Detecting and Quantifying Wash Trading on Decentralized Cryptocurrency Exchanges

Friedehelm Victor
Technische Universität Berlin
Berlin, Germany
friedelhem.victor@tu-berlin.de

Andrea Marie Weintraud
Technische Universität Berlin
Berlin, Germany
weintraud@campus.tu-berlin.de

ABSTRACT
Cryptosets such as cryptocurrencies and tokens are increasingly traded on decentralized exchanges. The advantage for users is that the funds are not in custody of a centralized external entity. However, these exchanges are prone to manipulative behavior. In this paper, we illustrate how wash trading activity can be identified on two of the first popular limit order book-based decentralized exchanges on the Ethereum blockchain, IDEX and EtherDelta. We identify a lower bound of accounts and trading structures that meet the legal definitions of wash trading, discovering that they are responsible for a wash trading volume in equivalent of 159 million U.S. Dollars. While self-trades and two-account structures are predominant, complex forms also occur. We quantify these activities, finding that on both exchanges, more than 30% of all traded tokens have been subject to wash trading activity. On EtherDelta, 10% of the tokens have almost exclusively been wash traded. All data is made available for future research. Our findings underpin the need for countermeasures that are applicable in decentralized systems.

CCS CONCEPTS
• Applied computing → Digital cash; Electronic funds transfer; Computer forensics.

KEYWORDS
wash trading, self-trades, measurement, crime, blockchain

1 INTRODUCTION
Despite several cycles of boom and bust, cryptocurrencies and other virtual assets in the form of blockchain-based tokens appear to remain popular. Founders of blockchain startups frequently choose to issue a token, e.g. to represent shares or voting rights in the startup. blockchain-based token systems became popular in 2017, and have evolved in recent years. In fact, some new developments even see heightened attention: A new group of services under the umbrella term Decentralized Finance (DeFi) are establishing themselves as a key use case for distributed ledger technologies (DLTs) in general. The aim is to design financial products, that are partly based on services from the traditional financial world yet completely new in other areas. The appeal lies in offering these financial products in a decentralized way, excluding middlemen as far as possible.

While the field now encompasses a multitude of services such as lending protocols, derivatives and insurances, so-called decentralized exchanges (DEX) were among the early drivers of the ecosystem. They allow for the exchange of virtual assets without having to rely on externally-controlled services such as centralized exchanges, many of which have been the victims of theft in recent years [22]. As the number of different DeFi services increases, so does the complexity and vulnerability of these systems, as demonstrated by the recent example of large so-called flash loan-based trades. Within a single transaction, a large cryptocurrency loan is taken and repaid while executing an attack vector against multiple services [24]. It is therefore necessary to gain a better understanding of how services in the DeFi ecosystem are being used. This can help to uncover manipulative behavior, so that these systems can be improved in the future.

In this work, we therefore focus on two of the first and most popular decentralized exchanges on the Ethereum platform, where we study the trading activity between user accounts. In particular, we identify a lower bound of suspicious trading behavior that closely follows the definition of wash trading, a type of market manipulation that is well known in the context of traditional financial instruments, and illegal in most countries. To perform wash trading, several users can collude and trade only amongst themselves. Thereby, they give the impression that they are buying and selling, but in reality they are not changing their own positions or taking any real market risk. These activities inevitably lead to increased (fake) trading volume, a metric that is observed among many traders and may influence a trader’s sentiment about a given virtual asset. The same effect can also be achieved with a single user operating multiple accounts. Since account creation on the Ethereum blockchain is virtually cost-free, and does not require providing identity information, this scenario is much more likely.

Although wash trading in cryptocurrencies has been studied on centralized exchanges [13], where trading happens off-chain, previous analysis is limited to exchange-reported trades, for which account level information is unavailable. In addition, it is very likely that these reported trades have never actually taken place, and the exchange operators are the main suspects. This is in contrast to trades on decentralized exchanges, where each trade carries a cost and is stored on the ledger along with the involved accounts.
To the best of our knowledge, no systematic analysis of wash trading activity exists for decentralized exchanges.

