FlashSyn: Flash Loan Attack Synthesis via Counter Example Driven Approximation

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

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

In this paper, we introduce a new framework for automated synthesis of adversarial transactions 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 approximation 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 novel counterexample driven approximation refinement technique. We implement our framework in a tool named FlashSyn. We evaluate FlashSyn on 16 DeFi protocols that were victims to flash loan attacks and 2 DeFi protocols from Damn Vulnerable DeFi challenges. FlashSyn automatically synthesizes an adversarial attack for 16 of the 18 benchmarks, demonstrating its effectiveness in finding possible flash loan attacks.

CCS CONCEPTS
• Security and privacy → Software security engineering; • Software and its engineering → Software testing and debugging.

KEYWORDS
program synthesis, program analysis, blockchain, smart contracts, vulnerability detection, flash loan

1 INTRODUCTION
Blockchain technology enables the creation of decentralized, resilient, and programmable ledgers on a global scale. Smart contracts, which can be deployed onto a blockchain, allow developers to encode intricate transaction rules that govern the ledger. These features have made blockchains and smart contracts essential infrastructure for a variety of decentralized financial services (DeFi). As of April 1st, 2023, the Total Value Locked (TVL) in 1,417 DeFi smart contracts had reached 50.15 billion [22].

However, security attacks are critical threats to smart contracts. Attackers can exploit vulnerabilities in smart contracts by sending malicious transactions, potentially stealing millions of dollars from users. Particularly, a new type of security threat has emerged, exploiting design flaws in DeFi contracts by leveraging large amounts of digital assets. These attacks, commonly referred to as flash loan attacks [54, 67, 82], typically involve borrowing the required large amount of assets from flash loan contracts. Among the top 200 costliest attacks recorded in Rekt Database, the financial loss caused by 36 flash loan attacks exceeds 418 million USD [54].

A typical flash loan attack transaction consists of a sequence of actions, or function calls to smart contracts. The first action involves borrowing a substantial amount of digital assets from a flash loan contract, while the last action returns these 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 created by the large volume of the 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 trick the contract to make unfavorable trades with the attacker [12, 59]. Although researchers have developed many automated program analysis and verification techniques [2, 36, 45, 56] 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: We present 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 in these contracts, FlashSyn automatically synthesizes an action sequence along with all action parameters to interact with the contracts to exploit potential flash loan vulnerabilities. Additionally, FlashSyn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
can analyze past blockchain transaction history, assisting users in identifying candidate actions for synthesis.

The primary challenge FlashSyn faces is that the underlying logic of DeFi actions is often too sophisticated for standard solvers to handle. Even if the action sequence was already known, a naive application of symbolic execution might not be able to find action parameters because it may need to extract overly complicated symbolic constraints causing the 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 challenges with its novel synthesis-via-approximation technique. Instead of attempting to extract accurate symbolic expressions from smart contract code, FlashSyn collects data points to approximate the effect of contract functions with numerical methods. FlashSyn then uses the approximated expressions to drive the synthesis. FlashSyn also incrementally improves the approximation with its novel counterexample driven approximation refinement techniques, i.e., if the synthesis fails because of a large deviation caused by the approximations, FlashSyn collects the corresponding data points as counterexamples to iteratively refine the approximations. The combination of these techniques allows the underlying optimizer of FlashSyn to work with more tractable expressions. 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 16 DeFi benchmark protocols that were victims to flash loan attacks and 2 DeFi benchmark protocols from Damn Vulnerable DeFi challenges [19]. FlashSyn synthesizes adversarial attacks for 16 out of the 18 benchmarks. For comparison, a baseline with manually crafted accurate action summaries only synthesizes attacks for 7 out of the 18.

**Contributions:** This paper makes the following contributions:

- **FlashSyn:** The first automated end-to-end program synthesis tool for detecting flash loan attack vulnerabilities. It enables approximate attack synthesis without diving into sophisticated logics of DeFi contracts.
- **Synthesis-via-approximation:** A novel synthesis-via-approximation technique to handle sophisticated logics of DeFi contracts.
- **Counterexample Driven Approximation Refinement:** A novel counterexample driven approximation refinement technique to incrementally improve the approximation during the synthesis process.
- **Experimental Evaluation:** We implemented FlashSyn in a tool and evaluated it on 16 protocols that were victims to flash loan attacks and 2 fictional flash loan attacks.

Our solution, FlashSyn has been adopted and further developed by Quantstamp, a leading smart contract auditing company for the detection of flash loan vulnerabilities in DeFi contracts [53, 60, 61].

2 **BACKGROUND**

**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 determining block order. Each block is constituted of transactions. Ethereum [11, 77] is the first blockchain to support, store, and execute Turing complete programs, known as smart contracts. Many new blockchains use the Ethereum virtual machine (EVM) for execution due to its popularity among developers.

**Smart Contracts:** 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 bytecode (e.g., EVM bytecode [11, 77]) that executes incoming transactions to change the storage and balance. Users interact with a smart contract by issuing transactions from their user accounts to the contract address. Smart contracts can also interact with other smart contracts as function calls. Currently, there are several human-readable high-level programming languages, e.g., Solidity [37] and Vyper [38], to write smart contracts that compile to the EVM bytecode.

**Decentralized Finance (DeFi):** DeFi is a peer-to-peer financial ecosystem built on top of blockchains [78]. The building blocks of DeFi are smart contracts that manage digital assets. A few DeFi protocols dominate the DeFi market and serve as references for other decentralized applications: stable coins (e.g., USDC and USDT), price oracles, decentralized exchanges, and lending and borrowing platforms. In DeFi, a special type of loan called flash loan allows lenders to offer loans to borrowers without upfront collateral deposits. The loan is only valid within a single transaction and must be repaid with fees before the completion of the transaction.

3 **ILLUSTRATIVE EXAMPLE**

We next present a motivating example to describe the complexity of flash loan attacks and our proposed approach to synthesize them.

**Background:** On October 26th, 2020, an attacker exploited the USDC and USDT vaults of Harvest Finance, causing a financial loss of about 33.8 million USD. In this section, we will focus on the attack on the USDC vault. The attacker repeatedly executed the same attack vector 17 times targeting the USDC vault. Fig. 1 summarizes the attack vector. The attack vector contains a sequence of actions that interact with the following contracts:

- **Uniswap:** Uniswap is a protocol with flash loan services.
- **Curve:** Curve is an exchange protocol for stable coins like USDT, USDC, and DAI, whose market prices are close to one USD. It maintains pools of stable coins and users can interact with these pools to exchange one kind of stable coins to another. For example, Y Pool in Curve contains both USDC and USDT. Users can put USDC into the pool to exchange USDT out. The exchange rate fluctuates around one, which is determined by the current ratio of USDC and USDT in the pool. Note that internally Y pool automatically deposits USDT and USDC to Yearn, keeps yUSDT and yUSDC tokens, and retrieves them back when the users withdraw. We omit this complication for simplicity.
- **Harvest:** Harvest is an asset management protocol and the victim contract of this attack. Users can deposit USDC and USDT into Harvest and receive rUSDC and rUSDT tokens which users can later use to retrieve their deposit back. Harvest will invest the deposited USDC and USDT from users to other DeFi protocols to generate profit. Note that the exchange rate between rUSDC and USDC is also not fixed. It is determined by a vulnerable closed

\[1\text{Yearn is a DeFi protocol that generates yield on deposited assets. yTokens of Yearn represent the liquidity provided in a Yearn product.}\]
source oracle contract, which ultimately uses Curve Y pool ratios to calculate the exchange rate.

