text{approximate}(A, \text{datapoints})$ 
7   |  $wlist := \text{actionsVectors}(A', len)$ 
8   | for  $p$  in  $wlist$  do
9     | if  $\text{IsFeasible}(p)$  then
10      | ( $p^*, \text{profit}$ ) := optimize( $p, \mathcal{P}$ )
11      | if  $\text{queryOracle}(q, p^*, \text{profit})$  then
12        |  $\text{answerlist.add}((p^*, \text{profit}))$ 
13      | else
14        |  $\text{datapoints} := \text{datapoints} \cup \text{CGDC}(p^*, q)$ 
15      | end
16    | else
17      | continue
18    | end
19  | end
20  |  $it := it + 1$ 
21 end
22 return  $\text{answerlist}$ 

```

input parameters on the actual smart contracts on the blockchain. If the query is successful, i.e., the actual profit closely matches the profit found by the optimizer, then we add the attack vector to the list of discovered attacks. Otherwise, if the query is not successful, we consider the attack vector to be counterexample, and we use it to generate new datapoints to refine the approximation of actions in subsequent iterations, the sub-procedure (line 14). We repeat the above process until the number of iterations reaches a fixed threshold It (line 5).

5.2 Pruning Attacks Vectors

We use the sub-procedure *isFeasible* to implement several heuristics to identify and prune undesired actions vectors. We discuss here some heuristics, and we elaborate more on others in the implementation section.

Heuristic 1: no duplicate adjacent actions. Using the observation that calling the same method in a DeFi smart contracts twice successfully is usually equivalent to calling the method once but with different parameters, we prune out actions vectors that contain two adjacent duplicates calls of the same method.

Heuristic 2: limited usage of a single action. Using the observation that for attack vectors that do not contain repetitions (e.g., for loop) a single action is only called limited number times. DeFi attacks usually involve calling actions from different DeFi platforms to manipulate the prices and

exchange rates, we fix a maximum threshold of how many instances of an action an actions vector can contain (this threshold is proportional to the length of the actions vector) and prune out actions vectors that do not meet this criterion.

5.3 Transition Functions Approximation

We now describe the *approximate* for approximating the actions that can be part of an attack vector. We note that smart contracts are deterministic since nodes in the blockchain relies on deterministic outcomes for any sequence of transactions to reach consensus. Thus, the state reached by contracts when executing a sequence of actions is unique. Therefore, as the effect of an action is solely dependent on the current state q and input parameters $params$, it is possible to write every element of the post-state as a function of the pre-state S and the input parameters $params$, i.e., $q' = f(q, params)$.

We use the collected datapoints which consist of a list of the tuple $(q', params, q)$ to estimate the function f^5 . To carry the estimation we use two different numerical methods to infer the approximation, i.e., polynomial approximation and multivariate interpolation.

5.3.1 Polynomial. In our polynomial approximation, we first adopt the polynomial features extraction technique to generate a new feature matrix consisting of all polynomial combinations of the features with degree less than or equal to a fixed threshold n . Then, we use linear regression to find polynomials' coefficients. Finally, we apply the obtained polynomials to all inputs and distinguish the number of misapproximated points.

5.3.2 Interpolation. Since not every multivariate function can be approximated using multivariate polynomials, we also use multivariate interpolation method as an alternative to multivariate polynomial approximation method. In particular, we adopt Nearest-neighbor interpolation method to build a nearest N dimensional interpolator based on the collected datapoints tuples. When the interpolator is called to evaluate on a new point, it searches for the nearest neighbor of the new point and return its estimated value.

5.4 Optimization

Given actions vector and the approximated methods called by those actions, in *optimize* we construct an optimization framework to find optimal values for the symbolic values in the actions vector.

Objective function: Given a blockchain state q and a address adr , the actions vector S transforms q to another state q' . The objective function in the optimization problem consists the increase of the tokens values in the balance of the address adr , i.e., $y = \mathcal{B}(q', adr) - \mathcal{B}(q, adr)$

Constraints: The optimization procedure is accompanied by constraints on the symbolic values that to be inferred. For instance, the balance of a token must always be non-negative, in other words, the adversary and the smart contracts cannot use more tokens than what they have in their balances, otherwise the transaction reverts. Also, for a symbolic parameter p_i that we want to infer a value for, we fix an upper bound value u_i .

In the following, we give an example of the shape of the optimization equation.

$$\begin{aligned} & \max_{p_0, p_1, \dots, p_n} y = \mathcal{B}(q', adr) - \mathcal{B}(q, adr) \\ \text{subject to: } & \begin{cases} 0 < p_i < u_i & i = 0, 1, \dots, n \\ \forall t \in T, adr' \in A. M(q', adr', t) \geq 0 \end{cases} \end{aligned} \quad (1)$$

⁵note that f can be a vector of functions where each function an element map the state q given the input parameters to an element of the state q' .

5.5 Counterexample Guided Data Collection (CGDC)

During synthesis, inevitably our optimization procedure explores some state space unexplored before, which challenges the accuracy of the approximations and discovers mismatch between the execution results and the estimated results. Thus, it is necessary to collect new data points based on the counterexamples that show the mismatch between the estimated and the actual values, to refine the approximations. This allows to the approximation accuracy and the erroneous counterexample is eliminated. Now we introduce a counterexample guided data collection (CGDC) technique, inspired of counterexample guided abstraction refinement [Clarke et al. 2000], to refine data points and approximations when an approximation error is identified.

We use C to denote the attack vector s.t. $q \xrightarrow{C} q'$ and we use q'_e to denote the estimated value for the state q' found by the optimizer and q'_a to the actual value obtained when executing C on the actual protocol on the blockchain. We use $\mathcal{P}_e(C) = \mathcal{B}(q'_e, \text{adr}) - \mathcal{B}(q, \text{adr})$ to denote the estimated profit and $\mathcal{P}_a(C) = \mathcal{B}(q'_a, \text{adr}) - \mathcal{B}(q, \text{adr})$ to denote the actual profit.

Definition 5.1. A counterexample is an attack vector C whose estimated profit $\mathcal{P}_e(C)$ is different from its actual profit $\mathcal{P}_a(C)$. Formally, $|\mathcal{P}_e(C) - \mathcal{P}_a(C)| \geq \varepsilon \cdot |\mathcal{P}_a(C)|$, where ε is a small constant representing accuracy tolerance.

Algorithm 2: Counterexample guided data collection.

Procedure: CGDC(C, q)

Input : A counterexample C and a blockchain state q .

Output : datapoints

```

1 datapoints := []
2 for  $k = \text{len}(C)$  to 1 do
3    $q'_e := \text{estimate}(q, C, k)$ 
4    $q'_a := \text{execute}(q, C, k)$ 
5   if isAccurate( $q'_e, q'_a$ ) then
6     returns datapoints
7   else
8     ( $a, \text{paras}$ ) :=  $C[k]$ 
9     datapoints[ $a$ ] := ( $q, \text{paras}, q'_a$ )
10  end
11 end
```

In Algorithm 2, we present the *CGDC* sub-procedure for collecting new datapoints from a counterexample. The procedure takes as inputs a counterexample C which is known to have an inaccurate profit estimation, and a blockchain state. The for loop at line 2 is used to locate approximation errors backward from the last action to the first action and collect new data points accordingly. In a loop iteration k , we check if the estimated functions of the action at the index k of C are accurate. First, we compute the estimated state q'_e reached by executing C until reaching the action indexed k (line 5) using the approximated of transition functions. Second, we compute the actual state q'_a reached by executing C until reaching the action indexed k (line 6) on the actual smart contracts on the blockchain. Then we compare the estimated and actual execution results (line 7). If the estimation is accurate, this indicates that transition functions of the action at the index k of C and its predecessors are accurate; so the procedure breaks the loop and returns the datapoints computed in the previous iterations (line 6). Otherwise, it indicates inaccurate transition functions of this action or/and its predecessors. Thus, we add a new data point associated with the

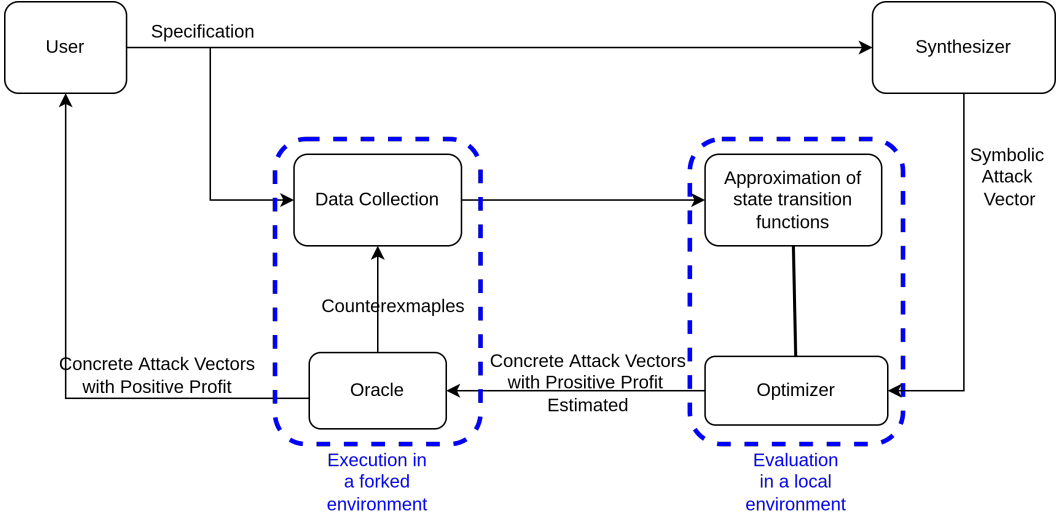


Fig. 4. The overview of our implementation

action at the index k of C (lines 8 and 9) and proceed to its next iteration of the loop to explore its predecessors.