Contributions in this paper:
- We present the first systematic analysis of wash trading behavior on the decentralized exchanges IDEX and Etherdelta, demonstrating the issue and the need for countermeasures.
- We empirically identify wash trade accounts and trade structures and discover that the majority of structures consist of one or two accounts, while complex ones also exist.
- We quantify a lower bound of the financial extent of the activities, finding that on both exchanges, a total of 159 million U.S. Dollars has been wash traded, that at least 30% of all tokens on both exchanges experienced wash trading activity, and that on EtherDelta, 10% of the traded tokens have almost exclusively been wash traded.

2 BACKGROUND
With the inception of Bitcoin [23], digital currencies have gained in popularity. Users can manage their currency within an account, also called a wallet, which consists of a public and private key pair. To send a transaction, users have to specify a target account address, which is derived from a public key, and sign it with their own private key. Today, users are not limited to transferring the native cryptocurrency of a particular blockchain. The emergence of blockchains featuring smart contracts has paved the way for the issuance of a wide variety of assets that are typically represented in the form of tokens. Ethereum [31] is a prime example for this, where thousands of such tokens have been issued. It is also a platform that has seen widespread use of decentralized exchanges, where tokens can be traded against Ether (ETH) and other tokens, thus offering an alternative to external, centralized exchanges.

2.1 Centralized and Decentralized Exchanges
A central exchange (CEX), sometimes also called a custodial exchange, keeps a user’s assets in their collective exchange wallets. These are high-value targets of attacks, as evidenced by numerous cryptocurrency exchange hacks. In order to trade on a CEX, users send their assets to an exchange-owned deposit wallet specifically created for a particular user. This wallet then forwards the assets to their main wallet and registers the received funds to be used for trading. All trading happens outside of the blockchain, until a user wants to withdraw their assets, at which point the exchange sends them back to the user wallet.

In contrast, a decentralized exchange (DEX) is typically implemented as a smart contract, which can allow for the non-custodial trading of cryptoassets. Users send their assets to the smart contract, can interact with it for trading, and withdraw again. In the past years, multiple types of DEX have been proposed, some of which are actively used.

2.2 Types of Decentralized Exchanges
The two main DEX variants are based on limit order books (LOB) or automated market makers (AMM). The following is a brief overview. For more details, we refer the reader to Lin et al. [21]

2.2.1 Limit Order Book-based DEX. In a LOB-based DEX, users trade with each other via an order book. As makers, they submit orders to the order book as an offer to buy or sell an asset at a certain price and/or volume. When filling such an order, either explicitly or automatically, the users act as takers. The order books can be managed on-chain or off-chain, but the actual settlement of the trades typically happens on-chain. Two popular examples of such LOB-based DEX are EtherDelta and IDEX, that both operate on the Ethereum blockchain. They both manage their order books off-chain, which requires an external service that users interact with. Nevertheless, they can withdraw their assets at any time by interacting with the smart contract directly. In the case of EtherDelta, users send trade instructions to the smart contract itself, whereas IDEX uses a separate account to trigger trade instructions at the smart contract. They both charge trading fees, which amount to 0.3% of the traded amounts, where IDEX splits the fees between maker and taker, and EtherDelta charges only the taker. In addition, the transaction fees need to be covered by the traders, effectively increasing the total trading fees to more than 0.3%.

2.2.2 Automated Market Maker DEX. AMM DEX such as Uniswap and Kyber differ significantly from an LOB DEX. Users do not trade peer-to-peer but against a liquidity pool or reserve. These are implemented using smart contracts. On Uniswap, each reserve or contract holds funds of a token and ETH. Uniswap features a pricing mechanism called a "constant product market maker" formula. Prices are determined by the share of token and ETH funds in a reserve in relation to an invariant of total liquidity, which remains stable during trading. Liquidity providers fill the pools and must always contribute equal amounts of a token and ETH [2].

