**Abstract**

Since the emergence of Ethereum, blockchain-based decentralized applications (DApps) have become increasingly popular and important. To balance the security, performance, and costs, a DApp typically consists of two layers: an on-chain layer to execute transactions and store crucial data on the blockchain and an off-chain layer to interact with users. A DApp needs to synchronize its off-chain layer with the on-chain layer proactively. Otherwise, the inconsistent data in the off-chain layer could mislead users and cause undesirable consequences, e.g., loss of transaction fees. However, transactions sent to the blockchain are not guaranteed to be executed and could even be reversed after execution due to chain reorganization. Such non-determinism in the transaction execution is unique to blockchain. DApp developers may fail to perform the on-chain-off-chain synchronization accurately due to their lack of familiarity with the complex transaction lifecycle.

In this work, we investigate the challenges of synchronizing on-chain and off-chain data in Ethereum-based DApps. We present two types of bugs that could result in inconsistencies between the on-chain and off-chain layers. To help detect such on-chain-off-chain synchronization bugs, we introduce a state transition model to guide the testing of DApps and propose two effective oracles to facilitate the automatic identification of bugs. We build the first testing framework, **DArcher**, to detect on-chain-off-chain synchronization bugs in DApps. We have evaluated **DArcher** on 11 popular real-world DApps. **DArcher** achieves high precision (99.3%), recall (87.6%), and accuracy (89.4%) in bug detection and significantly outperforms the baseline methods. It has found 15 real bugs in the 11 DApps. So far, six of the 15 bugs have been confirmed by the developers, and three have been fixed. These promising results demonstrate the usefulness of **DArcher**.

**CCS Concepts**

- Software and its engineering $\rightarrow$ Software testing and debugging.

**Keywords**

Software testing, Decentralized applications, DApps, Blockchain

**ACM Reference Format:**

Wuqi Zhang, Lili Wei, Shuqing Li, Yepang Liu, and Shing-Chi Cheung. 2021. **DArcher**: Detecting On-Chain-Off-Chain Synchronization Bugs in Decentralized Applications. In *ESEC/FSE ’21*, August 23–28, 2021, Athens, Greece. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3468264.3468546

1 INTRODUCTION

Decentralized Applications (DApps) are software applications running on a decentralized network like blockchain. Since the emergence of Ethereum [11], a blockchain platform that supports Turing-complete smart contracts, blockchain-based DApps have drawn much attention from both academia and industry. As of April 2021, on Ethereum, there are over 2.7 thousand DApps and 70 thousand active DApp users, issuing over 170 thousand transactions per day [44].

As shown in Fig. 1, DApps are typically deployed as web applications, consisting of two layers: on-chain and off-chain. The former comprises a set of smart contracts that store and update
We refer to such bugs that stem from synchronization bugs between the on-chain and off-chain states as on-chain-off-chain synchronization bugs.

On-chain-off-chain synchronization can be complicated. Changes are made to the on-chain state of a DApp by sending transactions to the blockchain. However, transaction executions are non-deterministic due to the decentralized nature of blockchain. On the one hand, transactions sent to the blockchain are not necessarily executed and could be dropped silently even after being acknowledged by the miners. On the other hand, executed transactions could be reversed as a result of chain reorganization. Such non-determinism can result in irreversible changes on the blockchain or financial losses. For example, in a DApp like Giveth, if a donation transaction is reversed on the blockchain while the DApp’s off-chain layer is unaware of the reverse and does not update the donation status, subsequent transactions to withdraw the donation will fail, causing the loss of transaction fees.

It is non-trivial to avoid on-chain-off-chain synchronization bugs even if developers have considered them in the development of DApps. Giveth developers have written more than 600 lines of code to track the lifecycle of all transactions and keep the consistency between the on-chain and the off-chain states. Augur implements a delicate rollback table to revert off-chain states when the transactions are reversed on the blockchain. In spite of the efforts that developers take, we still find on-chain-off-chain synchronization bugs in those DApps, as discussed in Section 6.5.

While several bug detection techniques have been proposed to assure the quality of DApps, they are either general tools for specifying test cases and assertions or focus only on the on-chain layer by finding defects in smart contracts. In other words, none of them can effectively pinpoint on-chain-off-chain synchronization bugs in DApps. This motivates us to investigate the on-chain-off-chain synchronization bugs and devise testing techniques to detect such bugs.

On-chain-off-chain synchronization bugs could occur in DApps on all blockchain platforms. Due to the impact of the Ethereum blockchain, our work focuses on Ethereum-based DApps. To ease presentation, we may simply refer to Ethereum-based DApps as DApps in the following.

There are two major challenges in detecting on-chain-off-chain synchronization bugs in DApps. First, we need to expose the non-determinism of the transaction executions, which is not considered by existing testing environments, so that on-chain-off-chain synchronization bugs can be better revealed. Second, the mapping between on-chain and off-chain states is usually unavailable and hard to specify. Hence, it is impractical to compare on-chain and off-chain states to check consistency directly. We need a decidable criterion to mechanically judge the consistency, thus identifying the existence of the bugs. To address the first challenge, we study the causes of the synchronization bugs and propose a state transition model for the lifecycle of transactions on the Ethereum blockchain. Guided by the model, we are able to trigger the scenarios where transactions are dropped or reversed to test DApps. To address the second challenge, we propose two test oracles based on the following key observation: the off-chain state of a DApp should stay the same if the corresponding on-chain state is unchanged, and a violation of this requirement would indicate the existence of bugs in the state synchronization process. With these oracles, we are able to define assertions only on the off-chain states to effectively reveal on-chain-off-chain synchronization bugs in DApps, without knowing the mapping between the on-chain and the off-chain states.