**Attack Actions:** The attack vector shown in Fig. 1 first flash loaned 18.3M USDT and 50M USDC, then called 4 methods (actions) to exploit the design flaw in Harvest. The first action, action 1, swaps 17,222,012 USDT for 17,216,703 USDC via Curve.Fi Y Pool. Action 1 reduces the estimated value of USDT based on the ratio in Y Pool as the amounts of USDT in Y Pool considerably increases. This in turn reduces Harvest Finance’s evaluation of its invested assets. Action 2 deposits 49,977,468 USDC into Harvest Finance USDC vault and due to the reduced evaluation of the invested underlying assets, the attacker receives 51,456,280 fUSDC back, which is abnormally large. Similar to action 1, action 3 then swaps 17,239,234 USDC back to 17,230,747 USDT via Curve.Fi Y Pool, which normalized the manipulated USDT/USDC rate in the pool. It also brings Harvest Finance’s evaluation of its invested underlying asset back to normal. Finally, action 4 withdraws 50,298,684 USDC (using 51,456,280 fUSDC) from Harvest Finance USDC vault. Assuming 1 USDC = 1 USD, the profit of the above attack vector is 307,420 USD.

This attack is a typical case of oracle manipulation. The exploiter manipulated the USDT/USDC rate in Curve.Fi Y Pool by swapping a large amount between USDC and USDT back and forth, which caused Harvest Finance protocol to incorrectly evaluate the value of its asset, leaving large arbitrage space for the exploiter. The actions sequence and particularly the parameters are carefully chosen by the attacker to yield best profit. There are multiple challenges FlashSyn faces to synthesize this attack.

**Challenge 1 - Sophisticated Interactions:** The attack involves several smart contracts that interact with each other and with other contracts outside the attack vector. The state changes caused by one action influence the behavior of other actions. This makes the synthesis problem of finding an attack vector more complicated as it requires finding roots of a quintic equation. In the actual implementation, D is calculated iteratively on the fly via Newton’s method (see extended version [13] Appendix A).

To demonstrate the complexity of DeFi protocols, we run an experiment with Manticore [56], a symbolic execution tool for smart contracts, to execute the function get_D, for computing D as shown in Fig. 2, with symbolic inputs and explore all possible reachable states. Manticore fails and throws a solver-related exception together with an out of memory error. We then simplified get_D by removing the outer for loop and bounding the length of xp to 2, Manticore still fails and throws the same error.

### 3.1 Apply FlashSyn

We will now show how FlashSyn synthesizes the Harvest USDC vault attack from the identified set of actions listed in Table 1.4 The first two input arguments to exchange specify the token types to be swapped. The third argument specifies the quantity to swap.5 Table 1 lists each action’s token flow, along with the number of data points collected initially (without counterexamples) and the total number of data points for polynomial and interpolation, respectively. The amounts of tokens transferred in/out for each action are calculated based on its contract’s member variables or read-only functions. We refer these variables and functions as states of an action. A prestate refers to the state before the execution of an action. Executing an action will also alter states, which are denoted as poststates. The states not altered by any action are ignored.

For example, exchange(USDT, USDC, v) leverages two states, balances[USDC] and balances[USDT], to calculate the amounts of token exchanges. Upon execution, this function also modifies

```plaintext
function get_D(uint[] xp) returns (uint):
  1 function get_D(uint[] xp) returns (uint):
  2 uint N_COINS = xp.length;
  3 uint S = sum(xp);
  4 // ...
  5 uint D = S;
  6 uint Ann = A * N_COINS; // A is a constant
  7 for (uint i = 0; i < 255; i = i + 1) {
    8 uint 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);
    }
    12 uint Dprev = D;
    13 D = ((Ann * S + D_P * N_COINS) * D) / ((Ann - 1) * S + D_P * N_COINS + 1) + D_P;
    14 if (abs(D - Dprev) <= 1) break;
  16 return 0;
```

Figure 1: Harvest USDC Vault Price Manipulation Attack.

Figure 2: get_D Method to Compute D.

instance, actions 2 and 4 swap an amount of token i to token j, while maintaining the following StableSwap invariant [23]:

\[
A \cdot n^n \sum_i x_i + D = A \cdot n^n \cdot D + \frac{D^{n+1}}{n+1} \cdot x_i
\]

where A is a constant, n is number of token types in the pool(4 for Curve.Fi Y Pool).6 \(x_i\) is token i’s liquidity, D is the total amount of tokens at equal prices. There does not exist a closed-form solution for D as it requires finding roots of a quintic equation. In the actual implementation, D is calculated iteratively on the fly via Newton’s method (see extended version [13] Appendix A).

3Ethereum address: 0x45f783cce6b7ff23b2ab2d70e416cdb7d6055f51.
4In Section 6, we present FlashFind to automatically find the set of candidate actions that are used here by FlashSyn to synthesize an attack vector.
5Note that in the implementation the actual name of the exchange method is exchange_underlying, 1 and 2 are used to identify the tokens USDC and USDT, respectively, and the method has a fourth argument to specify the minimal quantity expected to receive from the swapping.

6Ethereum address: 0x55655f5f5f5f5f5f5f5f5f5f5f5f5f5f5f5f5f5f.
these two states. Consequently, these two states act as both the prestates and poststates of exchange(USDT, USDC, v).

**Initial Approximation:** To generate the initial approximation of the state transition functions of each action, FlashSyn first collects data points where each data point is an input-output pair. The input is the action’s prestates and parameters, and the output is its poststates and the outputted values. To collect data points, FlashSyn executes the associated contracts on a private blockchain (a forked blockchain environment) with different parameters to reach input-output pairs with different prestates and poststates. FlashSyn then uses the collected data points to find the approximated state transition functions. We consider two techniques to solve the above multivariable approximation problem: linear regression based polynomial features and nearest-neighbor interpolation [55, 64]. The following example is one of exchange(USDT, USDC, v)’s state transition functions approximated by polynomials: $s_i' = 0.73244455x_{i1} - 0.23655202x_{i2} - 0.859155311x_{i3} + 27351279.416023515$ where $s_{i1}$ and $s_{i2}$ are the prestate and poststate balances[USDC], $s_{i3}$ is the prestate balance[USDT], and $v$ is the third argument of the action exchange(USDT, USDC, v).

**Enumerate and Filter Action Sequences:** After capturing an initial approximation of state transition functions, FlashSyn leverages an enumeration-based top-down algorithm to synthesize different action sequences. FlashSyn applies several pruning heuristics to filter unpromising sequences. For each enumerated action sequence, FlashSyn uses the approximated state transition functions to construct an optimization problem, consisting of constraints and an objective function that represents profit. FlashSyn then applies an off-the-shelf optimizer to obtain a list of parameters that maximize the profit estimated using approximated transition functions.

**Counterexample Driven Refinement:** After obtaining a list of parameters that maximize the estimated profit of an action sequence, FlashSyn proceeds to verify the synthesized attack vectors by executing them on a private blockchain and check their actual profits. If the difference between the actual profit and the estimated profit of an attack vector is greater than 5%, FlashSyn reports it as a counterexample, indicating inaccuracy of our approximated transition functions. To correct this inaccuracy, FlashSyn employs counterexample driven approximation refinement technique. FlashSyn utilizes the reported counterexamples to collect new data points and refine the approximations. The revised approximations are subsequently used to search for parameters in next loops. For example, FlashSyn with polynomial approximations collects 238 additional data points for the action exchange(USDT, USDC, v) throughout 7 refinement loops.