6 IMPLEMENTATION

In this section, we provide implementation details about different components of our tool.

6.1 Overview

Our tool's workflow consists of four basic steps as shown in Figure 4. In Step 1, it uses the specifications provided by users to collect a wide range of data points, and uses these data points to approximate state transition functions for each action. In Step 2, the synthesizer does a heuristic guided top-down search on sequences of actions and find all feasible symbolic attack vectors. In Step 3, the optimizer makes use of approximated state transition functions to find concrete attack vectors for a symbolic attack vector which is estimated to have a positive profit. In Step 4, it queries the oracle to check if the concrete attack vectors in step 3 can really yield positive profit. If not, those counterexamples are used to collect a new round of data points and make corrections to the approximated state transition functions.

6.2 State transition function approximation

As introduced in Section 5.3, given a set of data points $((S, p), (S', Q_{in}, Q_{out}))$, we are interested in finding approximating function f_1, f_2, f_3 , such that $S' \approx f_1(S, p)$, $Q_{in} \approx f_2(S, p)$, and $Q_{out} \approx f_3(S, p)$ ⁶. We consider finding f as a multivariate numerical approximation problem: given a set of input-output pairs $(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, find the best approximating function f , such that there are fewest misapproximated points, formally, minimize $\sum_{i=0}^n \llbracket |f(\mathbf{x}_i) - y_i| > \varepsilon \cdot |y_i| \rrbracket$, where ε is a constant representing tolerance.

6.2.1 Polynomial. We first adopt **PolynomialFeatures** method from **sklearn.preprocessing** library to generate a new feature matrix consisting of all polynomial combinations of the features

⁶Note f_i represent one or more functions, since there might be several poststates, token(s) consumed, or token(s) got.

with degree less than or equal to n . Then we use **LinearRegression** from **sklearn.linear_model** library to find polynomials' coefficients. Finally we apply the polynomial to all inputs and find the number of misapproximated points.

In practice, for any state transition function f , we tried $n = 1, 2, \dots, 6$, and choose the polynomial which gives us fewest misapproximated points as our desired approximating function.

However, not every multivariate function can be approximated using multivariate polynomials. Thus, multivariate interpolation method is also adopted as an alternative to multivariate polynomial approximation method.

6.2.2 Interpolation. We adopt Nearest-neighbor interpolation method from **scipy.interpolate** library to build a Nearest N dimensional interpolator based on given input-output pairs. When the interpolator is called to evaluate on a new point, it searches for the nearest neighbor of the new point and return its estimated value.

6.3 Optimizer Selection

scipy.optimize library provides many off-the-shelf local and global optimizers. However, local minimizers only explore a small neighborhood within the given ranges. It is not convincing in our settings as we aim to find the **existence** of a solution which makes objective function positive. Besides, local optimizers often require users to provide an initial guess of parameters, which will be used as a starting point for the algorithm to iteratively converge to the optimal solution. However, in our settings, it is difficult to find an initial guess for every symbolic attack, as each symbolic attack behave completely differently.

Among all global optimizers, we adopt simplicial homology global optimization(**shgo**) [Endres et al. 2018; Joe and Kuo 2008; Wales 2015] to find the optimal parameters. It allows users to specify (black-box) objective functions and constraints, which suits our settings as we do not have an explicit expression of interpolation functions.

6.4 Oracle

The oracle refers to a module able to execute transactions in a simulated environment and return execution results. We build our oracle on top of Foundry [Foundry 2022]. Foundry is a toolkit written in Rust for smart contract development. It supports parallel compilation of smart contracts, and cached data for a forked block. FlashSyn communicates with Foundry via command-lines. In our experiment, if the benchmark is an adversarial DeFi attack in the history of Ethereum Blockchain/Binance Smart Chain, we fork the blockchain at one block prior to the original adversarial transaction. If the benchmark is a case of security challenges(Damn Vulnerable DeFi⁷), we deploy corresponding DeFi protocols locally in advance.

6.5 Transaction Execution Batching

When building the oracle, we realize executing multiple transactions together is faster than executing them one by one. Thus, we batch transaction executions in our algorithm to reduce the number of queries to the oracle. Specifically, we batch line 12 - 16 in Algorithm 1, and run Synthesize Procedure iteratively. For each iteration, we find all feasible symbolic attack vectors and store corresponding lists of answer candidates(line 11). After the While loop, FlashSyn queries the oracle to execute all answer candidates together, find counterexamples, and collect new data points. Then we continue a new iteration of Synthesize Procedure with augmented approximations and a better time budget for the optimizer(line 11). The iterative process stops when there is no profit increase between the last two iterations.

⁷<https://www.damnulnerabledefi.xyz/>

7 EVALUATION

In this section, we describe a series of experiments designed to evaluate our tool's performance.

Benchmarks: In the following, we evaluate our tool on the fourteen existing DeFi attacks: six from Ethereum Blockchain, five from Binance Smart Chain, one from Fantom and two from Damn Vulnerable DeFi. Damn Vulnerable DeFi is CTF-like security playground for decentralized finances. It has twelve challenges which implicitly contain vulnerabilities. Two of the challenges are collected and slightly modified as benchmarks to evaluate our tool. We collect those benchmarks based on the applicability of our tool. These requirements are as follows: (1) The attack vector only involves calling existing DeFi protocol endpoints, without creating new contracts except the exploit itself. (2) The attack sequence can be executed atomically within a single transaction. The above requirements make Flashloan attacks, Arbitrage Attacks, Oracle Manipulation Attacks the best candidates of our benchmarks.

Experimental Setup: All of our experiments are conducted on a machine with Intel(R) Core(TM) i7-9850H 2.60GHz CPU and 32 GB of physical memory, running the Ubuntu 18.04 operating system with an NVIDIA GeForce GTX1650 GPU.

Initial Balance: When constructing optimization framework, prior works [Cao et al. 2021; Qin et al. 2021] optimized the parameters of getting initial balances needed to execute the attack vector. In this paper, to generalize our approach and simplify the problem, we assume the exploiter always have the enough initial balance at the beginning of the attack.

For the cases from Ethereum Blockchain and Binance Smart Chain, the attackers in history all borrowed a large flash loan at the start of the transaction and returned them in the end. To simulate a similar scenario, we issued tokens at the beginning of the attack as the initial balances. These initial balances are the same as or slightly larger than flash loan the attackers borrowed in history.

We assume zero cost to get these initial balances, because (1) some flashloan providers charge little. For example, dydx⁸ has a flashloan fee of 2 Wei no matter how much is flash loaned. (2) These attacks can also be conducted by big whales who have these balances already.

For the two cases from Damn Vulnerable DeFi, we assume the exploiter has the same initial balance as described in the problem statement.

Token Prices:

To compare profit of different concrete attack vectors, we assign each token a price. For the cases from Ethereum Blockchain and Binance Smart Chain, we choose the price of the most popular token involved in each attack to be 1.0 and assign other tokens a price based on popular DEX liquidities of each token. For the two cases from Damn Vulnerable DeFi, since the token "DVT" is made up, we assume 1000 DVT = 1 ETH to guarantee the values of DVT and ETH supplies are roughly the same in Uniswap.

7.1 Ethereum Blockchain

7.1.1 bZx1[Transaction] [Post Mortem]. On Feb. 15th, 2020, a flash loan transaction hit bZx protocol, triggering a logic bug inside bZx protocol and yields a profit of 1193.69 ETH (350k USD), which has been well studied by other researchers [Cao et al. 2021; Qin et al. 2021]. If the steps needed to get initial balances and return the flashloan are ignored, we identify two actions which fully reveal the vulnerabilities and form the attack sequence.

⁸<https://dydx.exchange/>

Table 3. Profit summary.