We implement our approach as a DApp testing framework, called DArcher, and evaluate it using 11 popular real-world DApps. These DApps vary in scale, ranging from hundreds to tens of thousands of lines of code, and have received at least 100 stars on GitHub. Our experiment results show that DArcher is effective. It is able to detect on-chain-off-chain synchronization bugs in all of the DApp subjects. We manually check warnings reported by DArcher, group the ones with the same root cause, and submit 15 issue reports on GitHub.

1To save transaction costs, blockchain typically only stores crucial data.

2Ethereum users need to pay fees for each transaction, no matter the transaction succeeds or not.

3https://github.com/Giveth/feathers-giveth/blob/45d585/src/blockchain/watcher.js

4https://github.com/AugurProject/augur/blob/85b570f/packages/augur-sdk/src/state/db/rollbackTable.ts
As a result, DApps usually store a simplified view of their on-chain state and some relevant calculation results at the off-chain layer. In this paper, we do not include such data in our definition of the off-chain state. Note that the off-chain layer may also store other data irrelevant to the on-chain state. During the lifecycle of each transaction, blockchain emits various events. DApps can monitor such events to keep their off-chain state synchronized with the corresponding on-chain state. This process is easy to implement if the transactions are executed deterministically on a centralized service or database. However, this is not the case on the blockchain, as will be explained in the next subsection.

2.2 Non-deterministic Transaction Execution

Transactions sent to the blockchain will be broadcast to miners throughout the network. Conceptually, miners on a blockchain collectively maintain a pool of transactions awaiting execution. A transaction is added to the transaction pool when it is received by a miner. A transaction is executed non-deterministically on the blockchain from two aspects. First, a transaction may not be executed after it has been sent by a DApp user. Second, an executed transaction can be reversed from the transaction history. In the following, we explain how such non-determinism arises.

2.2.1 Sent Transactions May Not Be Executed. A miner can select which transactions to execute from the pool when a new block has been mined. It is possible that a transaction is not selected and keeps staying in the pool. An old transaction in the pool can be invalidated by a new one with the same nonce, which is the index of a transaction sent from the user [10, 17]. The invalidated transactions will be dropped and never be executed on the blockchain. In addition, miners can silently drop transactions for various reasons (e.g., due to the size limit of the transaction pool).

2.2.2 Executed Transactions May Be Reversed. A newly mined block is not necessarily added to the blockchain. When multiple miners concurrently mine a new block [62], the blockchain will fork multiple chains of blocks. To resolve this problem, the blockchain will validate only the longest chain and invalidate the others. The process is known as chain reorganization [38]. If the execution of a transaction is recorded by a block in an invalidated chain, the transaction will be reversed and put back to the pool. In practice, chain reorganization can hardly affect blocks with a sufficient number of confirmations, which are the succeeding blocks on the same chain [18].

Such non-determinism in transaction execution complicates the interactions and synchronization between the on-chain and off-chain layers of a DApp. Bugs can arise if the DApp handles the non-determinism inappropriately. In Section 3, we will give a real-world example of such bugs.

3 MOTIVATING EXAMPLE

In this section, we present a code snippet adapted from Giveth [20], to illustrate the on-chain-off-chain synchronization in DApps, and an on-chain-off-chain synchronization bug [21].

Listing 1 shows a code snippet for the “withdraw donation” functionality in the front-end client of Giveth, which allows users to withdraw the donations they receive. The front-end client of Giveth uses web3.js [49], an official Ethereum JavaScript library, to send a transaction to an underlying smart contract “liquidPledging” and call its “withdraw” function, as seen in line 3. The client registers a callback for handling the “transactionHash” event, which occurs
when the transaction is added to the blockchain transaction pool (line 4). Since the transaction is to modify the on-chain state concerning the donation amount, the off-chain state is also updated accordingly (line 6). After the update, a corresponding withdraw record is created and stored in an off-chain centralized database (line 11).

The code in Listing 1 may not work correctly because the “withdraw donation” transaction may remain in the transaction pool indefinitely or be dropped without execution. The occurrence of the “transactionHash” event only signifies that a transaction has been added to the transaction pool rather than the execution of the transaction. Even if the transaction has been executed, it can still be reversed later. Therefore, it is possible that the off-chain state is updated while the on-chain state remains unchanged (i.e., when the withdraw transaction is dropped or reversed). Such state inconsistencies can lead to many undesirable consequences. For example, let us consider the following scenario. A user of Giveth sends a “withdraw donation” transaction to the smart contract, but the transaction is dropped by the blockchain. However, due to the bug, the Giveth client incorrectly shows that the donation has been withdrawn. The user may decide to donate the withdrawn cryptocurrency to another community. In such a scenario, the new donation transaction may fail, causing unnecessary loss of transaction fees and poor user experience.

It is difficult to expose on-chain-off-chain synchronization bugs using existing DApp testing tools. For example, Ganache [8] is one of the most commonly used blockchain environments for testing Ethereum-based DApps. Like testing on centralized databases, transactions sent to Ganache blockchain are executed immediately. Therefore, when testing Giveth using Ganache, the on-chain and off-chain states are always consistent. The bug mentioned above can never be detected.

To tackle this problem, we propose to model the non-determinism in transaction execution using the transaction lifecycle. Our approach simulates the non-deterministic transaction execution process and drives DApps to traverse each transaction’s lifecycle systematically. We also propose effective test oracles to detect inconsistencies between the on-chain and off-chain states automatically.

4 METHODOLOGY
This section first proposes a state transition model to capture the lifecycle of transactions on the Ethereum blockchain. We then discuss the challenges in on-chain-off-chain synchronization and identify two types of synchronization bugs that may arise in DApps. After that, we propose our framework, DÀrcher, to detect on-chain-off-chain synchronization bugs in DApps.