**Synthesized Attack:** In the Harvest USDC example, FlashSyn successfully found the following attack vector that yields an adjusted profit of 110051 USD using the interpolation technique with the counterexample driven refinement loop.

\[
\text{exchange}(USDT, USDC, 15192122) \cdot \text{deposit}(45105321) \cdot \text{exchange}(USDC, USDT, 11995404) \cdot \text{withdraw}(46198643)
\]

### 4 PRELIMINARY

**Labeled Transition Systems (LTS).** We use LTS to model behaviors of smart contracts. A LTS $A = (Q, \Sigma, q_0, \delta)$ over the possibly-infinite alphabet $\Sigma$ is a possibly-infinite set of states with an initial state $q_0 \in Q$, and a transition relation $\delta \subseteq Q \times \Sigma \times Q$.

**Execution.** An execution of $A$ is a sequence of states and transition labels (actions) $\rho = q_0, a_0, q_1, \ldots, a_{k-1}, q_k$ for $k > 0$ such that $\delta(q_i, a_i, q_{i+1})$ for each $0 \leq i < k$. We write $\rho(t_2 \cdot s_{j-1}) \cdot A q_j$ to denote the subsequence $q_1, a_1, \ldots, q_{j-1}, a_{j-1}, q_j$ of $\rho$.

**Invocation Label.** Formally, an invocation label $adv.m(\bar{u})$ consists of a method name $m$ of a contract address $adr$, accompanied by a vector $\bar{u}$ containing argument values.

**Operation Label.** An operation label $l = adv.m(\bar{u})$ corresponds to a state $Q_{adv}$ in a blockchain along with a return value $v$, and $l$ is a sequence of operation labels representing the “internal” calls made during the invocation of $m$. The distinguished invocation outcome $\perp$ is associated to invocations that revert.

**Interface.** The interface $Q_{adr}$ is the set of non-read-only operation labels in the contract $adv$. We assume w.l.o.g. that the preconditions are satisfied for all the operations in $Q_{adr}$; otherwise, the external invocation $adv.m(\bar{u})$ reverts. $Q_{adr}$ is a superset of the set of action candidates of FlashSyn.

**Smart Contract.** A smart contract at an address $adr$ is an LTS $\mathcal{C}_{adr} = (Q_{adr}, Q_{adr}, q_{0,adr}, \delta_{adr})$ over the interface $Q_{adr}$ where $Q_{adr}$ is the set states and $\delta_{adr}$ is the transition relation.

**Symbolic Actions Vector.** We define the notion of a symbolic actions vector $S = \ell_{adv1} \cdot \ldots \cdot \ell_{advn}$ s.t. $\ell_{advi} \in \Sigma$ for $1 \leq i < n$ as the sequence of operation labels (possibly from different contracts) associated with the execution $\rho$, i.e., $\rho = q_0, \ell_{adv1}, q_1, \ldots, \ell_{advn}, q_n$.

**Balance.** We define the balance of address $adv$ in a blockchain state $q$ as the mapping $B : Q \times A \Rightarrow V$ that maps the pair $(q, adv) \in Q \times A$ to the weighted sum of tokens the address $adv$ holds at $q$, i.e., $B(q, adv) = \sum_{t \in T} M(q, adv, t) \cdot P(t)$, where $T$ represents tokens held by $adv$, $M(q, adv, t)$ represents the amount of token $t$ held by $adv$ at the blockchain state $q$, and $P(q, t)$ represents the price of token $t$ at the blockchain state $q$.

**Attack Vector.** An attack vector by an adversary $adv$ consists of a symbolic actions vector $S$ where the symbolic arguments are replaced by concrete values (integer values) and $S$ transforms a blockchain state $q$ to another state $q'$ such that $B(q', adv) - B(q, adv) > 0$, i.e., the adversary $adv$ generates profit when the sequence of actions $S$ is executed with the concrete values.

**Problem formulation.** Given a specification $\varphi$ (which contains vulnerable contract addresses or action candidates) and a blockchain state $q$, the objective is to find an attack vector consisting of a concretization of the symbolic actions vector $S = \ell_{adv1} \cdot \ldots \cdot \ell_{advn}$ s.t. $\ell_{advi} \in \Sigma \otimes \varphi$ for $1 \leq i < n$, transforming the state $q$ to a state $q'$, and that maximizes the profit of an adversary $adv$, $B(q', adv) - B(q, adv)$.
5 FLASHSYN

Algorithm 1 gives the overall synthesis procedure of FlashSyn. FlashSyn first collects initial data points to approximate the actions in \( \text{Act} \) (line 3) where FlashSyn uses the state \( q \) as a starting blockchain state. Then, using the sub-procedure \text{Approximate} FlashSyn generates the approximations \text{ApproxAct} of the actions in \( \text{Act} \) using the collected data points (line 5). FlashSyn uses the sub-procedure \text{ActionsVectors} to generate all possible symbolic actions vectors of length less than len (line 6). FlashSyn then iterates over the generated actions vectors and uses some heuristics implemented in the sub-procedure \text{IsFeasible} to prune actions vectors (line 8). For instance, an actions vector containing two adjacent actions invoking the same method can be pruned to an actions vector where the two adjacent actions are merged. Afterwards, using the actions vector and approximated transition functions, the sub-procedure \text{Construct} constructs the optimization framework \( \mathcal{P} \) for the actions vector (line 9). Then, FlashSyn uses the optimization sub-procedure \text{Optimize} (line 10) to find the optimal concrete values to pass as input parameters to the methods in the actions vector that satisfy the constraints of \( \mathcal{P} \). FlashSyn then validate whether the attack vector generated by the optimizer indeed generates the profit with the sub-procedure \text{QueryOracle} to execute the generated attack vectors on the blockchain. If the query is successful, i.e., the actual profit closely matches the profit found by the optimizer, FlashSyn adds the attack vector to the list of discovered attacks. Otherwise, FlashSyn considers the attack vector to be a counterexample, and uses it to generate new data points to refine the approximation in the subsequent iterations, within the sub-procedure \text{CEGDC} (referenced in line 14 and introduced later in Section 5.3). FlashSyn repeats the process until the number of iterations reaches \( n \) (line 4).

\begin{algorithm}
\caption{Attack vectors synthesis procedure. Its inputs are actions \( \text{Act} \), the maximum length len, a blockchain state \( q \), and a threshold number of iterations \( n \). Its outputs are attack vectors that yield profits.}
\begin{algorithmic}[1]
\Procedure{Synthesize}{$\text{Act}, \text{len}, P, q, n$}
\For {$a \in \text{Act}$}
\State datapoints[$a$] $\leftarrow$ \text{DataCollect}(q, a);
\EndFor
\For {$i \in [0; n]$}
\State \text{ApproxA}ction $\leftarrow$ \text{Approximate}(\text{Act}, datapoints);
\State \text{wlist} $\leftarrow$ \text{ActionsVectors}(\text{ApproxAct}, len);
\EndFor
\For {$p \in \text{wlist}$}
\If {\text{IsFeasible}(p)}
\State \text{p'} $\leftarrow$ \text{Construct}(p, \text{ApproxA}ction);
\State \text{(p', profit)} $\leftarrow$ \text{Optimize}(p, \mathcal{P})
\State \text{if QueryOracle}(q, p', \text{profit})
\State \text{answerlist}.add($\text{p', profit}$);
\Else
\State datapoints = datapoints \cup \text{CEGDC}(p', q);
\EndIf
\EndFor
\State \Return \text{answerlist};
\EndProcedure
\end{algorithmic}
\end{algorithm}

5.1 Pruning Symbolic Actions Vectors

The sub-procedure \text{IsFeasible} implements some heuristics to prune undesired symbolic actions vectors.