Blockchain	Application	Actual profit	# Actions	Polynomial		Interpolation	
				Profit	Time	Profit	Time
Ethereum	bZx1	2209	2	2422	21	2365	49
	Harvest_USDC	338544	4	125843	332	738	2717
	Harvest_USDT	307416	4	24020	265	375	3404
	Warp	1705471.85	6	1541093	649	-	-
	Eminence	1674278	5	870709	481	-	-
	CheeseBank	3335773	8	1762771	4180	111011	120
Binance	ElevenFi	129743	5	103581	6	137699	35
	bEarnFi	13807(1 loop)	2	12495	28	11543	43
	ApeRocket	1345	6	352	1153	1100	1091
	Novo	24857	4	16401	80	19146	361
	WDoge	78	5	75	51	75	27
Fantom	OneRing	1534752	2	1408293	64	1838593	81
Damn Vulnerable	Puppet	93000	2	107215	64	93282	20
DeFi	PuppetV2	1000000	3	854058	245	672400	545

Table 4. Key Actions of bZx1 Attack

Protocol	Function	Token In	Token Out
bZx	mintWithEther(value: -)(address(this), 0)	ETH	/
uniswapV1	tokenToEthSwapInput(-, 1, 0xffffffff)	WBTC	ETH

Attack Root Cause: The core of bZx1 attack is a logic bug inside the `mintWithEther`. When `mintWithEther` is called with n ETH, it will automatically short sell $m \approx 4.33n$ ETH against WBTC through Uniswap V1. In other words, the exploiter spent fewer ETH but opened an unreasonably larger short position, which offered arbitrage opportunities. This was caused by the bug which skipped the sanity check, reported by peckshield⁹.

Real Attack Vector in History:

Action 1: `mintWithEther(value: 1300 ether)(address(this), 0)`

Action 2: `tokenToETHwapInput(112 * 1e8, 1, 0xffffffff)`

Adjusted profit: 1194

First the exploiter called `mintWithEther` with 1300 ETH, which triggers bZx protocol to swap 5637.62 ETH on Uniswap V1 for only 51.35 WBTC (at 109.79 ETH/WBTC). Then the exploiter converted its 112 WBTC into 6871.41 ETH (at 61.35 ETH/WBTC), a price substantially lower than market price. Overall, the adjusted profit is $(6871.41 - 1300) \cdot 1.0 - 112 \cdot 39.08 \approx 1194$

Best Attack Vector from FlashSyn:

Action 1: `mintWithEther(value: 1299 ether)(address(this), 0)`

Action 2: `tokenToETHwapInput(48 * 1e8, 1, 0xffffffff)`

Adjusted profit: 2319

Datapoints Statistics:

⁹<https://peckshield.medium.com/bzx-hack-full-disclosure-with-detailed-profit-analysis-e6b1fa9b18fc>

Table 5. bZx1 State Reading Functions

Protocol	State Reading Function	Note
uniswapV1	WBTC.balanceOf(UniswapWBTCAddress)	Uniswap's WBTC liquidity
uniswapV1	UniswapWBTCAddress.balance	Uniswap's ETH liquidity

Table 6. Number of Data Points for bZx1 Actions

Function	Start	End(Polynomial)	End(Interpolation)
mintWithEther	2045	2075	2167
tokenToEthSwapInput	561	584	627

7.1.2 *Harvest_USDC[Transaction][Post Mortem]*. Details are shown in Section 2

7.1.3 *Harvest_USDT[Transaction][Post Mortem]*. On October 26th, 2020, right after Harvest_USDC attack, the same exploiter performs a similar attack vector on USDT vault of Harvest.Fi. It repeatedly performs the same attack vector in 13 transactions. Harvest_USDT attack and Harvest_USDC attack caused a financial loss of about 34 million USD.

Table 7. Key Actions of Harvest_USDT Attack

Protocol	Function	Token In	Token Out
Curve.Fi	exchange_underlying(1, 2, -, 0)	USDC	USDT
Curve.Fi	exchange_underlying(2, 1, -, 0)	USDT	USDC
Harvest.Fi	deposit(-)	USDC	fUSDC
Harvest.Fi	withdraw(-)	fUSDC	USDC

Attack Root Cause: When a user deposits into or withdraw from Harvest vaults, the investment strategies of Harvest calculate the real-time value of assets invested in the underlying real-time protocols. The vaults use the calculation results to decide the shares to be issued to the user or the payout a user should receive. However, the assets of the vaults are deposited in shared public pools(such as Y Pool on Curve.Fi), which are subject to market effects like slippage. Thus, the real-time value of assets invested can be manipulated via market trades with a large volume, leaving a large arbitrage space.

Real Attack Vector in History:

Action 1: `exchange_underlying(2, 1, 17222012e6, 0)`

Action 2: `deposit(49977468e6)`

Action 3: `exchange_underlying(1, 2, 17239234e6, 0)`

Action 4: `withdraw(51456280e6)`

The exploiter first swaps a large amount of USDT to USDC via Y Pool on Curve.Fi, which pumps up the price of USDC(Action 1). As a consequence, Harvest's investment strategy calculates real-time value of its assets invested, which is abnormally high. This leads to more shares being issued to the exploiter when it deposits into the vault(Action 2). Then the exploiter swaps a large amount of USDC to USDT via Y Pool, which makes USDC/USDT price back to normal(Action 3). As a result, the real-time value of Harvest assets invested comes back to normal. This also rebalance Y pool and eliminate the exploiter's impermanent loss in Action 1.

Best Attack Vector from FlashSyn:Action 1: `exchange_underlying(2, 1, 12246094e6, 0)`Action 2: `deposit(22705078e6)`Action 3: `exchange_underlying(1, 2, 18808593e6, 0)`Action 4: `withdraw(23144531e6)`

FlashSyn finds the exact same symbolic attack vector as the real attack vector. But the synthesized attack vector uses smaller parameters for Action 2 and Action 4, which reduces the profit. Besides, the synthesized result swaps much more USDC in Action 3 than the USDT in Action 1, which keeps the imbalance of Y Pool and caused the impermanent loss to the exploiter. The adjusted profit of the above attack vector is 125843 USD.

Datapoints Statistics:

Table 8. Harvest_USDT State Reading Functions

Protocol	State Reading Function	Note
Curve.Fi	CURVE_yPool.balances(1)	Curve Y Pool USDC balance
Curve.Fi	CURVE_yPool.balances(2)	Curve Y Pool USDT balance
Harvest.Fi	fUSDC.underlyingBalanceInVault()	fUSDC underlying balance
Harvest.Fi	Strategy.investedUnderlyingBalance()	invested underlying balance
Harvest.Fi	fUSDC.totalSupply()	fUSDC total supply

Table 9. Number of Data Points for Harvest_USDT Actions

Action	Start	End(Polynomial)	End(Interpolation)
exchange_underlying(1, 2, -, 0)	7852	8349	8305
exchange_underlying(2, 1, -, 0)	8369	8874	8819
deposit(-)	7453	7710	7509
withdraw(-)	4496	5181	4683

7.1.4 *Warp[Transaction] [Post Mortem]*. On Dec. 18, 2020, Warp Finance suffered a flash loan attack which leads to about \$7.8 million loss, according to [Rekt News](#).

Table 10. Key Actions of Warp Attack

Protocol	Function	Token In	Token Out
uniswap	mint	WETH, DAI	LP
uniswap	swap	WETH	DAI
warp	provideCollateral(-)	LP	BL
warp	borrowSC(address(USDC), -)	BL	USDC
warp	borrowSC(address(DAI), -)	BL	DAI
uniswap	swap	DAI	WETH

Attack Root Cause: The core of Warp attack is a design flaw of calculating the price of LP tokens. The price of LP token is calculated as (amount(WETH) in the pool * WETH price + amount(DAI) in the pool * DAI price) / total supply of LP. Even though the developer uses Uniswap official price oracle to (correctly) calculate the prices of WETH and DAI, they failed to realize the amounts of

WETH and DAI can also be manipulated by flash loans. The exploiter took advantage of this point, pumped up the LP price, and falsely borrow excessive USDC and DAI from the pool.

Real Attack Vector in History:

Action 1: `mint(2900030e18)`

Action 2: `swapWETH2DAI(341217e18)`

Action 3: `provideCollateral(94349e18)`

Action 4: `borrowSC(address(USDC), 3917983e16)`

Action 5: `borrowSC(address(DAI), 3862646e16)`

Action 6: `swapDAI2WETH(47622329e18)`

Adjusted profit: 1693523

First the exploiter flash loans WETH and DAI and `mint` liquidity to Uniswap's WETH-DAI pair. Then the attacker swaps a huge amount of WETH into DAI in Uniswap to increase the total value of the WETH-DAI pool, pumping up the unit price of LP token. The exploiter then mortgages the previously obtained LP Token through the `provideCollateral` function. As the unit price of LP Token becomes higher, the LP Token mortgaged by the attacker can lend more stable coins to make profits.