4.1 Transaction Lifecycle
To detect on-chain-off-chain synchronization bugs, we model the non-determinism in the lifecycle of a transaction on the Ethereum blockchain using a state transition model as shown in Fig. 2. A transaction starts its lifecycle at the Created state by constructing the required arguments at the DApp’s off-chain layer. After that, the transaction is sent to the blockchain and transits to the Pending state, meaning that the transaction is added to the transaction pool awaiting execution on the blockchain. A transaction may remain in the Pending state indefinitely as miners are free to select more profitable transactions for execution.

A Pending transaction can be dropped and transit to the Dropped state in two cases. First, users may send a duplicated transaction to override the previous transaction or offer a higher fee to increase the chance of execution [33]. Second, the transaction may be deleted silently by miners due to the capacity limit of miners’ transaction pool or malicious behaviors of miners. In the latter case, the DApp is not informed that a transaction has been dropped. It is also hard for the DApp to proactively check whether a transaction is dropped on the blockchain or not.

A Pending transaction transits to the Executed state when it is executed and included in a block. As discussed in Section 2, an executed transaction can be reversed when the blockchain is reorganized. When it happens, the transaction will transit from

Listing 1: Withdraw Donation Functionality in the DApp Giveth
the *Executed* state to the *Reversed* state and be put back to the transaction pool, awaiting execution. Similar to the *Pending* state, transactions in the *Reversed* state can also be dropped.\(^6\)

While in theory executed transactions can be reversed, in practice, a transaction whose execution has been logged by a block with a sufficient number of confirmations can be considered finalized, i.e., the transaction will transit from the *Executed* state to the *Finalized* state. The number of confirmations required varies according to the security requirements of the DApp [45]. To avoid problems caused by reversed transactions, one common practice adopted by DApp developers is to use the result of a transaction at the off-chain layer only after the transaction transits to the *Finalized* state [18].

To estimate the frequency of transaction state transitions, we collect Ethereum traffic data from Etherscan [13] and an Ethereum full node maintained by us. We find that the average number of transactions submitted to Ethereum per second is 32.5. The average number of transactions executed by Ethereum per second (TPS) is 16.4, which indicates that around half of the transactions cannot be immediately executed after submission. We also monitor the Ethereum mainnet for over 4 months and observe that chain reorganization happens every 24.43 blocks, i.e., every 11.06 minutes.\(^7\) The average number of transactions reversed per hour due to such chain reorganizations is 16.69. These statistics show that the drop and reverse rates of transactions on Ethereum are non-negligible, and it is necessary for DApps to consider the non-deterministic transaction execution to avoid on-chain-off-chain synchronization bugs.

### 4.2 On-Chain-Off-Chain Synchronization Bugs

In Section 3, we have presented an on-chain-off-chain synchronization bug in a DApp named Giveth. On-chain-off-chain synchronization bugs occur when DApps fail to maintain the consistency between their on-chain states and off-chain states. In the Giveth bug, developers do not consider the situation in which a *Pending* transaction is *Dropped* by the blockchain, and inconsistencies are thus induced. Such inconsistencies can result in erroneous off-chain states, which might mislead users. Further operations based on these erroneous states can cause unexpected changes on the blockchain or waste transaction fees.

Since DApps are a new kind of software applications, the lack of understanding of state transitions in the transaction lifecycle may make it difficult for developers to assure proper on-chain-off-chain synchronization. We observe that developers often assume that the transactions submitted by their DApp would be *Executed* and eventually *Finalized* on the blockchain while ignoring the situations where (1) a *Pending* or *Reversed* transaction is dropped silently by the blockchain, or (2) an *Executed* transaction is reversed due to chain reorganization. As a result, on-chain-off-chain-synchronization bugs often arise. In this paper, we focus on studying on-chain-off-chain synchronization bugs triggered in the two situations. We refer to them as Type-I and Type-II bugs, respectively. Fig. 3 illustrates how these two types of bugs could occur with our transaction lifecycle model. We will introduce them in detail below.

#### 4.2.1 Type-I Bugs

As Fig. 3(a) shows, Type-I bugs occur because the off-chain state is prematurely updated when the transaction is in *Pending* state, but in fact, the transaction is later dropped. The bug in Giveth as shown in Listing 1 is a Type-I bug. Giveth immediately updates its off-chain state (lines 4–13) when the donation withdrawal transaction is submitted to the blockchain. It does not further check the state of the transaction and does not adequately deal with potential state changes.

#### 4.2.2 Type-II Bugs

As Fig. 3(b) shows, Type-II bugs occur because the off-chain state is updated after the transaction transits to the *Executed* state, but the executed transaction is later reversed due to chain reorganization. Issue #8260 of Augur [31], a popular DApp for prediction markets [30], is a Type-II bug. Augur updates its off-chain state when the transactions that create markets are executed but does not revert the off-chain state when the executed transactions are reversed. Consequently, the front-end client of Augur would display markets that do not exist on the blockchain. Any further operations on these markets will result in “Market Not Found” errors.

These two types of bugs are not well studied in the literature. Existing techniques either focused on testing smart contracts [22, 25, 32, 39, 54], which are only concerned with the on-chain layer of DApps, or do not fully consider the entire lifecycle of transactions when testing DApps [8]. Take the popular DApp testing environment Ganache [8] as an example. When using Ganache, DApp developers are encouraged to configure the testing blockchain to execute transactions immediately [23], while the possible state transitions from *Pending* to *Dropped* or from *Executed* to *Reversed* are ignored [24]. As a result, Type-I and Type-II bugs can easily slip into real-world DApps. This motivates us to propose *DArcher*, a testing framework to effectively detect the two types of on-chain-off-chain synchronization bugs in DApps.

### 4.3 The *DArcher* Testing Framework

We can see from the above discussion that we need to systematically emulate transaction state transitions to trigger on-chain-off-chain synchronization bugs. Guided by our transaction lifecycle model, *DArcher* controls the execution of each transaction in its testing environment to drive the transaction to traverse possible states. However, triggering bugs alone is not sufficient. Effective testing also requires oracles to judge the existence of bugs. In the following, we present our design of test oracles in *DArcher* and then explain how to leverage the oracles to detect on-chain-off-chain synchronization bugs during testing.