\textbf{Heuristic 1: no duplicate adjacent actions.} Two successive calls to the same method in a DeFi smart contract are usually equivalent to a single call with larger parameters. Thus, we discard actions vectors containing duplicate, successive actions.

\textbf{Heuristic 2: limited usage of a single action.} Using the observation that attack vectors do not contain repetitions, we fix a maximum number of calls to a single method an attack can contain and discard actions vectors that do not satisfy this criterion, e.g., an actions vector of length 4 cannot contain more than 2 calls to the same method.

\textbf{Heuristic 3: necessary preconditions.} Based on the observation that owning certain tokens is a necessary precondition for invoking some actions, FlashSyn prunes symbolic actions vectors that contain actions requiring tokens\(^6\) not yet owned. In Harvest USDC example, invoking \texttt{withdraw} method requires users own some share tokens (fUSDC) beforehand. The only action candidate that mints fUSDC for users is \texttt{deposit}; thus, this heuristic mandates that \texttt{deposit} must be called before invoking \texttt{withdraw}.

5.2 Optimization

Given a symbolic actions vector and their approximated transit functions, the sub-procedure \text{Construct} constructs an optimization framework to find optimal values for the parameters for the actions. Recall that given a blockchain state \( q \) and an address \( \text{adr} \), the actions vector \( S \) transforms \( q \) to another state \( q' \). The objective function in the optimization problem targets to increase the tokens values in the balance of the address \( \text{adr} \), i.e., \( y = B(q', \text{adr}) - B(q, \text{adr}) \). The optimization problem is accompanied by constraints on the symbolic values to be inferred. For instance, the balance of any token \( t \) for any address \( \text{adr}' \) must always be non-negative, i.e., the adversary and the smart contracts cannot use more tokens than what they have in their balances, otherwise the transaction reverts. In the following, we give the definition of the optimization problem.

\[\mathcal{P} : \max_{p_0, p_1, \ldots, p_n} y = B(q', \text{adr}) - B(q, \text{adr}) \]

subject to: \( \forall t \in T, \text{adr}' \in A. M(q', \text{adr}', t) \geq 0 \)

5.3 Counterexample Guided Data Collection

The optimization sub-procedure might explore parts of the states space not explored during the initial data points collection. This might challenge the accuracy of the approximations and result in mismatch between the estimated and the actual values. 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. Therefore, we propose counterexample guided data collection (CEGDC), inspired of counterexample guided abstraction refinement [17], to refine approximations when mismatches are identified.

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

\(^6\)Note a token can be standard tokens (ERC20, BEP20), or any other forms of tokens represented by integers such as asset tokens, debt tokens or share tokens.
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, \(|\mathcal{P}_e(C) - \mathcal{P}_a(C)| \geq \varepsilon \cdot (|\mathcal{P}_e(C)| + |\mathcal{P}_a(C)|)\), where \( \varepsilon \) is a small constant representing accuracy tolerance.

**Algorithm 2:** Counterexample guided data collection procedure. It takes a counterexample \( C \) and a state \( q \), and returns datapoints. \( k \in [n; 1] \) means that in the first iteration \( k = n > 0 \).

1. procedure CEGDC(C, q)
2. datapoints ← \{
3. for each \( k \in [\text{len}(C); 1] \):
4. \( q_k' \leftarrow \text{Estimate}(q, C, k) \);
5. \( q_k'' \leftarrow \text{Execute}(q, C, k) \);
6. if \( \text{IsAccurate}(q_k', q_k'') \)
7. returns datapoints ;
8. else
9. \((a, \text{paras}) \leftarrow C[k] \);  
10. datapoints[\( a \)] \leftarrow (q, \text{paras}, q_{k''}) ;
11. return datapoints;

In Algorithm 2, we present the sub-procedure CEGDC for collecting new data points from a counterexample. CEGDC takes as inputs a counterexample \( C \) which is known to have an inaccurate profit estimation, and a blockchain state. The for loop on line 3 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 \( k \), FlashSyn checks if the estimated methods of the action at the index \( k \) of \( C \) are accurate. First, FlashSyn computes the estimated state \( q_k' \) reached by executing \( C \) until reaching the action indexed \( k \) (line 4) using the approximated transition functions. Second, FlashSyn fetches the actual state \( q_k'' \) reached by executing \( C \) until reaching the action indexed \( k \) (line 5) on the actual smart contracts on the blockchain. Then, FlashSyn compares the estimated and actual execution results (line 6). If the estimation is accurate, this indicates that the 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 data points computed in the previous iterations (line 7). Otherwise, it indicates inaccurate transition functions of this action or/and its predecessors. Thus, we add a new data point associated with the action at the index \( k \) of \( C \) (lines 9 and 10) and proceed to the next iteration of the loop to explore the action predecessors.

### 5.4 FlashFind: Action Candidates Identification

Flash loan attacks typically focus on victim contracts containing functions capable of transferring tokens,\(^7\) which can be invoked by regular users. The attacker manipulates the transfer amount under specific conditions to make profit.

We designed and built a tool FlashFind to assist users of FlashSyn to select action candidates likely to be involved in an attack vector. Given a set of target smart contracts, FlashFind identifies action candidates in the following steps.

**5.4.1 Selecting Action Candidates from Contract Application Binary Interfaces (ABIs).** The Application Binary Interface (ABI) serves as an interpreter enabling communication with the EVM bytecode. For all verified smart contracts, their ABIs are publicly available.

An ABI typically comprises (public or external) function names, argument names/types, function state mutability, and return types. During the process of selecting action candidates, certain functions can be safely ignored: (1) Functions with the view or pure mutability can be excluded. (2) Functions that can only be invoked by privileged users, such as `transferOwnership` and `changeAdmin`, can also be disregarded since they are unlikely to be accessed by regular users. (3) Token permission management functions/parameters, such as the function `approve` or parameter `deadline`, are excluded.\(^8\) These functions/parameters solely control whether a transaction will be reverted or not and do not affect the behaviors of contracts.

**5.4.2 Learning Special Parameters from Transaction History.** After selecting a set of action candidates from contracts’ ABIs, some non-integer parameters (e.g., bytes, string, address, array, enum) can still be unknown. FlashFind collects past transactions of the target contracts and extracts function execution level trace data from these transactions, and utilizes the trace data to learn the special parameters from previous function calls made to the contract.

**5.4.3 Local Execution and Intra-dependency Analysis.** After learning special parameters, each action candidate is executed at a given block to verify its executability. An action candidate may not be executable due to various reasons: (1) the function is disabled by the owner or admin; (2) internal function calls to other contracts are disabled by the owners or admins of those contracts; (3) the function is not valid under current blockchain states. The inexecutability due to these reasons cannot be identified by static analysis and can only be determined by executions. All such inexecutable functions are filtered out.