Although AMM DEX have recently gained in popularity, we focus on LOB-based DEX in this work, as they have a longer history of over 3 years.

2.3 Defining Wash Trades
Wash trading has been prohibited in the U.S. by the Commodity Exchange Act in 1936 [10] [11]. The Commoditi Futures Trading Commission (CFTC) defines it as "Entering into, or purporting to enter into, transactions to give the appearance that purchases and sales have been made, without incurring market risk or changing the trader’s market position" [9]. It is also referred to as Round Trip Trading. Actors colluding in or arranging transactions, such that they do not incur market risk, are thereby deemed wash trades. The definition also indicates that actors executing a series of transactions among themselves, after which they end up at the same market position that they had initially, can be considered wash trades. In 2013, the CME Group has issued a Market Regulation Advisory Notice. It provides some guidance as to what may constitute a wash trade. It notes two important criteria: a) the intent of actors to execute fictitious trades that are not subject to market risk, and b) “a wash result - meaning the purchase and sale of the same instrument at the same price, or a similar price, for accounts with the same beneficial ownership” [18, p. 5].
3 RELATED WORK

Academic studies that are related to our work mostly concern wash trading in traditional financial markets and other types of market manipulation in cryptocurrencies. While these will be addressed in detail in the following, other related works concern the graph-based study of token transfer networks [12, 26, 29] and the identification of users with multiple accounts in Ethereum [27].

3.1 Cryptocurrency Price Manipulation

The increasing attention for cryptocurrencies has come with a range of manipulative activities. Several papers have studied the phenomenon of cryptocurrency pump and dump schemes [19, 20, 28, 32] and manipulation of Bitcoin prices [16, 17], both on centralized exchanges. Suspicious trader networks and manipulative transaction patterns have been identified on the Mt. Gox exchange [11]. Much less work exists on manipulation on DEX, though substantial evidence for front running attacks was found [14].

3.2 Wash Trade Detection

Before academic work has focused on the detection of wash trading, a related phenomenon called collusive cliques has been studied using graph clustering and nearest neighbor algorithms [3]. Other methods that have been used include spectral clustering [15], Hidden Markov Models [6] and correlation statistics on aggregated order volume time series [30].

Cao et al. were among the first to explicitly study the detection of wash trades in financial markets [7]. They point out that previous work had focused on collusive and correlated trading behavior rather than the specific pattern of wash trading. Their detection algorithm identifies subsets of a given set of trades that lead to no position change of all involved traders. From a topological perspective, these trades form a closed cycle. To identify these, the algorithm checks the power set of a set of trades, resulting in a runtime of O(2^n).

4 DATASETS AND PREPARATION

In this section, we detail what data we have collected and how it was preprocessed. We also provide information on how we model token trade graphs, and provide high-level insights about the datasets.

We collected all Ethereum blocks, transactions and events related to the EtherDelta and IDEX exchanges, corresponding to a time frame between 02/09/2017 and 05/04/2020. Given the transactions and events, along with the smart contract addresses of IDEX and EtherDelta, information about the executed trades can be obtained.

To extract trades that have been executed on IDEX, we have parsed all successful transactions targeted at the IDEX smart contract1, which call the function trade, identified by the shortened Keccak256 hash \(0xef34b588\). The parameters of this function call include information on the participating accounts, which token has been bought and sold, traded amounts and fees.

Similarly, EtherDelta trades can be extracted by parsing the trade events triggered by the EtherDelta smart contract2. Every successful EtherDelta trade emits a \(\text{trade}\) event, identified by the event topic \(0xef34b588\) starting with \(0xef\), which is triggered by the smart contract when a corresponding transaction has been executed against it. The event also contains information about the participating accounts, exchanged tokens and amounts.