Best Attack Vector from FlashSyn:

Action 1: `swapWETH2DAI(479381e18)`

Action 2: `mint(316661e18)`

Action 3: `provideCollateral(64243e18)`

Action 4: `borrowSC(address(USDC), 3847320e16)`

Action 5: `borrowSC(address(DAI), 3573555e16)`

Action 6: `swapDAI2WETH(56946702e18)`

Adjusted profit: 2528667

It is quite surprising that the attack vector from FlashSyn gets a higher profit compared to the attacker in history. The reason behind this is the attack vector given by FlashSyn spends more on manipulating Uniswap pair and spends less on the LP Token mortgaged. In this attack, to get the same lending allowance, it is more economical to manipulate the Uniswap pair than collateralize more LP tokens.

Datapoints Statistics:

Table 11. Warp State Reading Functions

Protocol	State Reading Function	Note
uniswap	<code>getReserves()[0]</code>	Uniswap's DAI liquidity
uniswap	<code>getReserves()[1]</code>	Uniswap's WETH liquidity

Table 12. Number of Data Points for Warp Actions

Function	Start	End(Polynomial)	End(Interpolation)
<code>mint</code>	6718	6718	6720
<code>provideCollateral(-)</code>	968	968	971
<code>borrowSC(address(USDC), -)</code>	19	19	19
<code>borrowSC(address(DAI), -)</code>	17	17	18

7.1.5 *Eminence*[*Transaction*] [*Post Mortem*]. On Sept 29, 2020, an attacker found a design flaw inside the newly deployed Eminence system, which allowed a flash loan attack and cause a financial loss of about 15 million USD.

Table 13. Key Actions of Eminence Attack

Protocol	Function	Token In	Token Out
Eminence	buy(-)	DAI	EMN
eAAVE	buy(-)	EMN	eAAVE
Eminence	sell(-)	EMN	DAI
eAAVE	sell(-)	eAAVE	EMN

Attack Root Cause: The root cause of Eminence attack is a logic bug during the interactions of Eminence and eAAVE contracts. Eminence contract has DAI as reserves and EMN tokens as shares. eAAVE contract has EMN tokens as reserves and eAAVE tokens as shares. However, when you mint eAAVE you burn EMN without removing DAI, which falsely increase the price of EMN.

Real Attack Vector in History:

Action 1: [Eminence.buy\(15000000e18\)](#)

Action 2: [eAAVE.buy\(691825807e18\)](#)

Action 3: [Eminence.sell\(691825807e18\)](#)

Action 4: [eAAVE.sell\(572431e18\)](#)

Action 5: [Eminence.sell\(691825227e18\)](#)

Adjusted profit: 1674278 DAI

First the exploiter [buy](#) EMN tokens with 15M DAI. Then the exploiter [buy](#) eAAVE tokens with 691825807 EMN, which is close to half of EMN got from Action 1. At this time, EMN tokens are burned so the price of EMN is falsely high. Then the attacker [sell](#) the remaining 691825807 EMN to get excessive DAI. Finally, the attacker [sell](#) eAAVE to get EMN and [sell](#) EMN to get DAI.

Best Attack Vector from FlashSyn:

Action 1: [Eminence.buy\(17355204e18\)](#)

Action 2: [eAAVE.buy\(787036921e18\)](#)

Action 3: [Eminence.sell\(787036921e18\)](#)

Action 4: [eAAVE.sell\(590277e18\)](#)

Action 5: [Eminence.sell\(787036921e18\)](#)

Adjusted profit: 1757458 DAI

Datapoints Statistics:

Table 14. Eminence State Reading Functions

Protocol	State Reading Function	Note
Eminence	totalSupply()	Eminence total supply
Eminence	reserveBalance()	Eminence reserve
eAAVE	totalSupply()	eAAVE total supply
eAAVE	reserveBalance()	eAAVE reserve

Table 15. Number of Data Points for Eminence Actions

Function	Start	End(Polynomial)	End(Interpolation)
deposit	5918	6027	6071
getReward	11775	12036	12072
withdrawAll	10764	10865	11148

7.1.6 *Cheesebank[Transaction][Post Mortem]*. On November 6th, 2020, a flash loan attack drained about \$3.3 million of USDC/USDT/DAI from Cheesebank contracts by exploiting a design flaw of measuring asset price from an AMM-based oracle.

Table 16. Key Actions of CheeseBank Attack

Protocol	Function	Token In	Token Out
Uniswap	mint	ETH	LP
Uniswap	swap	ETH	Cheese
CheeseBank	refresh	/	/
CheeseBank	CheeseETH.mint(-)	LP	LQ
CheeseBank	CheeseUSDC.borrow(-)	LQ	USDC
CheeseBank	CheeseUSDT.borrow(-)	LQ	USDT
CheeseBank	CheeseDAI.borrow(-)	LQ	DAI
Uniswap	swap	Cheese	ETH

Attack Root Cause: The root cause of CheeseBank attack is due to a flawed approach to measure the price of collaterals based on the AMM-based oracle Uniswap. Cheesebank contract only queries the instant states of Uniswap to calculate collateral prices. Thus, with a flashloan-based manipulation of collateral price on Uniswap, the exploitation manages to make a series of malicious borrow operations, leading to \$3.3 million of USDC/USDT/DAI loss

Real Attack Vector in History:

Action 1: `mint(50e18)`

Action 2: `swapETH2Cheese(20000e18)`

Action 3: `refresh()`

Action 4: `CheeseETH.mint(2833e18)`

Action 4: `CheeseUSDC.borrow(2068252e6)`

Action 5: `CheeseUSDT.borrow(1237995e6)`

Action 6: `CheeseDAI.borrow(87586e18)`

Action 7: `swapCheese2ETH(288822e18)`

With an ETH flash loan, the exploiter first mints UNI_V2 LP tokens. Then the exploiter raises the Cheese token price by swapping 20k WETH to 288k Cheese. This is the crucial step in this incident since the Cheese Bank uses the amount of WETH in a liquidity pool to estimate the price of the corresponding LP token. Then the exploiter calls `refresh` to update the price of LP token, and collateralize LP tokens(Action 4) to get falsely large borrow allowance, which allows the bad actor to drain all the USDC, USDT, and DAI withheld by Cheese Bank by three `borrow()` calls.

Best Attack Vector from FlashSyn:

Action 1: `mint(83e18)`

Action 2: `swapETH2Cheese(11209e18)`

Action 3: `mint(16e18)`

Action 4: `refresh()`

Action 5: `CheeseETH.mint(4002e18)`

Action 6: `CheeseUSDC.borrow(1895288e6)`

Action 7: `swapCheese2ETH(245907e18)`

Data Points Statistics

Table 17. CheeseBank State Reading Functions

Protocol	State Reading Function	Note	End(Interpolation)
Uniswap	<code>getReserves()[0]</code>	Uniswap Pair's Cheese reserve	1127
Uniswap	<code>getReserves()[1]</code>	Uniswap Pair's WETH reserve	267
Uniswap	<code>totalSupply()</code>	Uniswap Pair's LP total supply	31

Table 18. Number of Data Points for CheeseBank Actions

Function	Start	End(Polynomial)	End(Interpolation)
<code>refresh</code>	1084	1255	1127
<code>mint</code>	248	497	267
<code>CheeseDAI.borrow(-)</code>	14	350	31

7.2 Binance Smart Chain

7.2.1 *ElevenFi*[[Transaction](#)][[Post Mortem](#)]. On June 22nd, 2021, a series of flash loan attacks from two attackers were perpetrated on the Eleven Finance's NeverSellVaults.

Table 19. Key Actions of ElevenFi Attack

Protocol	Function	Token In	Token Out
Eleven.Fi	<code>addLiquidity(-, 0, uint(-1))</code>	BUSD	<code>nrvFUSDT</code>
Eleven.Fi	<code>deposit(-)</code>	<code>nrvFUSDT</code>	<code>11nrvFUSDT</code>
Eleven.Fi	<code>emergencyBurn()</code>	<code>11nrvFUSDT</code>	<code>nrvFUSDT</code>
Eleven.Fi	<code>withdraw(-)</code>	<code>11nrvFUSDT</code>	<code>nrvFUSDT</code>
Eleven.Fi	<code>removeLiquidityOneToken(-, 1, 0, uint(-1))</code>	<code>nrvFUSDT</code>	BUSD

Attack Root Cause: The root cause of the attack is a logic error inside the `emergencyBurn`. When the `emergencyBurn` is called, the staked token `nrvFUSDT` is sent out but the share token `11nrvFUSDT` is not burned. As a result, after calling `emergencyBurn`, the attacker is able to call `withdraw` to withdraw the staked token `nrvFUSDT` a second time, which later can be converted into BUSD via `deposit`.