FlashFind automatically collects storage read/write information during the execution of these functions and infers the Read-After-Write (RAW) dependencies\(^9\) between different action candidates. An action A has a RAW dependency (or equivalently, is RAW dependent) on action B if the execution of action A reads the storage written by action B. From the RAW dependencies, it is possible to observe that certain functions behave independently, meaning they do not have any RAW dependencies on other functions, and other functions do not have any RAW dependencies on them either. Consequently, these independent functions can be safely ignored.

After analyzing ABIs, transaction history, and local executions, FlashFind generates a list of action candidates with only their integer arguments left undetermined. These action candidates are then input into FlashSyn for further synthesis.

### 6 IMPLEMENTATION

FlashSyn is implemented in Python. Figure 3 shows an overview of our implementation. The components `Runner`, `Synthesizer`,

---

\(^7\)Here, tokens refer to various forms of DeFi tokens, including stable coins, debt tokens, share tokens, liquidity tokens, asset tokens, etc.

\(^8\)Previous works [44, 50] have extensively researched access control vulnerabilities. We exclude them from the scope of this work.

\(^9\)To simplify the search process of FlashSyn, these permissions are assumed to be granted maximally.

\(^10\)This RAW dependency information is also employed in FlashSyn’s initial data collection to expand the range of data points.

\(^11\)It is important to note that this step excludes any tx.origin/msg.sender-related storage reads/writes, as such storage accesses do not alter the global state of the protocol and are therefore unlikely to impact the functional behaviors of actions.
We built (by simple inspection of smart contract codes), or the action which require an initial guess of parameters. Under our settings, it is not feasible to scipy.optimize.shgo find an initial guess for every symbolic attack, as each attack behaves differently.

Multiple symbolic actions vectors in parallel. 18 parameters.

If the estimation is inaccurate, the attack vector is treated as a Runner cut by approximator. Obtained attack vector that yields a positive profit is then executed by Approximator. After Runner validates the results, Approximator ranks symbolic actions vectors and drops the ones with low priority scores. The actions vector with high priority score will be searched with higher strengths. Specifically, if a symbolic actions vector yields a positive profit \( P \) in iteration \( k \), its priority score of iteration \( k + 1 \) is \( P' \). If a symbolic actions vector does not yield a positive profit in iteration \( k \), it is given a small priority score between 1 and 10 based on Optimizer results. An actions vector will also be dropped when its priority score does not increase between iterations. When all actions vectors are dropped, the whole synthesis procedure stops, and FlashSyn returns all the profitable attack vectors it found.

(5) FlashFind. We also implemented FlashFind as an optional component. FlashFind uses TrueBlocks [65] and blockchain explorers [9, 34, 42] to collect past transactions of the target contracts. FlashFind employs Phalcon [1] and Foundry [39] to extract function level trace data from those transactions and perform analysis on storage accesses. Our evaluation shows that FlashFind is able to identify action candidates and helps FlashSyn discovers alternative attack vectors (see RQ4 in Section 7).

FlashSyn does not require prior knowledge of a vulnerable location or contract. Given a set of DeFi lego user interface contracts, action candidates and their special parameters such as strings are given by the users or automatically extracted from transaction history using FlashFind. FlashSyn utilizes these action candidates to synthesize attack vectors and search for optimal numerical values. Note that these action candidates are not necessarily the ones that contain the vulnerability. Rather, they serve as user interfaces for interacting with the protocol. The vulnerability may reside in any contract invoked through a user interacting with the protocol, and is generally available for invoking these functions. Such information is essential for any user interacting with the protocol, and is generally available in the documents of DeFi protocols.

7 EVALUATION

We aim to answer the following research questions:

RQ 1: How effective is FlashSyn in synthesizing flash loan attack vectors?
Table 2: Benchmark of Attacks Used in the Evaluation.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>#C</th>
<th>LoC</th>
<th>Vulnerability Type</th>
<th>Loss in History</th>
</tr>
</thead>
<tbody>
<tr>
<td>b2x1</td>
<td>6</td>
<td>4964</td>
<td>pump &amp; arbitrage</td>
<td>25</td>
</tr>
<tr>
<td>Harvest_USDT</td>
<td>6</td>
<td>6446</td>
<td>manipulate oracle</td>
<td>28</td>
</tr>
<tr>
<td>Harvest_USDC</td>
<td>6</td>
<td>4095</td>
<td>manipulate oracle</td>
<td>29</td>
</tr>
<tr>
<td>Eminence</td>
<td>2</td>
<td>489</td>
<td>design flaw</td>
<td>27</td>
</tr>
<tr>
<td>ValueDeFi</td>
<td>8</td>
<td>7043</td>
<td>manipulate oracle</td>
<td>31</td>
</tr>
<tr>
<td>Cheesebank</td>
<td>12</td>
<td>1246</td>
<td>manipulate oracle</td>
<td>26</td>
</tr>
<tr>
<td>Warp</td>
<td>11</td>
<td>13139</td>
<td>manipulate oracle</td>
<td>32</td>
</tr>
<tr>
<td>Yarn</td>
<td>5</td>
<td>2200</td>
<td>forced investment</td>
<td>33</td>
</tr>
<tr>
<td>inverseFi</td>
<td>7</td>
<td>5734</td>
<td>manipulate oracle</td>
<td>30</td>
</tr>
<tr>
<td>bEarnFi</td>
<td>3</td>
<td>3007</td>
<td>asset mismatch</td>
<td>3</td>
</tr>
<tr>
<td>AutoShark</td>
<td>6</td>
<td>8052</td>
<td>design flaw</td>
<td>4</td>
</tr>
<tr>
<td>ElevenFi</td>
<td>7</td>
<td>5613</td>
<td>design flaw</td>
<td>6</td>
</tr>
<tr>
<td>ApoRocket</td>
<td>7</td>
<td>1562</td>
<td>design flaw</td>
<td>3</td>
</tr>
<tr>
<td>WDoge</td>
<td>2</td>
<td>788</td>
<td>deflationary token</td>
<td>8</td>
</tr>
<tr>
<td>Novo</td>
<td>4</td>
<td>7080</td>
<td>design flaw</td>
<td>7</td>
</tr>
<tr>
<td>OneRing</td>
<td>14</td>
<td>5386</td>
<td>design flaw</td>
<td>41</td>
</tr>
<tr>
<td>Puppet</td>
<td>2</td>
<td>742</td>
<td>manipulate oracle</td>
<td>20</td>
</tr>
<tr>
<td>PuppetV2</td>
<td>1</td>
<td>161</td>
<td>manipulate oracle</td>
<td>21</td>
</tr>
<tr>
<td>ETH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>82.5 million USD</td>
</tr>
</tbody>
</table>

*: #C denotes number of the victim protocol’s contracts invoked in exploits.
*: LoC denotes total number of lines of code in the contracts identified by #C, excluding closed-source contracts.
†: The logic designs of one or more functions in the victim contracts is flawed, with highly specific case-by-case vulnerabilities.

RQ 2: How well does the synthesis-via-approximation technique perform compared to precise baselines?

RQ 3: How much does counterexample driven approximation refinement improve FlashSyn’s results?

RQ 4: How effective is the combination of FlashFind and FlashSyn to synthesize attack vectors end-to-end?