#### 4.3.1 Test Oracles

It is challenging to design oracles for detecting on-chain-off-chain synchronization bugs. Although the on-chain and off-chain states in DApps need to be consistent, they are not necessarily identical. Off-chain states are often maintained as a simplified version of the corresponding on-chain states to facilitate front-end user actions. For instance, in our motivating example,
Off-chain State Update

shows an overview of Reversed
indicate the existence of Type-II bugs.

2
1

4.3.2 Lifecycle Emulation & Assertion Checking. As discussed earlier, in existing blockchain testing environments such as Ganache [8], transactions are directly Executed and Finalized once submitted to the blockchain. These transactions are never dropped or reversed, and thus Type-I and Type-II bugs cannot be triggered. To address the limitation, DArcher implements a blockchain environment that can control the state of transactions. Instead of executing a transaction immediately after it is submitted to the blockchain, DArcher drives the transaction to traverse lifecycle states in the following order: Created → Pending → Executed → Reversed → Executed → Finalized. Such a traversal allows our proposed oracles to be evaluated for each transaction to detect the two types of on-chain-off-chain synchronization bugs. Specifically, when there is a state transition, DArcher fetches and stores the off-chain state of the DApp under testing. After the state traversal terminates, the fetched off-chain state is checked against the Assertions 1 and 2. If there is any assertion violation, DArcher will report a bug. More details of bug detection will be further introduced in Section 5.

5 IMPLEMENTATION

Fig. 4 shows an overview of DArcher. Given a DApp, the front-end developer fires UI events to exercise the DApp to explore the functionalities that involve sending transactions. Once a transaction is sent, DArcher leverages a controlled blockchain to execute it and traverse its lifecycle states in the order mentioned in Section 4.3.2. During the state traversal process, DArcher keeps fetching the off-chain state of the DApp, whenever there is a state change. The state consistency analyzer then checks such collected off-chain states to detect bugs.

We have open-sourced DArcher on GitHub. In the following, we present more details on how DArcher is implemented.

https://github.com/Troublor/darcher
5.1 The Front-End Explorer
As mentioned earlier, the front-end explorer fires UI events to exercise DApps. In practice, developers can use any applicable tool or manually write test cases for purpose. In the current implementation of DA Archer, we choose to integrate a popular web testing tool, Crawljax [34], to generate GUI events to test DApps, which are often web-based applications. Crawljax infers a state-flow graph when testing a web application. GUI events are fired at the states that can transit to unvisited states. The state-flow graph is updated whenever new states are discovered during testing. The exploration will stop when all of the states in the graph have been visited. Such model-based testing can help exercise DApps to interact with the blockchain.

5.2 The Controlled Blockchain
We implement a controlled blockchain in DA Archer based on Geth [12], a popular Ethereum client. When the DApp under test submits a transaction to our controlled blockchain, DA Archer will drive the transaction to traverse its lifecycle states according to the order mentioned in Section 4.3.2.

5.3 The State Consistency Analyzer
As mentioned in the overview, during testing, DA Archer keeps collecting the off-chain state data for bug detection. Different DApps may maintain the off-chain states in different ways, e.g., using various databases such as MySQL and MongoDB, or browser storage such as IndexedDB and localStorage. It is hard to design a tool to automatically identify runtime values representing the off-chain states for all DApps. As a workaround, we build our off-chain state fetcher for different data storage, including databases, browser storage, and HTML elements. Users of our tool only need to configure a few rules to specify the variables and fields that constitute the off-chain state of their DApps (e.g., specifying column names in databases or regular expressions to include or exclude table columns).

To minimize the runtime overhead, DA Archer does not instrument DApps. Since it is hard to determine when a DApp finishes updating its off-chain state at runtime, DA Archer would wait for a period of time when a transaction’s lifecycle state changes before fetching the off-chain state. The waiting time is configurable in DA Archer.

After the state traversal process completes, DA Archer checks the fetched off-chain states to identify the two types of on-chain-off-chain synchronization bugs according to the two oracles proposed in Section 4.3.1.

6 Evaluation
We evaluate DA Archer on real-world web-based DApps. Specifically, we investigate the following research questions:

- **RQ1 (Bug Detection Capability)**: Can DA Archer effectively detect on-chain-off-chain synchronization bugs in real DApps?
- **RQ2 (Efficacy of Our Oracles)**: How effective are our proposed oracles? Can on-chain-off-chain synchronization bugs be detected using existing oracles?
- **RQ3 (Usefulness)**: Can DA Archer detect on-chain-off-chain synchronization bugs that are useful to developers?

6.1 Subjects
To collect subjects for evaluation, we search for Ethereum DApps on GitHub using the keyword `ethereum` with constraints `stars:>=100`, `language:javascript` or `language:typescript`. The first constraint is to ensure the popularity of the selected subjects. The second constraint is to help find web-based DApps. The search returns 254 projects. We manually check each project to exclude those that are not web applications or those that are libraries, blockchain clients, block explorers, transaction trackers, frameworks, operating systems, and so on. Projects that are tagged as deprecated or archived are also excluded because they are not actively maintained. Furthermore, we exclude those that are claimed to be examples, tutorials, or starters, as we are interested in real DApps rather than toy examples. After this filtering process, 21 DApps remain. Finally, DA Archer is a dynamic testing framework, which requires executing transactions of smart contracts on a local controlled blockchain. Although a number of DApps provide the source code of smart contracts and web applications, many of them lack instructions to deploy their smart contracts on a local blockchain. They expect users to deploy the open-source web applications on Ethereum public testnets or mainnet, where contracts have been well-deployed. Therefore, we exclude those projects that we fail to deploy on our local blockchain by following the provided instructions. After the above filtering, we collect a total of 11 popular real-world DApps for experiments. This is in line with the finding in an empirical study made by Wu et al. [58] that few DApps are fully open-source. Table 1 provides the information of these subjects. The selected subjects differ in scale. The smallest DApp has less than 1,000 lines of code, while the largest DApp has more than 30,000 lines of code. The purposes of the subjects are also quite diversified.