Real Attack Vector in History:

Action 1: `addLiquidity(130001e18, 0, uint(-1))`

Action 2: `deposit(130947e18)`

Action 3: `emergencyBurn()`

Action 4: `withdraw(130947e18)`

Action 5: `removeLiquidityOneToken(261894e18, 1, 0, uint(-1))`

The exploiter first provides BUSD liquidities to Eleven Finance protocol, and get issued the staked token nrvFUSDT. Next it calls `deposit` to deposit nrvFUSDT into ElevenNeverSellVault contract of Eleven Finance, and get issued the share token 11nrvFUSDT. Then it calls `emergencyBurn` to get back its nrvFUSDT fund without burning the share token 11nrvFUSDT. Finally, it calls `withdraw` to burn 11nrvFUSDT and withdraw nrvFUSDT from ElevenNeverSellVault. In this way, it gets more staked token nrvFUSDT than what it deserves. Finally, it calls `removeLiquidityOneToken` to remove the liquidity of nrvFUSDT from ElevenNeverSellVault and get BUSD.

Best Attack Vector from FlashSyn:

Action 1: `addLiquidity(140000e18, 0, uint(-1))`

Action 2: `deposit(140000e18)`

Action 3: `emergencyBurn()`

Action 4: `withdraw(140000e18)`

Action 5: `removeLiquidityOneToken(280000e18, 1, 0, uint(-1))`

FlashSyn finds the exact same symbolic attack vector as the real attack vector in history, with only a small difference in the parameters.

Data Points Statistics

Table 20. ElevenFi State Reading Functions

Protocol	State Reading Function	Note
BUSD	<code>BUSD.balanceOf(SwapAddress)</code>	Eleven.Fi's BUSD balance
Eleven.Fi	<code>nrvFUSDT.totalSupply()</code>	Total supply of nrvFUSDT

Table 21. Number of Data Points for ElevenFi Actions

Action	Start	End(Polynomial)	End(Interpolation)
<code>addLiquidity(-, 0, uint(-1))</code>	2066	2189	2073
<code>deposit(-)</code>	/	/	/
<code>emergencyBurn()</code>	/	/	/
<code>withdraw(-)</code>	/	/	/
<code>removeLiquidityOneToken(-, 1, 0, uint(-1))</code>	681	923	701

7.2.2 *bEarnFi[Transaction][Post Mortem]*. On May 16th 2021, BvaultsBank contract of bEarn.Fi protocol was exploited and about \$11M funds were drained from the pool. The exploiter got its initial balances via recursive flash loans from eight different fund pools.

Table 22. Key Actions of bEarnFi Attack

Protocol	Function	Token In	Token Out
bEarn.Fi	<code>deposit(13, -)</code>	BUSD	/
bEarn.Fi	<code>emergencyWithdraw(13)</code>	/	BUSD

Attack Root Cause: When users interact with bEarnFi's **Bvaultsbank** contract to deposit and withdraw funds, meanwhile, **Bvaultsbank** contract interacts with the associated strategy **BvaultsStrategy** contract to deposit and withdraw funds. However, there is a difference between withdraw logic of Bvaultsbank and that of BvaultsStrategy. They use the same parameter but different asset

denomination. When a user sends a withdrawal request of 100 BUSD, **Bvaultsbank** will withdraw 100 ibBUSD from **BvaultsStrategy**. Note 1 ibBUSD is an interest bearing asset and more expensive than 1 BUSD. The difference is treated as a profit and will be credited to the exploiter when it withdraws the funds next time.

Real Attack Vector in History:

Action 1: `deposit(13, 7800000e18)`

Action 2: `emergencyWithdraw(13)`

Action 3: `deposit(13, 7800000e18)`

Action 4: `emergencyWithdraw(13)`

The exploiter first deposits 7804k BUSD into **Bvaultsbank** contract, which are immediately sent to **BvaultsStrategy** contract (Action 1). Then it called `emergencyWithdraw(13)` to directly withdraw its funds, which turns to be interpreted as withdrawing 7804k ibBUSD ($\approx 7818k$ BUSD). Note the exploiter only gets back 7804k BUSD in Action 2, the leftover is stored in **BvaultsStrategy**. In the next round, the exploiter does the same thing again, but this time, with leftover from last time, **BvaultsStrategy** credits the exploiter with $\approx 7818k$ BUSD. The adjusted profit is about 14k BUSD.

Best Attack Vector from name:

Action 1: `deposit(13, 7799710e18)`

Action 2: `emergencyWithdraw(13)`

Action 3: `deposit(13, 7003934e18)`

Action 4: `emergencyWithdraw(13)`

FlashSyn's synthesized result is very close to the real attack sequence in history. It deposits a slightly smaller amount in Action 3, which reduces the profit a little bit. The adjusted profit for the above attack vector is 12495 BUSD. **Data Points Statistics:**

Table 23. State Reading Functions

Protocol	State Reading Function	Note
bEarn.Fi	BvaultsStrategy.sharesTotal()	Total number of shares of BVault
bEarn.Fi	BvaultsStrategy.wantLockedTotal()	Locked values of BVault
bEarn.Fi	Bvaultsbank.userInfo(13, address(this))	Number of the user's share of BVault
BUSD	BUSD.balanceOf(BvaultsStrategyAddress)	BvaultStrategy's BUSD balance

Table 24. Number of Data Points for bEarnFi Actions

Action	Start	End(Polynomial)	End(Interpolation)
deposit(13, -)	4988	4992	5082
emergencyWithdraw(13)	4007	4011	4101

7.2.3 ApeRocket[Transaction] [Post Mortem]. On July 14, 2021, ApeRocket Finance was hacked due to an issue in the reward minting process. The hacker attacked the AutoCake contract just 10 hours after its deployment. The same attack vector was executed twice, causing a total of 883 BNB financial loss.

Table 25. Key Actions of ApeRocket Attack

Protocol	Function	Token In	Token Out
ApeRocket	deposit	CAKE	Share, Principle
CAKE	transfer	CAKE	/
ApeRocket	harvest	/	/
ApeRocket	getReward	Share, Principle	CAKE, SPACE
ApeRocket	withdrawAll	Share, Principle	CAKE, SPACE
ApeSwap	swapExactTokensForTokens	Space	WBNB

Attack Root Cause: The core of ApeRocket attack is a design flaw of calculating the harvested reward. `harvest` will regard the current CAKE balance of ApeRocket AutoCake contract as rewards. However, the attacker can manipulate the calculated rewards by manually transfer CAKE to AutoCake contract. The manipulated rewards will be paid to the attacker in SPACE when `getReward` is executed. In this way, excessive SPACE will be minted.

Real Attack Vector in History:

Action 1: `deposit(509143 * 1e18)`
 Action 2: `CAKE.transfer(address(AutoCake), 1105857 * 1e18)`
 Action 3: `harvest()`
 Action 4: `getReward()`
 Action 5: `withdrawAll()`
 Action 6: `swapExactTokensForTokens(503997 * 1e18)`
 Adjusted profit: 1345

First the exploiter called `deposit` 509143 CAKE to AutoCake contract to get the share. Then the exploiter manually transferred 1105857 CAKE to AutoCake contract as the reward, which resulted `getReward()` to mint excessive SPACE as the rewards to the exploiter. Next the exploiter called `withdrawAll` to withdraw CAKE and some small SPACE reward. Finally, the exploiter swapped SPACE to WBNB to get the profit.

Best Attack Vector from FlashSyn:

Action 1: `deposit(246484 * 1e18)`
 Action 2: `CAKE.transfer(address(AutoCake), 391211 * 1e18)`
 Action 3: `harvest()`
 Action 2: `CAKE.transfer(address(AutoCake), 977148 * 1e18)`
 Action 5: `withdrawAll()`
 Action 6: `swapExactTokensForTokens(620312 * 1e18)`
 Adjusted profit: 1258

Datapoints Statistics:

Table 26. ApeRocket State Reading Functions

Protocol	State Reading Function	Note
CAKE	<code>CAKE.balanceOf(AutoCakeAddress)</code>	AutoCake CakeBalance
ApeRocket	<code>MasterChef.userInfo(0, address(AutoCake))</code>	AutoCake Staked
ApeRocket	<code>AutoCake.totalShares()</code>	Total Shares of AutoCake
ApeSwap	<code>ApePair.getReserves()[0]</code>	ApePair WBNB reserve
ApeSwap	<code>ApePair.getReserves()[1]</code>	ApePair Space reserve

Table 27. Number of Data Points for ApeRocket Actions

Function	Start	End(Polynomial)	End(Interpolation)
deposit	5918	6027	6071
getReward	11775	12036	12072
withdrawAll	10764	10865	11148

7.2.4 *Novo[Transaction] [Post Mortem]*. On May 29, 2022, Novo protocol on Binance Smart Chain experienced a flash loan attack that drained liquidity from the liquidity pool and caused 278 BNB \$83K in damage.