Scope: FlashSyn focuses on flash loan attacks that generate positive profit by sequentially invoking functions within existing DeFi contracts. Security attacks that require exploiting other vulnerabilities such as re-entrance or conducting social engineering are outside the scope of FlashSyn and our evaluation. The goal of FlashSyn is to prove the existence and the exploitability of flash loan vulnerabilities. Consequently, activities such as getting and repaying the flash loan are not part of our synthesis task.

Benchmarks: We investigated historical flash loan attacks that span from 02/14/2020 to 06/16/2022 across Ethereum, Binance Smart Chain (BSC), and Fantom (FTM) and attempted to reproduce each of them in our environment. In the end, we reproduced 16 attacks that are within our scope and collected them as our benchmark attacks. These attacks invoked 2-14 contracts of the victim protocol in the nested invocation tree per attack, consisting of a total of 489 to 13,139 lines of code, reflecting the multifaceted nature of the DeFi protocols exploited in real-world flash loan attacks. Also, protocols in our benchmark contain up to 15 action candidates from which FlashSyn needs to find an attack vector. Altogether, the 16 historical flash loan attacks in our benchmark have caused over 82.5 million US dollars in losses and include widely-known cases such as Harvest, b2x, and Eminence. Additionally, we include 2 fictional attacks from the Damn Vulnerable DeFi (DVD) challenges [19].

Ground Truth: For historical flash loan attacks, we forked the corresponding blockchain at one block prior to the attack transaction and replayed the attacker’s attack vector as the ground truth. For DVD benchmarks, we select community solutions as ground truth. Note that in a flash loan attack, if the same attack vector is repeated multiple times, we remove the loop and only consider the first attack vector as the ground truth.

Precise Baseline: To demonstrate the effectiveness of synthesis-via-approximation techniques, we implemented a baseline synthesizer that works with manual summaries of smart contract actions. Specifically, we manually inspected all benchmarks whose relevant smart contracts that are all open-source and for each benchmark we allocated more than 4 manual analysis hours to extract the precise mathematical summaries. The baseline synthesizer then uses the manually extracted precise summaries to drive the synthesis.

Environment Setup: We assume that the flash loan providers are generally available, and we do not consider the borrow and the return as part of the synthesis task. To facilitate FlashSyn experimentation, we manually annotated the prestates and poststates for each action. The details of this annotation process are described in the extended version [13] Appendix C. Although this manual effort is required, it’s worth noting that automation of this step is possible. Techniques such as dynamic taint analysis and forward symbolic execution can be employed to automatically identify which storage variables influence the change in token balances, thereby streamlining the annotation of prestates and poststates. The experiments are conducted on an Ubuntu 22.04 server, with an AMD Ryzen Threadripper 2990WX 32-Core Processor and 128 GB RAM.

Experiment Overview. To answer RQ 1, we apply FlashSyn to the 18 benchmarks with the same set of candidate actions in ground truths. For each candidate action, the prestates and poststates are annotated for FlashSyn to drive the approximated formula for this action. We set a timeout of 3 hours for FlashSyn-poly and 4 hours for FlashSyn-inter. FlashSyn does not know a priori whether a benchmark has an attack vector with a positive profit, and it does not set any bounds on the profit. It tries iteratively to synthesize an attack vector with a maximum profit. FlashSyn’s refinement loop is guided by intermediate results and FlashSyn stops when it cannot improve the profit or the above timeouts are reached. To answer RQ 2, we replace Approximator component of FlashSyn with manually extracted precise mathematical summaries, and conduct the same experiment with 4 hours timeout. To answer RQ 3, we evaluate FlashSyn with different initial data points and with CEGDC enabled/disabled. To answer RQ 4, we first use FlashFind to identify candidate actions from given contract addresses which the hacker used in history, manually annotate them as in RQ 1, and then apply FlashSyn with this new set of candidate actions to synthesize attack vectors under the same setting as in RQ 1. The results for RQ 1+RQ 2, RQ 3, RQ 4 are summarized in Table 3, Table 4, and Table 5, respectively.

RQ1: Effectiveness of FlashSyn. Table 3 summarizes the results of the experiment. The first five columns of Table 3 list benchmark information including the number of actions to be approximated and the length of the ground truths. The four
columns under FlashSyn-poly list data concerning the synthesis using polynomial approximations. The four columns under FlashSyn-inter list data concerning the synthesis using interpolation based approximation.

Our results show that FlashSyn can effectively synthesize flash loan attack vectors. FlashSyn-poly (resp., FlashSyn-inter) synthesizes profitable attack vectors for 16 (resp., 13) benchmarks with an average normalized profit (w.r.t. the ground truth profit) of 0.945 (resp., 0.641). For three cases (ApeRocket, ElevenFi, and AutoShark) the profits found by FlashSyn-poly are within 99% of the profits in the original attack vectors. Surprisingly in another three cases (bZx1, Warp, and OneRing) the profits found by FlashSyn are bigger than the profits in the original attack vectors. For instance, in the Warp case the profit is roughly double the ground truth profit (see extended version [13] Appendix B for Warp case study). On average, FlashSyn-poly is x2 faster than FlashSyn-inter, because we used parallelism in FlashSyn-poly which is not possible for FlashSyn-inter.

For 10 benchmarks, FlashSyn successfully discovers new profitable symbolic actions vectors that are different from the ground truths. These vectors either exploit the same vulnerability but in a different order of actions, or represent arbitrage opportunities that were not exploited by the original attackers. For the remaining 6 benchmarks, FlashSyn discovers exactly the same symbolic actions vectors as the ground truths but with different parameters. Note that FlashSyn is not able to solve Yearn and InvverseFi which are not shown in Table 3. These two benchmarks put high requirements on the precision of the approximation and small miss-approximation errors caused FlashSyn to miss finding attack vectors and accurate parameters.

To evaluate the efficacy of the pruning heuristics introduced in Section 5.1, we conduct experiments comparing the search space sizes when using FlashSyn with and without the application of some of the heuristics. Our results indicate that Heuristic 1 leads to an average reduction of 57% in the search space size. Subsequently, Heuristic 2 further reduces the remaining search space by an additional 34%, and Heuristic 3 contributes an additional reduction of 65% to the remaining search space.

To compare FlashSyn with existing static analyzers, we manually select contracts containing vulnerabilities in the benchmarks and apply the popular smart contracts static analyzer Slither [35] to them. In the experiments, we identify contracts that contain