6.2 Experiment Design
In this subsection, we explain how we set up our experiments, derive off-chain states, and establish the ground truth to validate...
bug detection results. We also introduce the baselines, against which DArcher will be compared.

**Experiment Setup.** We deploy the subjects locally on our controlled blockchain. We set the DArcher’s waiting time for DApps to update their off-chain states to be 15 seconds, which is aligned with the average block interval time on Ethereum mainnet [13]. We run DArcher on each DApp for one hour. We observe that this is sufficient for DArcher to reach the saturation of test coverage during experiments. We repeat the testing process ten times for each DApp to mitigate the randomness of Crawjax.

**Off-Chain State Derivation.** One challenge in the experiments is to identify the data fields that compose off-chain states. As we have discussed in Section 5.3, different DApps may maintain off-chain states in different ways. Essentially, DArcher requires developers to manually specify which data fields or variables in the DApp constitute the off-chain state. However, such specification is unavailable for our evaluation. As a workaround, we manually explore each DApp and derive the off-chain state by including those data fields that are updated when transactions are directly executed (i.e., going through the states *Created* → *Pending* → *Executed* → *Finalized*). The intuition behind is that we assume developers have tested their DApp to ensure that the off-chain state is properly updated when transactions undergo such “normal” executions. Note that DArcher does not integrate this off-chain state derivation mechanism in that the aforementioned assumption does not necessarily hold for all DApps.

**Result Validation.** To evaluate the precision and recall of DArcher for each DApp, we manually reproduce all transactions generated during testing and make them go through the lifecycle to identify potential inconsistencies between the on-chain and off-chain states. This process is independently performed by two authors. The results are cross-checked for consistency. It is worth mentioning that the recall metric we used here is for evaluating DArcher’s capability of catching Type-I and Type-II bugs once they arise during the processing of transactions. We do not aim to evaluate how many of all possible on-chain-off-chain synchronization bugs can be detected by DArcher since it is hard to obtain the ground truth. Not only that, the Crawjax front-end explorer may not be able to trigger all possible transactions during testing.

**Baselines.** Since there is no prior work for detecting on-chain-off-chain synchronization bugs in DApps, we construct two baseline methods by replacing oracles used by DArcher with the ones used for detecting smart contract vulnerabilities [25] and web application faults [4]. Specifically, Baseline-I would report bugs if the execution of a transaction violates the assertions defined by ContractFuzzer [25]. Baseline-II would report bugs if runtime errors occur in the JavaScript console during the testing process [4]. Except for the differences in oracles, the two baseline methods use exactly the same tests for each DApp and traverse the lifecycle of each transaction in the same way as DArcher. We do not compare DArcher with existing blockchain testing environments such as Ganache [8], since none of them is able to emulate the lifecycle of transactions, and thus neither Type-I nor Type-II bugs can be triggered by them.

### 6.3 RQ1: Bug Detection Capability

Table 2 presents the results of DArcher and baselines, including the overall test coverage, precision, recall, and accuracy. Since DArcher aims to test the correctness of synchronization between the on-chain and off-chain layers in DApps during the transaction execution process, we measure the test coverage in terms of how well transaction submission API call sites are covered during testing. Such API call sites are essentially the “starting point” of on-chain-off-chain synchronization. Their coverage could indicate the diversity of the tests generated by the Front-End Explorer. Note that since some functionalities relying on Ethereum public blockchain services (e.g., Uniswap [52]) are unavailable on the controlled blockchain of DArcher, API call sites for such functionalities are not considered in our coverage measurement. In the following, we will discuss the experiment results in detail.

**Coverage.** DArcher triggers a total of 3,134 transactions and covers 70% of the transaction submission API call sites when testing the 11 DApps. It achieves an overall accuracy of 89.4% in deciding whether a transaction handling process contains bugs or not.

**False Positives.** Altogether, there are 385 transactions violating Assertion 1 and 1,862 transactions violating Assertion 2. There is no transaction violating both assertions. Among these transactions, there are 16 exhibiting Type-I bugs, which are found to be false positives (FPs). There are no false positives for those detected to exhibit Type-II bugs. We find that all of the 16 FPs are related to the DApp Giveth. They arise from the unexpected delays in updating the off-chain states in Giveth’s cache server. For instance, when a transaction reaches the *Pending* state, Giveth makes a partial update to the off-chain state, with a flag indicating that the transaction is awaiting execution. The flag should be cleared when the transaction is *Finalized*. However, it is not cleared within 15 seconds, so that DArcher fetches the same off-chain state as the one fetched when the transaction is *Pending*. In this case, DArcher reports a violation of Assertion 1, but in fact, it is a false positive. If DArcher waits for a longer period before fetching the off-chain state, the violation will not be reported. Despite the FPs, DArcher still achieves an overall precision of 99.3% in detecting on-chain-off-chain synchronization bugs in the collected DApps.

**False Negatives.** There are 52 and 264 transactions exhibiting Type-I and Type-II bugs, respectively, which are missed by DArcher.

---

11We are able to measure the coverage of transaction submission API call sites because DApps use designated APIs [6, 47, 48] provided by the official libraries (e.g., web3.js [49]) of Ethereum to interact with the blockchain.