Table 28. Key Actions of Novo Attack

Protocol	Function	Token In	Token Out
Pancake	swapExactTokensForTokensSupportingFeeOnTransferTokens1	WBNB	Novo
Novo	transferFrom	/	/
Pancake	sync	/	/
Pancake	swapExactTokensForTokensSupportingFeeOnTransferTokens2	Novo	WBNB

Attack Root Cause: The core of Novo attack is an implementation mistake inside `transferFrom`. The approval checks inside `transferFrom` are commented out, which means anybody can transfer any amount of Novo from any address. The attacker removed Novo from NovoWBNB Pancake Pair thus pumped the price of Novo. Then the attacker swap a small amount of Novo to almost all WBNB in NovoWBNB pair.

Real Attack Vector in History:

Action 1: `swapExactTokensForTokensSupportingFeeOnTransferTokens1(1720 * 1e16)`

Action 2: `transferFrom(address(pair), address(0), 113951614 * 1e9)`

Action 3: `sync()`

Action 4: `swapExactTokensForTokensSupportingFeeOnTransferTokens2(4749070 * 1e9)`

Adjusted profit: 24857 WBNB

First the exploiter called `swapExactTokensForTokensSupportingFeeOnTransferTokens` to swap WBNB for some Novo tokens. Then the attacker executed `transferFrom` to remove most of Novo from NovoWBNB Pancake pair and executed `sync` to update the reserves of NovoWBNB Pancake pair. Finally the exploiter called `swapExactTokensForTokensSupportingFeeOnTransferTokens` to swap Novo for almost all WBNB in NovoWBNB pair.

Best Attack Vector from FlashSyn:

Action 1: `swapExactTokensForTokensSupportingFeeOnTransferTokens1(1124 * 1e16)`

Action 2: `transferFrom(address(pair), address(0), 116514550 * 1e9)`

Action 3: `sync()`

Action 4: `swapExactTokensForTokensSupportingFeeOnTransferTokens2(2284484 * 1e9)`

Adjusted profit: 16401 WBNB

Datapoints Statistics:

Table 29. Novo State Reading Functions

Protocol	State Reading Function	Note
Pancake	<code>NovoWBNBPair.getReserves()[0]</code>	PancakeSwap Pair's Novo Reserve
Pancake	<code>NovoWBNBPair.getReserves()[1]</code>	PancakeSwap Pair's WBNB Reserve
Pancake	<code>Wdoge.balanceOf(address(NovoWBNBPair))</code>	PancakeSwap Pair's Novo Balance
Pancake	<code>WBNB.balanceOf(address(NovoWBNBPair))</code>	PancakeSwap Pair's WBNB Balance

Table 30. Number of Data Points for Novo Actions

Function	Start	End(Polynomial)	End(Interpolati
<code>swapExactTokensForTokensSupportingFeeOnTransferTokens1</code>	2859	2890	2867
<code>swapExactTokensForTokensSupportingFeeOnTransferTokens2</code>	1818	1916	1834

7.2.5 *WDoge[Transaction] [Post Mortem]*. On April 24, 2022, the Wiener DOGE project on Binance Smart Chain experienced a flash loan attack that resulted in a loss of US\$30,000.

Table 31. Key Actions of WDoge Attack

Protocol	Function	Token In	Token Out
Pancake	<code>swap</code>	WBNB	Wdoge
Pancake	<code>swap</code>	Wdoge	WBNB
Wdoge	<code>transfer(address(WdogeWBNBpair), -)</code>	Wdoge	/
Pancake	<code>skim(address(this))</code>	/	Wdoge
Pancake	<code>sync()</code>	/	/

Attack Root Cause: The root cause of the incident is due to the inconsistency between Wiener DOGE contract's charging mechanism and the Uniswap pool. Wdoge tokens are deflationary tokens. Suppose A transfers x Wdoge tokens to B, A lost $x + 0.04x$ Wdoge, while B received $0.9x$, and $0.1x$ will be redistributed among developers, fee wallet and other holders. However, if the uniswap pair's Wdoge reserve is less than Wdoge balance, the attacker can call `skim` to force the pair to send extra Wdoge to another address, and drain Wdoge reserve due to Wdoge's special burn mechanism.

Real Attack Vector in History:

Action 1: `swapWBNB2Wdoge(2900e18)`

Action 2: `transfer(address(WdogeWBNBpair), 5224718e24)`

Action 3: `skim(address(this))`

Action 4: `sync()`

Action 5: `SwapWdoge2WBNB(4466647e24)`

Adjusted profit: 78

Best Attack Vector from FlashSyn:

Action 1: `swapWBNB2Wdoge(2859e18)`

Action 2: `transfer(address(WdogeWBNBpair), 5156250e24)`

Action 3: `skim(address(this))`

Action 4: `sync()`

Action 5: `SwapWdoge2WBNB(4609375e24)`

Adjusted profit: 75

Datapoints Statistics:

Table 32. WDog State Reading Functions

Protocol	State Reading Function	Note
Pancake	WdogeWBNBpair.getReserves()[0]	PancakeSwap Pair's Wdoge Reserve
Pancake	WdogeWBNBpair.getReserves()[1]	PancakeSwap Pair's WBNB Reserve
Pancake	Wdoge.balanceOf(address(WdogeWBNBpair))	PancakeSwap Pair's Wdoge Balance
Pancake	WBNB.balanceOf(address(WdogeWBNBpair))	PancakeSwap Pair's WBNB Balance

Table 33. Number of Data Points for WDog Actions

Function	Start	End(Polynomial)	End(Interpolation)
skim	8100	8101	8226

7.3 Fantom

7.3.1 *OneRing[Transaction] [Post Mortem]*. On March 21st, 2022, OneRing Finance on Fantom blockchain was attacked by a flashloan attack. Assets worth about 2 million USD was stolen according to the official [post mortem](#).

Table 34. Key Actions of OneRing Attack

Protocol	Function	Token In	Token Out
OneRing	depositSafe(-, address(USDC))	USDC	OShare
OneRing	withdraw(-, address(USDC))	OShare	USDC

Attack Root Cause: The hack was possible because a design flaw of calculating the price of LP tokens(OShare). Right after contract depolyment, the exploiter borrowed and called `depositSafe` to deposit \$80 million USDC to increase the price of OShare. Then the attacker called `withdraw` to burn OShare and drain the USDC reserve of the contract.

Real Attack Vector in History:

Action 1: `depositSafe(80000000e6, address(USDC))`

Action 2: `withdraw(41965511e18, address(USDC))`

Adjusted profit: 1534752 USDC

Best Attack Vector from FlashSyn:

Action 1: `depositSafe(101586914e6, address(USDC))`

Action 2: `withdraw(51780430e18, address(USDC))`

Adjusted profit: 1838593 USDC

Datapoints Statistics:

Table 35. OneRing State Reading Functions

Protocol	State Reading Function	Note
OneRing	balanceWithInvested()	OneRing's total balance
OneRing	totalSupply()	OShare total supply

Table 36. Number of Data Points for OneRing Actions

Function	Start	End(Polynomial)	End(Interpolation)
depositSafe	4534	4654	4683
withdraw	1529	1652	1687

7.4 Damn Vulnerable DeFi

7.4.1 Puppet[Problem][Sample Solution]. In this challenge, there is a huge lending pool of Damn Valuable Tokens(DVT), where users are required to collateralize twice the borrow amount in ETH. The lending pool initially has 100000 DVT in liquidity. There is also a DVT uniswap v1 exchange, initially with 10 ETH and 10 DVT in liquidity. Beginning with 25 ETH and 1000 DVT, our goal is to steal tokens from the lending pool.

Table 37. Key Actions of Puppet Attack

Protocol	Function	Token In	Token Out
Uniswap	tokenToEthSwapInput(-, 1, 0xffffffff)	DVT	ETH
Puppet	borrow(-)	ETH	DVT

Attack Root Cause: Since the lending refers to an uniswap v1 exchange as its oracle, if there are much more DVT than ETH in the uniswap exchange, then we only need a small amount of ETH to borrow the entire amount of DVT from the lending pool. It is a typical oracle manipulation attack.

Attack Vector given by community:

Action 1: `tokenToEthSwapInput(1000e18, 1, 0xffffffff)`

Action 2: `borrow(100000e18)`

The exploiter first converts all its DVT(1000 DVT) to ETH, making the uniswap exchange has 1010 DVT and less than 1 ETH. Because of the manipulated liquidities inside the uniswap exchange, the exploiter is able to borrow 100000 DVT from the lending pool using a small amount of ETH. If we assume 1 ETH = 1000 DVT, then the adjusted profit is 93000 DVT.