The trade timestamps are obtained from the block the transaction or event was found in. To assess the trade volume in U.S. Dollars, we obtained daily exchange rates of Ether to Dollar from Etherscan. All data is available on Zenodo3, and the code is on GitHub4.

4.1 Data Preprocessing

We have preprocessed the data as follows: As the internal trade amount representation is based on very large integer amounts for precision, we first converted them to floating point values for easier handling. Secondly, the prices of the involved assets are computed based on exchanged amounts, and U.S. Dollar values are joined. We then removed trades containing missing data and possibly unsuccessful IDEX transactions as indicated by the status field which has been introduced with the Ethereum Byzantium fork. We also removed trades in which two tokens were exchanged, leaving trades between Ether and some token. We do so in order to be able to study wash trading of individual tokens traded against a common base currency. These steps remove only a very small fraction of the trades (IDEX: 3, 612 trades, EtherDelta: 3, 095 trades). An overview of the resulting trade datasets is depicted in Table 1.

| Table 1: Overview of IDEX and EtherDelta datasets. |
|-----------------|-----------------|-----------------|
|                | IDEX            | EtherDelta       |
| Start date (UTC) | 09/27/2017 10:57pm | 02/09/2017 11:56pm |
| End date (UTC)   | 05/04/2020 1:22pm  | 05/04/2020 1:22pm  |
| Number of trades | 5,340,537        | 3,573,512        |
| Number of tokens | 249,911          | 323,598          |
| Number of tokens | 1,206            | 6,551            |

1IDEX smart contract address: 0xa0a0d7bce7e4db8f48e0f1e3fa5f6f4050c20
2EtherDelta smart contract address: 0x8d1a97cbb4f7a1fe0395050f3235ce681
3https://zenodo.org/record/4540223
4https://github.com/friedhelmvictor/lob-dex-wash-trading-paper
4.2 Modeling Token Trade Graphs

Each trade in the data contains two participating accounts, and a token amount that was exchanged for a certain Ether amount. Therefore, we can model all trades of a certain type of token as a token trade graph. For the remaining chapters, we define a token trade graph \( G(V, E) \) to be a directed multigraph, where \( V \) is the set of trading account addresses, and \( E \) is the set of trades. The direction of each edge is given by the token flow — from the token seller to the token buyer. Note that this means each trade is represented as only one directed edge, even though both vertices send and receive assets. Finally, each edge also has a weight, which corresponds to the traded token volume.

This setup allows us to observe cycles with matching volumes in a token trade graphs, which is fundamental to the idea of wash trading. To illustrate, Figure 1 showcases a few potential scenarios of wash trading activity within a token trade graph. Each vertex is one trading account, trades belonging to the wash trade set are colored red, and vertices belonging to the set of wash trading actors are colored in a strong yellow. The rest of the graph is slightly faded. Figure 1a shows a loop, i.e. a self-trade. This means a single account executed a trade with itself, which can also be observed in our data. In Figure 1b, three actors trade the same amount in a cycle, slightly modified in Figure 1c, where the trades between trader 4 and 2 are split into two trades. Finally, Figure 1d shows a more complicated scenario. There is a cycle along vertices 1 to 5, but also three partly overlapping sub-cycles \( \{1, 5, 4, 2\} \), \( \{2, 5, 4\} \) and \( \{2, 5, 4, 3\} \). Summing the trades, each trader buys and sells exactly 120 tokens, which leads to no change in their market positions. Also, note that not all trades even among the wash trading actors belong to the wash result, e.g. 30 token from trader 5 to 4. Indeed, actors may try to hide their illegitimate activities among other, legitimate trades.

4.3 High Level Insights

We briefly explore the datasets to illustrate their structure. Particularly interesting is the relationship between the number of trade partners and the number of executed trades, which is visualized in Figure 2. Each cross is one trading account, positioned by the number of trade partners and number of trades participated in. Surprisingly, some traders perform a large number of trades with only a few other accounts.