### Table 1: The DApps Used in Our Experiments

| DApp                  | Stars LOC (JS) Commits Purpose                  |
|-----------------------|------------------------------------------------|
| AgroChain [1]         | 106 291 40 Agricultural Supply Chain           |
| Augur [30]            | 264 68,972 40,971 Prediction Markets           |
| DemocracyEarth [16]   | 1,505 29,149 3,394 Governance for DAOs         |
| ETH Hot Wallet [46]   | 197 7,295 284 Wallet                           |
| Ethereum Voting Dapp  | 348 238 31 Voting                              |
| Giveth [20]           | 257 30,708 3,131 Charitable Donation           |
| Heiswap [50]          | 100 1,687 87 Anonymous Transfer                |
| MetaMask [15]         | 4,091 81,798 11,072 Wallet                    |
| Multisender [46]      | 165 1,421 33 One-to-Many Transfer              |
| PublicVotes [42]      | 149 628 34 Voting                              |
| TodoList Dapp [2]     | 127 1,076 28 Todo List                         |

---
After investigating the corresponding transactions, we find two major reasons for the false negatives (FNs). First, DArcher may fetch the off-chain state before it is inappropriately updated, in the same way as mentioned in the previous Giveth example. Second, transactions may depend on each other. An on-chain-off-chain synchronization bug may occur if a transaction \( v \) is executed based on the interim result of another transaction \( u \), while \( u \) is reversed and dropped. FN occurs if the DApp does not update the off-chain state for transaction \( v \) (neither Assertion 1 or 2 will be violated if the off-chain state is unchanged). Despite such cases, the recall of DArcher is still quite high and reaches 87.6% in our experiments.

Answer to RQ1: DArcher can effectively detect on-chain-off-chain synchronization bugs in DApps with high precision (99.3%), recall (87.6%), and accuracy (89.4%).

6.4 RQ2: Efficacy of Our Oracles

We answer RQ2 by comparing DArcher with the two baseline methods that employ existing oracles. The results of baselines are also presented in Table 2.

As we can see from the table, Baseline-I only generates warnings for seven transactions. After inspecting the seven transactions whose execution violates vulnerability assertions, we find that the underlying smart contract contains the exception disorder vulnerability according to the bug definition by Jiang et al. [25]. However, these seven warnings are all FPs in terms of on-chain-off-chain synchronization bugs. Smart contract vulnerability oracles cannot detect any on-chain-off-chain synchronization bugs in the experiment. This is because the detection of on-chain-off-chain synchronization bugs requires the examination of both on-chain and off-chain layers of a DApp, whereas the oracles for smart contract vulnerabilities examine only the on-chain layer.

Baseline-II is able to reveal some on-chain-off-chain synchronization bugs. For example, a bug [31] in Augur results in a runtime error with message "Uncaught (in promise) Error: execution reverted" when a transaction is reversed. However, the runtime error oracle is not effective compared to our proposed oracles. Our experiments show that the overall precision, recall, and accuracy of Baseline-II are only 56.9%, 10.1% and 20.7%, respectively, which are significantly worse than DArcher. In addition, the runtime error messages generated may not provide useful information about the root causes of on-chain-off-chain synchronization bugs. For instance, the message "Error: PollingBlockTracker - encountered error fetching block:" generated in the testing of MetaMask gives little hint of the occurrence of a synchronization bug.

Answer to RQ2: Our proposed oracles significantly outperform the existing ones in terms of detecting on-chain-off-chain synchronization bugs.

6.5 RQ3: Usefulness

We answer RQ3 by reporting bugs detected by DArcher to the developers and communicating with them. Although DArcher reports warnings for thousands of transactions in the experiments, lots of them are repeated explorations of the same functionalities in DApps. To avoid overwhelming developers, we group the warnings with the same root cause into a single issue to report to the developers. Table 3 lists the IDs of the GitHub issues, in which we report on-chain-off-chain synchronization bugs to the DApp developers.

In total, we have reported 15 bugs, of which six have been confirmed by developers, and three have been fixed. Developers provide positive feedback on our reported bugs. For example, developers of Giveth respond "Syncing two backend (cache server and blockchain) is a delicate and complex job, and we hope it is solved soon", and the bugs in Giveth were fixed one and a half month after we reported them.

Answer to RQ3: Among 15 bugs reported to developers, six have been confirmed and three have been fixed. Responses from developers show that DArcher is useful in detecting on-chain-off-chain synchronization bugs.

7 DISCUSSIONS

7.1 Synchronization Strategies

In the evaluation, we observe that some DApps have considered the possibility that pending transactions can be silently dropped and that executed transactions can be reversed. However, on-chain-off-chain synchronization bugs are still detected in these DApps. This indicates that on-chain-off-chain synchronization is non-trivial and error-prone, highlighting the detection capability of DArcher.

To better understand how developers of DApps synthesize on-chain and off-chain states, we further investigate the synchronization strategies adopted in the 11 DApps used in our evaluation. We observe three common strategies as follows.
### Table 2: Experiment Results of DArcher and Baselines

| DApp                  | API Call Site Coverage | Total Txs | DArcher | DArcher | Contract Vulnerability Oracle | Baseline-II |
|-----------------------|------------------------|-----------|---------|---------|-------------------------------|-------------|
|                       | Total                   |           | TP      | FP      | FN    | Pre.  | Rec.  | Acc. | TP      | FP     | FN    | Pre.  | Rec.  | Acc. |
| AgroChain             | 562.0%                  | 341      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| Augur                 | 66.7%                   | 144      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| DemocracyEarth        | 100.0%                  | 79       | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| ETH Hot Wallet        | 100.0%                  | 140      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| Ethereum Voting Dapp  | 100.0%                  | 376      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| Giveth               | 63.5%                   | 353      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| Heiswap               | 100.0%                  | 323      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| MetaMask              | 100.0%                  | 225      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| Multisender           | 100.0%                  | 334      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| PublicVotes           | 100.0%                  | 209      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |
| ToodList Dapp         | 100.0%                  | 435      | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% | 0.00%   | 0.00%   | 100.0% | 0.00% | 0.00% | 100.0% |

**Note:**
- **Aggressive Updating:** The aggressive updating strategy is another choice for DApp developers. This strategy is intended to keep the off-chain state closely synchronized with the on-chain state, which means the DApp updates its off-chain state when transactions are executed, and reverts the off-chain state when transactions are reversed. For instance, Augur adopts this strategy in its implementation. When a transaction is sent to the blockchain, that is, in the *Pending* state, the off-chain state is not updated. The update only takes place when the transaction is executed. If the transaction is reversed due to blockchain reorganization, Augur will also revert its off-chain state accordingly. This strategy offers better user experience than the passive waiting strategy. Users can see the updates of the off-chain state immediately when their transactions are executed or reversed. However, it is more error-prone to revert the off-chain state, which might involve many data fields, when transactions are reversed on the blockchain.