Best Attack Vector from FlashSyn:

Action 1: `borrow(1e18, 1, 0xffffffff)`

Action 2: `tokenToEthSwapInput(784e18, 1, 0xffffffff)`

Action 3: `borrow(99999e18, 1, 0xffffffff)`

Action 4: `tokenToEthSwapInput(1e18, 1, 0xffffffff)`

Compared with the attack vector given by the community, our synthesized attack vector exchanges fewer DVT tokens to ETH(Action 2), and also borrows all DVT from the lending pool. Because of the token prices we assume, exchanging all DVT is no longer the best choice. There is a tradeoff between swapping more DVT to ETH in the uniswap exchange and spending more ETH to borrow all DVT from the lending pool. FlashSyn has chosen to swap 784 DVT to ETH. The adjusted profit is 107215 DVT.

Data Points Statistics:

Table 38. Puppet State Reading Functions

Protocol	State Reading Function	Note
Uniswap	<code>dvt.balanceOf(address(uniswapExchange))</code>	Uniswap DVT liquidity
Uniswap	<code>address(uniswapExchange).balance</code>	Uniswap ETH liquidity
Puppet	<code>dvt.balanceOf(address(puppetPool))</code>	Puppet Pool DVT balance

Table 39. Number of Data Points for Puppet Actions

Action	Start	End(Polynomial)	End(Interpolation)
tokenToEthSwapInput(-, 1, 0xffffffff)	2456	2483	2464
borrow(-)	1096	1155	1104

7.4.2 PuppetV2[Problem][Sample Solution]. Similar to Puppet challenge in the previous chapter, there is a lending pool of DVT, where users are required to collateralize three times the borrow amount in Wrapped ETH. There is an uniswap v2 pair, initially with 10 WETH and 100 DVT in liquidity. Beginning with 20 ETH and 10000 DVT, our goal is still to steal tokens from the lending pool.

Table 40. Key Actions of PuppetV2 Attack

Protocol	Function Name	Token In	Token Out
Uniswap	swapExactTokensForETH(-)	DVT	ETH
WETH	deposit(-)	ETH	WETH
Puppet	borrow(-)	WETH	DVT

Attack Root Cause: Similar to Puppet challenge, the lending refers to an uniswap v2 pair as its oracle, if liquidities of DVT and WETH are manipulated, an exploiter can spend a small amount of WETH to borrow the entire amount of DVT from the lending pool. It is a typical oracle manipulation attack.

Attack Vector given by community:

Action 1: `swapExactTokensForETH(10000e18)`

Action 2: `deposit(29e18)`

Action 3: `borrow(1000000e18)`

The exploiter first converts all its DVT(10000 DVT) to ETH, making the uniswap v2 pair has 10100 DVT and less than 1 ETH(Action 1). Then the exploiter wraps 29 ETH(Action 2) to prepare for borrowing all of DVT in the lending pool(Action 3). If we assume 1 ETH = 1 WETH = 1000 DVT, then the adjusted profit is 1000000 DVT.

Best Attack Vector from FlashSyn:

Action 1: `swapExactTokensForETH(9999e18)`

Action 2: `deposit(28e18)`

Action 3: `borrow(9908470e18)`

Our synthesized attack vector is very similar to the community's one. And they behave the same way. The adjusted profit of our synthesized attack vector is 854058 DVT.

Data Points Statistics:

Table 41. PuppetV2 State Reading Functions

Protocol	State Reading Function	Note
Uniswap	<code>uniswapV2Pair.getReserves()[0]</code>	Uniswap DVT liquidity
Uniswap	<code>uniswapV2Pair.getReserves()[1]</code>	Uniswap WETH liquidity
PuppetV2	<code>dvt.balanceOf(address(puppetV2Pool))</code>	PuppetV2 Pool DVT reserve

Table 42. Number of Data Points for PuppetV2 Actions

Action	Initial Round	End(Polynomial)	End(Interpolation)
swapExactTokensForETH(-)	667	704	730
deposit(-)	/	/	/
borrow(-)	63	147	138

8 RELATED WORK

Parametric optimization: Some researchers manually extracted formulas of state transitions of some DeFi smart contracts, defined related parameter constraints, and used an off-the-shelf optimizer to find the parameters which yield the best profit for several flash loan attacks [Cao et al. 2021; Qin et al. 2021]. However, this technique requires expert knowledge of the underlying DeFi protocols and a strong mathematical background, which gives users extra burden. Besides, this kind of techniques assumes the sequence of actions is known. Their task is only to optimize corresponding parameters. The technique is sufficient to check whether an attack vector is optimal but might not be enough to detect the possibility of such attacks before it happens.

Symbolic execution: Mythril [ConsenSys 2022], Oyente [Luu et al. 2016], ETHBMC [Frank et al. 2020], SmartCopy [Feng et al. 2019], and Manticore [Mossberg et al. 2019] rely on symbolic execution [King 1976] to explore the program states of a smart contract, looking for an execution path that violates a user-defined invariant. However, these tools do not scale well and suffer from path explosion problem as other traditional symbolic execution tools do. They are not universally applicable to DeFi smart contracts, such as Y Pool discussed in Section 2.

Static analysis: Slither [Feist et al. 2019], Securify [Tsankov et al. 2018], Zeus [Kalra et al. 2018] and SmartCheck [Tikhomirov et al. 2018] apply static analysis technique to detect specific type of vulnerabilities in smart contracts. They detect pre-defined patterns of vulnerabilities and over-approximate program states, which inevitably cause false positives and false negatives. These tools are also only able to analyze smart contracts locally in nature, and fail to consider interactions between smart contracts and real-time on-chain states.

9 CONCLUSION

Towards preventing flash loan attacks on DeFi protocols, we have proposed an automated synthesis framework based on numerical approximation. Given smart contracts of a DeFi protocol, we use data driven synthesis to approximate the functional behaviors of the smart contracts. To collect the synthesis data points, our current implementation assumes manually-specified functions to approximate and the sets of states that those functions modify. We consider the cost of this manual effort to be offset by a significant benefit: not dealing with the complex implementations of the functions in the DeFi protocols and using approximation to find adversarial attacks. Our experiments show that our proposed framework is practical, the approximation is precise, and we can synthesize adversarial attacks in a reasonable time period.

REFERENCES

- Lexi Brent, Neville Grech, Sifis Lagouvardos, Bernhard Scholz, and Yannis Smaragdakis. 2020. Ethainter: a smart contract security analyzer for composite vulnerabilities. In *Proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation*. 454–469.
- Yixin Cao, Chuanwei Zou, and Xianfeng Cheng. 2021. Flashot: a snapshot of flash loan attack on DeFi ecosystem. *arXiv preprint arXiv:2102.00626* (2021).
- BNB Smart Chain. 2020. Binance Smart Chain White Paper. <https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md>

- Sang-Min Choi, Jiho Park, Quan Nguyen, and Andre Cronje. 2018. Fantom: A scalable framework for asynchronous distributed systems. *arXiv preprint arXiv:1810.10360* (2018).
- Edmund Clarke, Orna Grumberg, Somesh Jha, Yuan Lu, and Helmut Veith. 2000. Counterexample-guided abstraction refinement. In *International Conference on Computer Aided Verification*. Springer, 154–169.
- ConsenSys. 2022. Mythril. <https://github.com/ConsenSys/mythril> Accessed: 2022-06-06.
- CENTRE Consortium. 2022. USD Coin (USDC). <https://www.circle.com/en/usdc>
- DefiLlama. 2022. DefiLlama. <https://defillama.com/> Accessed: 2022-06-01.
- Michael Egorov. 2019. StableSwap-efficient mechanism for Stablecoin liquidity. Retrieved Feb 24 (2019), 2021.
- Stefan C Endres, Carl Sandrock, and Walter W Focke. 2018. A simplicial homology algorithm for Lipschitz optimisation. *Journal of Global Optimization* 72, 2 (2018), 181–217.
- Etherscan. 2022. The Ethereum Blockchain Explorer. <https://etherscan.io/>
- Josselin Feist, Gustavo Grieco, and Alex Groce. 2019. Slither: a static analysis framework for smart contracts. In *2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*. IEEE, 8–15.
- Yu Feng, Emina Torlak, and Rastislav Bodik. 2019. Precise attack synthesis for smart contracts. *arXiv preprint arXiv:1902.06067* (2019).
- Ethereum Foundation. 2022a. Serpent Programming Language. <https://eth.wiki/archive/serpent> Accessed: 2022-05-24.
- Ethereum Foundation. 2022b. Solidity Programming Language. <https://docs.soliditylang.org/en/v0.8.14/> Accessed: 2022-05-24.
- Ethereum Foundation. 2022c. Vyper Programming Language. <https://vyper.readthedocs.io/en/stable/> Accessed: 2022-05-24.