---

**Table 3: Summary of FlashSyn Results.** AC denotes the number of action candidates. AP denotes the number of action candidates to approximate. GL and GP denote the length and the profit of the ground truth attack vector, respectively. IDP and TDP denote the initial and total number of collected of data points, respectively. Time denotes the time spent in seconds.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>AC</th>
<th>AP</th>
<th>GL</th>
<th>GP</th>
<th>IDP</th>
<th>TDP</th>
<th>Profit</th>
<th>Time</th>
<th>Precise Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>bZx1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>114</td>
<td>519</td>
<td>5844</td>
<td>2392</td>
<td>422</td>
<td>6373</td>
</tr>
<tr>
<td>Harvest_USDT</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>383448</td>
<td>8000</td>
<td>9325</td>
<td>110139</td>
<td>670</td>
<td>10289</td>
</tr>
<tr>
<td>Harvest_USDC</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>307416</td>
<td>8000</td>
<td>8912</td>
<td>59614</td>
<td>677</td>
<td>10914</td>
</tr>
<tr>
<td>Eminence</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>1674278</td>
<td>8000</td>
<td>8780</td>
<td>1507174</td>
<td>1191</td>
<td>8104</td>
</tr>
<tr>
<td>ValueDeFi</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8618002</td>
<td>12000</td>
<td>19975</td>
<td>8378194</td>
<td>4691</td>
<td>15758</td>
</tr>
<tr>
<td>CheeseBank</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>3270347</td>
<td>2679</td>
<td>2937</td>
<td>1946291</td>
<td>4391</td>
<td>2715</td>
</tr>
<tr>
<td>Warp</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1693523</td>
<td>6000</td>
<td>6000</td>
<td>2773345</td>
<td>1164</td>
<td>6000</td>
</tr>
<tr>
<td>bEarnFi</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>18077</td>
<td>4000</td>
<td>4854</td>
<td>13770</td>
<td>470</td>
<td>4652</td>
</tr>
<tr>
<td>AutoShark</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>1381</td>
<td>2753</td>
<td>2753</td>
<td>1372</td>
<td>5484</td>
<td>2753</td>
</tr>
<tr>
<td>ElevenFi</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>129741</td>
<td>4000</td>
<td>4070</td>
<td>129658</td>
<td>409</td>
<td>4526</td>
</tr>
<tr>
<td>ApeRocket</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>1345</td>
<td>4000</td>
<td>6402</td>
<td>1333</td>
<td>733</td>
<td>6235</td>
</tr>
<tr>
<td>Wdige</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>78</td>
<td>2000</td>
<td>2001</td>
<td>75</td>
<td>272</td>
<td>2080</td>
</tr>
<tr>
<td>Novo</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>24857</td>
<td>4000</td>
<td>4164</td>
<td>20210</td>
<td>702</td>
<td>4031</td>
</tr>
<tr>
<td>OneRing</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1534752</td>
<td>4000</td>
<td>4710</td>
<td>1814882</td>
<td>558</td>
<td>4218</td>
</tr>
<tr>
<td>Puppet</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>89000</td>
<td>6000</td>
<td>6301</td>
<td>89000</td>
<td>1203</td>
<td>6452</td>
</tr>
<tr>
<td>PuppetV2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>953100</td>
<td>4491</td>
<td>4836</td>
<td>747799</td>
<td>2441</td>
<td>5061</td>
</tr>
</tbody>
</table>


†: FlashSyn’s results include at least one attack vector that differs from the ground truth.

---

**Table 4: Summary of FlashSyn Results under Different Settings (see extended version [13] Appendix F for detailed results).** n+x: n denotes the settings of initial number of data points and +x denotes whether FlashSyn uses counterexample driven loops.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>FlashSyn-poly</th>
<th>FlashSyn-inter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 200+x 500 1000 1000+x 2000 2000+x</td>
<td>200 200+x 500 1000 1000+x 2000 2000+x</td>
</tr>
<tr>
<td>Avg. Time (s)</td>
<td>632 893 1120 1747 842 1397 982 1594</td>
<td>2601 3509 3180 3917 3022 3845 3200 3754</td>
</tr>
<tr>
<td>Avg. Data Points</td>
<td>584 1042 1432 2376 2795 3571 5445 6367</td>
<td>584 1338 1432 2450 2795 3656 5445 6248</td>
</tr>
<tr>
<td>Avg. Norm. Profit</td>
<td>0.793 0.829 0.846 0.922 0.762 0.786 0.717 0.945</td>
<td>0.539 0.555 0.630 0.634 0.535 0.580 0.594 0.641</td>
</tr>
<tr>
<td>Benchmarks Solved</td>
<td>15 15 15 16 15 15 15 16</td>
<td>13 13 14 14 13 13 13 13</td>
</tr>
</tbody>
</table>
the root cause of the vulnerabilities as the target contracts for Slither to analyze. Note that in practice, identifying target contracts for Slither is much harder than that for FlashFind. For FlashFind, the target contracts are simply user-interface contracts. In contrast, identifying the contract with the actual vulnerability, such as a contract invoked in a deeply nested call chain, can be tedious. Slither fails to detect vulnerabilities for all 18 benchmarks, among them Slither fails to parse 2 benchmarks (Novo and Yearn). The possible reasons include: (i) Slither’s inability to reason across multi-contract interactions, common in flash loan attacks; and (ii) its lack of context awareness, such as not detecting Uniswap [70] when used as an oracle.

**RQ2: Comparison with Precise Baseline.** The last column of Table 3 lists data when the approximation component of FlashSyn is replaced with precise mathematical summaries for actions. Note that 4 benchmarks are partially closed-source (cs), and 5 benchmarks are too complicated (cx), thus we are not able to extract mathematical precise summaries for them. For others, we list the profit generated using the manually extracted mathematical expressions in the synthesizer and optimizer.

Our results in Table 3 show that the synthesis-via-approximation approach performs well compared to precise baselines. For the 9 cases that the precise baseline failed due to either close source (cs) or complicated contract logics (cx), FlashSyn found attack vectors that generate positive profits. On average, for the 7 cases that the precise baseline succeeds, the best profit from FlashSyn is 0.97 of the profit returned by the precise baseline. In particular, for *Warp* and *PuppetV2* FlashSyn synthesizes an attack vector with a profit higher than that obtained by the precise approach. This is because the approximations used in FlashSyn are simpler than the mathematical summaries used in precise baseline. This enables the optimizer to converge faster and find better parameter values within the fixed time budget.

**RQ3: Counterexample Driven Approximation Refinement.** Table 4 summarizes the evaluation of FlashSyn under different settings. In particular, we evaluated FlashSyn with 200, 500, 1000, and 2000 initial data points threshold per action to be approximated without and with counterexample loop. The Avg. rows in Table 4 are calculated based on the 16 benchmarks excluding *Yearn* and *InverseFi*. The Avg. Norm. Profit is calculated as the average of normalized profits, i.e., profit / ground truth profit.

For FlashSyn-poly, the results in Table 4 show that only with counterexample loop we are able to solve the 16 benchmarks (Columns 500× and 2000×). Also, the maximum average of normalized profits is achieved with counterexample loop (Column 2000×) which improved from 0.717 (Column 2000) without counterexample loop to 0.945 with counterexample loop. For FlashSyn-inter, the maximum average of normalized profits is also achieved with counterexample loop (Column 2000×) which improved from 0.594 (Column 2000) without counterexample loop to 0.641 with counterexample loop.

**RQ4: Effectiveness of FlashFind.** In this experiment, we evaluate the combination of FlashFind and FlashSyn on the 14 benchmarks that FlashSyn was able to synthesize a profitable attack vector in Table 3 excluding the two fictional DVD benchmarks.13 In particular, only contract addresses are provided to FlashFind and FlashFind identifies candidate actions for FlashSyn to synthesize attack vectors with the 2000 initial data points threshold per action configuration. Table 5 presents the results.