### Table 3: Real Bugs Detected by DArcher

| DApp                  | Issues | DApp | Issues |
|-----------------------|--------|------|--------|
| AgroChain             | #7, #8 | Heiswap | #29 |
| Augur                 | #8260\(^f\) | MetaMask | #10120\(^f\) |
| DemocracyEarth        | #561   | Multisender | #34\(^e\) |
| ETH Hot Wallet        | #41    | PublicVotes | #11, #12 |
| Ethereum Voting Dapp  | #28    | ToodList Dapp | #5\(^f\) |
| Giveth               | #1103\(^f\), #1605\(^f\), #1792\(^f\) | | |

\(^f\)The reported bugs have been confirmed and fixed.

\(^e\)The reported bugs have been confirmed.

\(^\)The developers have asked for Pull Requests.

Other issues have not received responses from developers.

**Periodic Polling.** The most straightforward way to synchronize on-chain and off-chain states is to poll the on-chain state periodically. For instance, DemocracyEarth [16] registers a daemon task that periodically checks the on-chain state and updates the off-chain state accordingly. This strategy effectively keeps the off-chain state consistent with the on-chain state during the lifecycle of transactions. Nevertheless, periodic polling is inefficient if the DApp is complicated. If redundant synchronization work is performed repeatedly, a lot of communication and computation overheads will result.

**Passive Waiting.** The passive waiting strategy copes with the non-determinism of transaction execution by updating the off-chain state only when a transaction reaches the *Finalized* state in its lifecycle. That is to say, as long as the non-determinism still exists, i.e., transactions are not yet finalized, the DApp does not update its off-chain state. For instance, Giveth [20] adopts this strategy by counting the number of confirmations for each transaction after it is executed. Only when a transaction has enough confirmations will Giveth update its off-chain state in the centralized cache server. This strategy could save a lot of communication and computation overheads compared to the periodic polling strategy since the DApp only needs to track the transactions that are not yet finalized. However, it could induce an inevitable delay in the DApp, influencing user experiences because after the transaction is executed, users must wait until there are enough confirmations.

**7.2 Limitations and Future Work**

Our work is subject to two limitations. First, precise identification of the off-chain states is important to the test effectiveness of DArcher. In our experiments, we assume that the off-chain states are appropriately updated when the transactions are completed in a straightforward manner. This may not always hold. Second, DArcher assumes that each update of an off-chain state, if it indeed happens, would be completed within a fixed time period, which is to be manually specified. If the period is set to be too large, the time efficiency of DArcher is compromised. If the period is set to be too small, DArcher may miss the detection of some on-chain-off-chain synchronization bugs. Furthermore, the period can vary across DApps. A possible solution to these two limitations is to analyze the source code of DApps to determine what comprises the off-chain states and when they will be updated so that the efficiency and effectiveness of each transaction’s analysis could be improved. However, the dynamic and reflective nature of JavaScript [43] imposes other challenges to perform a sound and complete analysis of DApps. As such, we leave these two limitations to be addressed in our future work.

Furthermore, as discussed in Section 6.3, it could be the case that the consequence of on-chain-off-chain synchronization bugs...
is not reflected in DApps’ off-chain states, for instance, sending a transaction dependent on the interim result of another transaction, which gets reversed or dropped. Future work can be made to detect the on-chain-off-chain synchronization bugs exhibited in such scenarios.

### 7.3 Threats to Validity
The limited number of DApp subjects poses an external threat to the validity of our evaluation results. We mitigate this threat by selecting popular real-world DApps of different sizes and purposes from GitHub to improve their representativeness. More subjects are needed to address this threat in future research fully. Leveraging Crawljax to exercise DApps induces another external threat in that Crawljax is a randomized tool and may not trigger all possible transactions. We mitigate this threat by repeating the experiments ten times in order to increase the diversity of explored functionalities.

Threats to internal validity may arise from the way we interpret the experiment results. The specification of off-chain states for each DApp poses a threat, which we mitigate by mechanically deriving off-chain states with the assumption discussed in Section 6.2. During the reproduction of transactions, we confirm that our assumption holds for all DApp subjects. We also report our detected bugs to the DApp developers and ask for their feedback to confirm the effectiveness and usefulness of DArcher.

## 8 RELATED WORK

This section briefly reviews the existing work related to the problem that DArcher aims to address.

### 8.1 DApp Development and Testing
In 2017, Porru et al. [41] introduced the concept of blockchain-oriented software engineering and pointed out the challenges and research directions on the development and testing. Since then, DApps, as a kind of blockchain-oriented software, started to attract attention from software engineering researchers. Wessling et al. [55, 56] discussed the design choices of the architecture of a software that involves blockchain, showing the benefits and drawbacks of DApps. Wu et al. [57, 58] conducted empirical studies on Ethereum-based DApps to show the popularity, growth, development practice, cost, and the open-source status quo. They pointed out the research direction in the synchronization between the on-chain and off-chain layers of a DApp, but no further study has been conducted. Wu et al. [59] proposed a framework, Kaya, for testing DApps. However, their framework only provides tools to facilitate the manual creation and execution of test cases for DApp. Unlike our work, Kaya does not target any bugs specific to DApps and does not propose oracles to help automatically reveal bugs in DApps.