- Ethereum Foundation. 2022d. Yul Programming Language. <https://docs.soliditylang.org/en/latest/yul.html> Accessed: 2022-05-24.
- The Near Foundation. 2021. The NEAR White Paper. <https://near.org/papers/the-official-near-white-paper/>
- Foundry. 2022. Foundry. <https://github.com/foundry-rs/foundry/> Accessed: 2022-05-31.
- Joel Frank, Cornelius Aschermann, and Thorsten Holz. 2020. {ETHBMC}: A Bounded Model Checker for Smart Contracts. In *29th USENIX Security Symposium (USENIX Security 20)*. 2757–2774.
- Gustavo Grieco, Will Song, Artur Cygan, Josselin Feist, and Alex Groce. 2020. Echidna: effective, usable, and fast fuzzing for smart contracts. In *Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis*. 557–560.
- Stephen Joe and Frances Y Kuo. 2008. Constructing Sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing* 30, 5 (2008), 2635–2654.
- Sukrit Kalra, Seep Goel, Mohan Dhawan, and Subodh Sharma. 2018. Zeus: analyzing safety of smart contracts.. In *Ndss*. 1–12.
- Yeonsoo Kim, Seongho Jeong, Kamil Jezek, Bernd Burgstaller, and Bernhard Scholz. 2021. An {Off-The-Chain} Execution Environment for Scalable Testing and Profiling of Smart Contracts. In *2021 USENIX Annual Technical Conference (USENIX ATC 21)*. 565–579.
- James C King. 1976. Symbolic execution and program testing. *Commun. ACM* 19, 7 (1976), 385–394.
- Chenxing Li, Peilun Li, Dong Zhou, Zhe Yang, Ming Wu, Guang Yang, Wei Xu, Fan Long, and Andrew Chi-Chih Yao. 2020. *A Decentralized Blockchain with High Throughput and Fast Confirmation*. USENIX Association, USA.
- Tether Limited. 2022. Tether Token. <https://tether.to/en/>
- Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor. 2016. Making smart contracts smarter. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 254–269.
- Jack McKay. 2022. DeFi-ing Cyber Attacks. (2022).
- Mark Mossberg, Felipe Manzano, Eric Hennenfent, Alex Groce, Gustavo Grieco, Josselin Feist, Trent Brunson, and Artem Dinaburg. 2019. Manticore: A user-friendly symbolic execution framework for binaries and smart contracts. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 1186–1189.
- Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review* (2008), 21260.
- Andrei-Dragoş Popescu et al. 2020. Decentralized finance (defi)–the lego of finance. *Social Sciences and Education Research Review* 7, 1 (2020), 321–349.
- Kaihua Qin, Liyi Zhou, Benjamin Livshits, and Arthur Gervais. 2021. Attacking the defi ecosystem with flash loans for fun and profit. In *International Conference on Financial Cryptography and Data Security*. Springer, 3–32.
- Kevin Sekniqi, Daniel Laine, Stephen Buttolph, and Emin Gün Sirer. 2020. Avalanche Platform. <https://www.avalabs.org/whitepapers>
- SlowMist. 2022. The 10 most common attacks. <https://hacked.slowmist.io/en/statistics/?c=all&d=all> Accessed: 2022-06-01.
- Sergei Tikhomirov, Ekaterina Voskresenskaya, Ivan Ivanitskiy, Ramil Takhaviev, Evgeny Marchenko, and Yaroslav Alexandrov. 2018. Smartcheck: Static analysis of ethereum smart contracts. In *Proceedings of the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain*. 9–16.

- Petar Tsankov, Andrei Dan, Dana Drachler-Cohen, Arthur Gervais, Florian Buenzli, and Martin Vechev. 2018. Securify: Practical security analysis of smart contracts. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 67–82.
- DJ Wales. 2015. Perspective: Insight into reaction coordinates and dynamics from the potential energy landscape. *The Journal of chemical physics* 142, 13 (2015), 130901.
- Shuai Wang, Chengyu Zhang, and Zhendong Su. 2019. Detecting nondeterministic payment bugs in Ethereum smart contracts. *Proceedings of the ACM on Programming Languages* 3, OOPSLA (2019), 1–29.
- Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151, 2014 (2014), 1–32.
- Karl Wüst and Arthur Gervais. 2018. Do you need a blockchain?. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE, 45–54.

A APPENDIX

```

1 function exchange_underlying(int128 i, int128 j, uint256 dx, uint256 min_dy) {
2     uint256[] memory rates = _stored_rates();
3     uint N_COINS = rates.length;
4     uint256[] memory precisions = PRECISION_MUL;
5     uint256 rate_i = rates[i] / precisions[i];
6     uint256 rate_j = rates[j] / precisions[j];
7     uint256 dx_ = dx * PRECISION / rate_i;
8     uint256 dy_ = _exchange(i, j, dx_, rates); // key step: call get_y and then call get_D
9     uint256 dy = dy_ * rate_j / PRECISION;
10    assert(dy >= min_dy, "Exchange resulted in fewer coins than expected");
11    bool[] memory tethered = TETHERED;
12    uint256 ok = 0;
13    if(tethered[i])
14        USDT(underlying_coins[i]).transferFrom(msg.sender, address(this), dx);
15    else
16        assert_modifiable(ERC20(underlying_coins[i]).transferFrom(msg.sender, address(this), dx));
17    ERC20(underlying_coins[i]).approve(coins[i], dx);
18    yERC20(coins[i]).deposit(dx);
19    yERC20(coins[j]).withdraw(dy_);
20    // y-tokens calculate imprecisely - use all available
21    dy = ERC20(underlying_coins[j]).balanceOf(address(this));
22    assert(dy >= min_dy, "Exchange resulted in fewer coins than expected");
23    if(tethered[j])
24        USDT(underlying_coins[j]).transfer(msg.sender, dy);
25    else
26        assert_modifiable(ERC20(underlying_coins[j]).transfer(msg.sender, dy));
27    log.TokenExchangeUnderlying(msg.sender, i, dx, j, dy);
28 }

```

Fig. 5. Source code of exchange_underlying(). Original code is in Vyper, rewritten in Solidity.


```

1 function get_y(int128 i, int128 j, uint256 x, uint256[] calldata _xp) returns (uint256) {
2     // x in the input is converted to the same price/precision
3     uint N_COINS = _xp.length;
4     assert( i != j && i >= 0 && j >= 0 && uint(i) < N_COINS && uint(j) < N_COINS );
5     uint256 D = get_D(_xp);
6     uint256 c = D;
7     uint256 S_ = 0;
8     uint256 Ann = A * N_COINS; // A is a constant member variable selected by the owner of the contract
9     uint256 _x = 0;
10    for(uint _i = 0; _i < N_COINS; _i = _i + 1) {
11        if(_i == uint(i)) {
12            _x = x;
13        } else if(_i != uint(j)) {
14            _x = _xp[_i];
15        } else
16            continue;
17        S_ += _x;
18        c = c * D / (_x * N_COINS);
19    }
20    c = c * D / (Ann * N_COINS);
21    uint256 b = S_ + D / Ann; // - D
22    uint256 y_prev = 0;
23    uint256 y = D;
24    for(uint _i = 0; _i < 255; _i = _i + 1) {
25        y_prev = y;
26        y = (y*y + c) / (2 * y + b - D);
27        // Equality with the precision of 1
28        if(y > y_prev) {
29            if (y - y_prev <= 1)
30                break;
31        } else {
32            if(y_prev - y <= 1)
33                break;
34        }
35    }
36    return y;
37 }

```

Fig. 6. Source code of `get_y()`. Original code is in Vyper, rewritten in Solidity.

```

1 function get_D(uint256[] calldata xp) returns (uint256) {
2   uint N_COINS = xp.length;
3   uint256 S = 0;
4   for(uint _i = 0; _i < xp.length; _i = _i + 1) {
5     uint _x = xp[_i];
6     S += _x;
7   }
8   if(S == 0)
9     return 0;
10  uint256 Dprev = 0;
11  uint256 D = S;
12  uint256 Ann = A * N_COINS; // A: a constant selected by the owner of the contract
13  for(uint _i = 0; _i < 255; _i = _i + 1){
14    uint256 D_P = D;
15    for(uint _j = 0; _j < xp.length; _j = _j + 1){
16      uint _x = xp[_j];
17      D_P = D_P * D / (_x * N_COINS + 1); // +1 to prevent /0
18    }
19    Dprev = D;
20    D = (Ann * S + D_P * N_COINS) * D / ((Ann - 1) * D + (N_COINS + 1) * D_P);
21    // Equality with the precision of 1
22    if(D > Dprev) {
23      if(D - Dprev <= 1)
24        break;
25    } else {
26      if(Dprev - D <= 1)
27        break;
28    }
29  }
30  return D;
31 }

```

Fig. 7. Source code of get_D(). Original code is in Vyper, rewritten in Solidity.