FlashFind successfully identifies a reasonable number of action candidates for 11 out of the 14 benchmarks from given contract addresses. Among them, FlashFind identifies additional candidate actions for 7 benchmarks. For instance, FlashFind identifies 6 additional candidate actions for *OneRing*. The remaining 3 benchmarks contain action candidates whose arguments are non-primitive types, and FlashFind identifies an excessive and impractical number of choices from transaction history.14

Even with the extra candidate actions FlashSyn was able to synthesize profitable attack vectors for all 11 benchmarks. Surprisingly in 6 benchmarks, FlashSyn finds attack vectors that contains new action candidates from FlashFind that are not in the ground truth. There are two possibilities: First, the new action candidates identified by FlashFind are functionally similar to one action in ground truths (e.g, *withdraw* and *withdrawSafe* for *OneRing*). Replacing old actions with new ones gives new attack vectors. Second, the new action candidates represent another way of draining assets which the attacker failed to identify. For example, in the *Warp* benchmark, the attacker only invoked *borrowUSDC*(*USDC*, v) and *borrowSC*(*DAI*, v) to drain USDC and DAI [32, 63], however, FlashFind identifies *borrowUSDT(v)* as another candidate action, which could have been used to drain USDT as well in the same transaction.

**Impact.** One author of this paper collaborated with Quantstamp for applying FlashSyn for 3 months. We discovered two zero-day flash loan vulnerabilities in two protocols under audit.

**Threats to Validity:** The internal threat to validity mainly lies in human mistakes in the study. Note that all authors have extensive smart contract security analysis experience and software engineering expertise in general. To further reduce this threat, we manually check the balance changes for the best results given by FlashSyn in each benchmark. We verify that with the help of on-chain exchanges, these attack vectors can generate a post-balance strictly larger than initial capital (see extended version [13] Appendix E). The external threat to validity mainly lies in the subjects used in our study. The flash loan attacks we study might not be representative. We mitigate this risk by using diverse and reputable data sources, including academic papers [12, 59] and an industrial database [67].

**Limitations:** Like most synthesis tools, FlashSyn faces scalability challenges. The search space grows exponentially with the number of actions and attack vector length. A practical approach is to assess protocols on a module by module basis. By focusing only on inter-dependent actions within, we can maintain both the number of actions and the attack vector length at manageable levels, thereby mitigating the scaling issue.

---

13DVD benchmarks do not have historical transactions that FlashFind can use.
14In such cases, we believe experienced security analysts could manually identify special parameters and further reduce the number of parameter choices.
Table 5: Summary of Evaluation Results of Combining FlashSyn with FlashFind. AC is the number of action candidates. AP is the number of action candidates to approximate. GL and GP are the length and the profit of the ground truth attack vector, respectively. IDP and TDP are the initial and total number of collected data points, respectively. Time is measured in seconds.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>GL GP</th>
<th>FlashFind + FlashSyn-poly</th>
<th>FlashSyn-poly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AC AP IDP</td>
<td>TDP</td>
<td>Profit</td>
</tr>
<tr>
<td>bZx1</td>
<td>2 1194</td>
<td>3 3</td>
<td>5192</td>
</tr>
<tr>
<td>Harvest_USDT</td>
<td>4 338448</td>
<td>15 15</td>
<td>30000</td>
</tr>
<tr>
<td>Harvest_USDC</td>
<td>4 307416</td>
<td>15 15</td>
<td>30000</td>
</tr>
<tr>
<td>Eminence</td>
<td>4 1674278</td>
<td>4 4</td>
<td>8000</td>
</tr>
<tr>
<td>ValueDeFi</td>
<td>6 8618002</td>
<td>6 6</td>
<td>12000</td>
</tr>
<tr>
<td>Warp</td>
<td>6 1693523</td>
<td>8 5</td>
<td>7772</td>
</tr>
<tr>
<td>bEarnFi</td>
<td>4 18077</td>
<td>2 2</td>
<td>4000</td>
</tr>
<tr>
<td>ApeRocket</td>
<td>6 1345</td>
<td>11 5</td>
<td>10000</td>
</tr>
<tr>
<td>Wdoge</td>
<td>5 78</td>
<td>7 2</td>
<td>4000</td>
</tr>
<tr>
<td>Novo</td>
<td>4 24857</td>
<td>6 2</td>
<td>4000</td>
</tr>
<tr>
<td>OneRing</td>
<td>2 1534752</td>
<td>8 8</td>
<td>16000</td>
</tr>
</tbody>
</table>

‡: FlashSyn’s results include at least one attack vector that differs from the ground truth.
§: FlashSyn’s results include at least one attack vector that contains an action not present in the ground truth.

8 RELATED WORK

Parametric optimization: For some flash loan attack cases, researchers [12, 59] manually extracted math formulas of function candidates, manually defined related parameter constraints, and used an off-the-shelf optimizer to search for parameters which yield the best profit. However, this technique requires significant manual efforts and expert knowledge of the underlying DeFi protocols. Consequently, it becomes impractical for checking a large number of potential attack vectors. Note that our benchmark set contains significantly more flash loan attacks than prior work [12, 59], i.e., 18 versus 2 in [59] and 9 in [12].

Static Analysis: Slither [35], Securify [69], Zeus [48], Park [83] and SmartCheck [68] apply static analysis techniques to verify smart contracts. There are also several works that use symbolic execution [49] to explore the program states of a smart contract, looking for an execution path that violates a user-defined invariant, e.g., Mythril [18], Oyente [52], FairCon [51], ETHBMC [40], SmartCopy [36], and Manticore [56]. These techniques tend to operate with one contract at a time and therefore cannot handle flash loan attacks that involve multiple contracts. These techniques also may suffer from the complicated logics of the DeFi contracts, and cause path explosion. FlashSyn uses its novel synthesis-via-approximation techniques to avoid these issues.

Fuzzing: ContractFuzzer [46], sFuzz [57], ContraMaster [75], SMARTIAN [16] and ItyFuzz [66] introduce novel fuzzing techniques to discover vulnerabilities in smart contracts. However, these techniques either only work on one contract or focus on specific vulnerabilities like re-entrancy. Moreover, flash loan attack vectors can contain up to 8 actions and 7 integer parameters, which is unlikely to be found by random fuzzing.

Flash Loan Attacks: Prior works [12, 59, 76] study specific flash loan attacks and manually analyze the attack vectors. Some tools [62, 73, 80] are designed to monitor flash loan attacks after they happened. Some researchers investigate other usage of flash loans including arbitrage [74] or wash trading [43].

To the best of our knowledge, FlashSyn is the first tool that automatically synthesizes complicated flash loan attack traces. It shows successes in real-world DeFi protocols under audit.

9 CONCLUSION

We have proposed an automated synthesis framework based on numerical approximation to generate the flash loan attack vectors on DeFi protocols. Our results of FlashSyn show that the proposed framework is practical and FlashSyn can automatically synthesize attack vectors for real world attacks. Our results also highlight the effectiveness of the synthesis-via-approximation approach. The approach helps FlashSyn to overcome the challenges posed by complicated functions in DeFi protocols and our results demonstrate that using approximations of these functions is sufficient to drive the synthesis process. FlashSyn has been adopted by a top smart contract auditing company to detect flash loan vulnerabilities. The paper also points out a new promising direction, synthesis-via-approximation, for solving trace synthesis problems when facing complicated functions.

10 DATA AVAILABILITY

The benchmarks, source code and experimental data of our artifacts are publicly accessible on [14] and have been archived on [15]. Additionally, additional materials of this research are available at the extended version of this paper [13] and the website https://sites.google.com/view/FlashSyn.

ACKNOWLEDGEMENT

We thank anonymous reviewers for their insightful comments on the early version of the paper. We thank Dr. Martin Derka and auditors from Quantstamp for adopting and testing FlashSyn on projects under audit. This research is supported by NSERC Discovery Grant and MITACS Accelerate program. Sidi Mohamed Belllahi is also supported by NSERC Postdoctoral Fellowship.