### 8.2 Smart Contract Testing
Lots of studies have been conducted over the past several years to analyze and test smart contracts. Various approaches have been proposed to detect vulnerabilities with symbolic execution [32, 35], fuzzing [25, 26, 28, 39, 60], static analysis [5, 14, 22, 36, 51, 54, 61], mutation testing [7], or machine learning [29] techniques. Multiple empirical studies have also been conducted to review and verify the effectiveness and efficiency of smart contract vulnerability analysis tools [9, 19]. However, these studies only focus on testing the on-chain layer of a DApp, that is, smart contracts. As shown in our evaluation, in which we adopt contract vulnerability oracles from ContraFuzzer [25], testing smart contracts only and neglecting the testing of the interaction between on-chain and off-chain layers cannot ensure the correctness of a DApp. Our work, instead, takes the interaction between the on-chain and off-chain layers into consideration and can detect bugs during the synchronization of the two layers.

### 8.3 Test Case Generation for Web Applications
Since DArcher targets web-based DApps, test case generation in web applications is relevant to our work. Mesbah et al. [34] proposed Crawljax, which can derive a state flow graph for web applications and generate tests to traverse the graph. Since Crawljax is well-maintained and capable of automatically generating test cases with good coverage to explore web applications, we integrate it into DArcher to trigger the interaction with smart contracts for testing DApps. Other test generation tools, such as SubWeb [3] and DIG [4], are usually built based on Crawljax, with the aim to further increase the coverage of generated tests. Since these tools require web applications to have page navigational models, which are not available in our DApp subjects, we did not integrate them into our framework. Nevertheless, the DApp Front-End Explorer component of DArcher is loosely coupled with other components. It is convenient to migrate to other web testing tools to explore functionalities of DApps and detect on-chain-off-chain synchronization bugs based on our proposed oracles.

## 9 CONCLUSION
In this paper, we study the development challenges of DApps caused by the non-deterministic transaction execution on the blockchain and the on-chain-off-chain synchronization bugs thus induced. We propose DArcher, an automated testing framework to detect two common types of such bugs in DApps. Our experiments on 11 real-world DApps show that DArcher is effective in detecting on-chain-off-chain synchronization bugs and significantly outperforms the baseline methods in terms of precision, recall, and accuracy. Feedbacks from real-world DApp developers also confirm the effectiveness and usefulness of DArcher.

## ACKNOWLEDGMENTS
We sincerely thank Yuqing Xiao for the early-stage discussions towards this idea; Jialun Cao and Lu Liu for the discussions about the feasibility of our methodology; Yongqiang Tian and Jiarrong Wu for their efforts of proof-reading; and anonymous reviewers for their constructive comments. This work was supported by the National Natural Science Foundation of China (Grant No. 61933021), Hong Kong RGC/GRF (Grant No. 16207120) and Guangdong Provincial Key Laboratory (Grant No. 2020B12101001), China. Lili Wei was supported by the Postdoctoral Fellowship Scheme of the Hong Kong Research Grant Council.

## REFERENCES
[1] Kerala Blockchain Academy. 2021. AgroChain: Agricultural Supply Chain Dapp with Micro-Finance. https://github.com/Kerala-Blockchain-Academy/AgroChain.
[55] Florian Wessling, Christopher Ehmke, Marc Hesenius, and Volker Gruhn. 2018. How Much Blockchain Do You Need? Towards a Concept for Building Hybrid DApp Architectures. In Proceedings of the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB ’18). Association for Computing Machinery, New York, NY, USA, 44–47. https://doi.org/10.1145/3194113.3194121

[56] Florian Wessling, Christopher Ehmke, Ole Meyer, and Volker Gruhn. 2019. Towards Blockchain Tactics: Building Hybrid Decentralized Software Architectures. In 2019 IEEE International Conference on Software Architecture Companion (ICSA-C ’19). 254–237. https://doi.org/10.1109/ICSA-C.2019.00048

[57] Kaidong Wu. 2019. An Empirical Study of Blockchain-Based Decentralized Applications. arXiv:1902.04969 [cs] (Feb. 2019). arXiv:1902.04969 [cs]

[58] Kaidong Wu, Yun Ma, Gang Huang, and Xuanzhe Liu. 2019. A First Look at Blockchain-Based Decentralized Applications. arXiv:1909.00939 [cs] (Sept. 2019). arXiv:1909.00939 [cs]

[59] Zhenhao Wu, Jiashuo Zhang, Jianbo Gao, Yue Li, Qingshan Li, Zhi Guan, and Zhong Chen. 2020. Kaya: A Testing Framework for Blockchain-Based Decentralized Applications. In 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME ’20). 826–829. https://doi.org/10.1109/ICSME46990.2020.00103

[60] Valentin Wüstholz and Maria Christakis. 2020. Targeted Greybox Fuzzing with Static Lookahead Analysis. In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering (ICSE ’20). ACM, Seoul South Korea, 789–800. https://doi.org/10.1145/3377811.3380388

[61] Yinxing Xue, Mingliang Ma, Yun Lin, Yulei Sui, Jiaming Ye, and Tianyong Peng. 2020. Cross-Contract Static Analysis for Detecting Practical Reentrancy Vulnerabilities in Smart Contracts. In Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering (ASE ’20). Association for Computing Machinery, New York, NY, USA, 1029–1040. https://doi.org/10.1145/3324884.3416553

[62] Shijie Zhang and Jong-Hyouk Lee. 2020. Analysis of the Main Consensus Protocols of Blockchain. ICT Express 6, 2 (June 2020), 93–97. https://doi.org/10.1016/j.ictexp.2019.08.001