Secondly, we can explore traded Ether amounts. Related works use the indicator of an unnatural trade size distribution (such as a uniform distribution) as a signal for the existence of wash trading on centralized exchanges. Figure 3 shows the trade size distribution for both DEX analyzed in this paper. Higher amounts are traded less frequently, and there exist local peaks at round numbers such as 5 Ether. We conclude that these distributions appear to conform to natural trading behavior, and are likely unsuitable as an indicator for wash trading on DEX.
In order to detect wash trades, we follow legal definitions and criteria of wash trading by the CFTC and CME. Our aim is therefore to identify sets of trades between collusive trading accounts that lead to no change in the individual position of each participating trader. In other words, for each account within a set of trading accounts, the total amount of purchased assets equals the total amount of sold assets, such that the involved traders essentially hold the same position they had initially. In this section, we present a volume matching algorithm that identifies such trades, which needs to be supplied with a set of trades and a time window parameter. Due to the very high number of trades in our datasets, it is not feasible to check the power set of the trade sets as proposed with previous wash trade detection methods, as this results in exponential time complexity. For this reason, we follow a two-step process:

1. **Account Candidate Set Generation:** We determine candidate sets of potentially collusive traders via an iterative counting of strongly connected components (SCCs) in each token trade graph.
2. **Trade Volume Matching:** For time windows of trades within frequently occurring SCCs, we determine if there exists a trade subnet that leads to no position change for each trader.

### 5.1 Candidate Set Generation with SCCs

We saw in Figure 1, that wash trade scenarios contain at least one cycle. However, an approach of detecting all cycles would not be able to capture the scenario in Figure 1d, where the wash trading structure consists of a cycle with sub-cycles. Therefore, we propose to use the concept of strongly connected components to identify maximal subsets of vertices that are connected via cycles.

A directed graph $G$ is strongly connected if there exists a directed path from every vertex $v$ to every other vertex $u$ in $V$. If $G$ as a whole is not strongly connected, it can consist of strongly connected components (SCCs). An SCC is defined as a maximal subset of vertices $C \subseteq V$ such that there exists a directed path from every vertex $u$ to $v$ and from $v$ to $u$, where $u, v \in C$ [25]. An SCC contains one or multiple cycles, which can be thought of as round trip trades. As such cycles can appear coincidentally, we are particularly interested in trades that are repeatedly part of such SCCs. Therefore, for each token trade graph, we perform an iterative counting of SCCs as illustrated in Algorithm 1 and Figure 4.

First, a directed token trade multigraph is simplified, turning the number of multi edges between two vertices into an edge weight. We then perform multiple iterations of determining all SCCs in the graph, counting how often they have been seen over the iterations.

#### Algorithm 1: Iterative SCC Counting

| Data: | Directed token trade multigraph $G(V, E)$ |
|--------|-------------------------------------|
| Result: | Mapping $M$ of sets of vertices $V$ to counts |
| $G \leftarrow$ simplify($G$, edgeweight = edge count) |
| $M \leftarrow$ Mapping |
| repeat while $|E| > 0$ do |
| $SCCList \leftarrow computeSCCs(G)$ |
| $foreach SCC \in SCCList$ do |
| $SCCV \leftarrow getVertexSet(SCC)$ |
| $M.set(SCCV) \leftarrow M.get(SCCV) + 1$ |
| $endforeach$ |
| $foreach e \in E$ do |
| $weight(e) \leftarrow weight(e) - 1$ |
| $if weight(e) = 0$ then |
| $E \leftarrow E \setminus \{e\}$ |
| $endforeach$ |
| $return (M)$ |

After each iteration, we decrease all edge weights by 1, and remove edges that would end up with an edge weight of 0. When no edges remain, the process is complete. The result are sets of vertices of SCCs with counts indicating how often they have occurred.

After performing Algorithm 1 on all token trade graphs of each DEX, we can visualize the complementary cumulative distribution function (CCDF) of SCC counts in Figure 5. We select those SCCs as our candidate set which occur at least $100$ times. On IDEX, this corresponds to the top $3\%$, on EtherDelta the top $1\%$.

![Figure 5: CCDF of identified SCCs. A small fraction of SCCs appear at least 100 times, indicating at least 100 round trip trades. We choose these as our candidate set.](image-url)
5.2 Volume Matching

After identifying candidate sets of SCCs consisting of trading accounts that repeatedly trade in circular ways, we look at the trades between these accounts, for individual tokens. The goal is to find subsets of trades with matching volumes that lead to no overall position change for the individual accounts. In this section, we therefore formalize position changes based on multiple trades, adapted from Cao et al. [7, 8] for the context of crypto tokens. Finally, we describe our volume matching algorithm of linear time complexity.

5.2.1 Formalizing trader position changes for wash trade detection.

In order to detect trades that lead to no individual position change, we need to sum up trades in such a way that the traded volumes are summed up per trading account involved in these transactions. A volume can be bought or sold, so the position an individual account holds can increase or decrease.

We can formalize a trade \( T \) as in Equation (1), where trader \( a_j \) sells volume \( V \) to trader \( a_i \). We denote trader positions by their account addresses \( a \), since one real-world trader may control several account addresses. The signs indicate the selling or buying position. Since the algorithm that is based on this formalization is run for sets of trades within a particular token, we do not need to indicate the token here.

\[
T = (+a_i, -a_j)V.  
\]

Two (or more) trades are summed up and lead to signed trader positions that are close to \( 0 \). The next challenge is to test various trade sets. Given a set of trades, the algorithm by Cao et al. [8] tests all possible subsets (i.e. the power set), trying to find configurations such that the signed sums of each trader equal zero. However, this leads to an exponential time complexity in \( O(2^n) \), in dependence on the size of the given trade set \( n \). In Section 5.1, we have proposed to generate SCC candidate sets per token trade graph in order to reduce the trade sets to be analyzed. However, these SCCs can still execute thousands of trades in a particular token, so an exponential runtime is not feasible. We therefore propose to further split these trades using a time window \( t \). Also, instead of checking all \( 2^n \) subsets of a window trade set, we propose to check \( n \) subsets, iteratively removing the last trade in the set.

Algorithm 2 illustrates this concept for a set of trades \( S_T \). The signed trader positions are calculated by summing up all given trades (Line 3). The algorithm then iterates through the trades. In each step, it checks whether each trader has a current position \( p_i \) of at most a margin \( m \) of the mean trade size. If the current trade set \( S_W \) meets this condition, a wash trade set is found. If not, the last trade is repeatedly removed from the set (Lines 9-10).

\[
\begin{align*}
\sum_{k=1}^{2} T_k &= T_1 + T_2 \\
&= (+a_i, -a_j)V_1 + (+a_j, -a_i)V_2 \\
&= \{[+a_i]V_1, [-a_j]V_1, [-a_j]V_1, [+a_j]V_2\} \\
&= \{[a_i]p_i, [-a_j]p_j\} \\
&= P
\end{align*}
\]

In other words, when summing up all trades in \( S_W \), all traders \( a_i \) involved in \( S_W \) must hold a relative position of \( p_i = 0 \), allowing for a small deviation \( \alpha \). We define this deviation as a percentage (or margin) \( m \) of the mean trading volume of \( S_W \), e.g. \( m = 0.01 \):

\[
v = m \cdot \frac{\sum_{k \in S_W} V_k}{|S_W|}
\]

We need to choose such a margin for two reasons: In order to make a wash trade less obvious, traders might not buy and sell exactly the same volumes when wash trading. Secondly, each trade incurs a cost, which may decrease a trader’s capital with every trade and therefore also be reflected in slightly decreasing wash trade volumes. We set this margin to be 1%.

5.2.2 Wash Trade Detection Algorithm. Based on the definitions, we can now determine whether a given set of trades yields trader positions that are close to 0. The next challenge is to test various trade sets. Given a set of trades, the algorithm by Cao et al. [8] tests all possible subsets (i.e. the power set), trying to find configurations such that the signed sums of each trader equal zero. However, this leads to an exponential time complexity in \( O(2^n) \), in dependence on the size of the given trade set \( n \). In Section 5.1, we have proposed to generate SCC candidate sets per token trade graph in order to reduce the trade sets to be analyzed. However, these SCCs can still execute thousands of trades in a particular token, so an exponential runtime is not feasible. We therefore propose to further split these trades using a time window \( t \). Also, instead of checking all \( 2^n \) subsets of a window trade set, we propose to check \( n \) subsets, iteratively removing the last trade in the set.

Algorithm 2: Trade Volume Matching

Data: Set of trades \( S_T = \{(+a_{i_1}, -a_{j_1})V_1, ..., (+a_{i_n}, -a_{j_n})V_n\} \), margin \( m \)

Result: Set of trades that constitute a wash result \( S_W \)

1. \( S_W \leftarrow S_T \) // assume all trades are in wash result
2. \( n \leftarrow |S_W| \)
3. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // compute positions according to Equation (2)
4. while \( n > 1 \) do
5. \( v_{mean} \leftarrow \frac{\sum_{k \in S_W} V_k}{|S_W|} \) // compute mean trade size
6. if \( \{p_i \leq m \cdot v_{mean}\} \forall [a_i]p_i \in P \) then
7. return \( S_W \)
8. end
9. \( n \leftarrow n - 1 \)
10. \( S_W \leftarrow S_W[1, ..., n] \) // remove last trade
11. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // update positions
12. end

Algorithm 2: Trade Volume Matching

Data: Set of trades \( S_T = \{(+a_{i_1}, -a_{j_1})V_1, ..., (+a_{i_n}, -a_{j_n})V_n\} \), margin \( m \)

Result: Set of trades that constitute a wash result \( S_W \)

1. \( S_W \leftarrow S_T \) // assume all trades are in wash result
2. \( n \leftarrow |S_W| \)
3. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // compute positions according to Equation (2)
4. while \( n > 1 \) do
5. \( v_{mean} \leftarrow \frac{\sum_{k \in S_W} V_k}{|S_W|} \) // compute mean trade size
6. if \( \{p_i \leq m \cdot v_{mean}\} \forall [a_i]p_i \in P \) then
7. return \( S_W \)
8. end
9. \( n \leftarrow n - 1 \)
10. \( S_W \leftarrow S_W[1, ..., n] \) // remove last trade
11. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // update positions
12. end

Algorithm 2: Trade Volume Matching

Data: Set of trades \( S_T = \{(+a_{i_1}, -a_{j_1})V_1, ..., (+a_{i_n}, -a_{j_n})V_n\} \), margin \( m \)

Result: Set of trades that constitute a wash result \( S_W \)

1. \( S_W \leftarrow S_T \) // assume all trades are in wash result
2. \( n \leftarrow |S_W| \)
3. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // compute positions according to Equation (2)
4. while \( n > 1 \) do
5. \( v_{mean} \leftarrow \frac{\sum_{k \in S_W} V_k}{|S_W|} \) // compute mean trade size
6. if \( \{p_i \leq m \cdot v_{mean}\} \forall [a_i]p_i \in P \) then
7. return \( S_W \)
8. end
9. \( n \leftarrow n - 1 \)
10. \( S_W \leftarrow S_W[1, ..., n] \) // remove last trade
11. \( P \leftarrow \sum S_W = \{[a_i]p_i, ..., [a_j]p_j\} \) // update positions
12. end
We now study wash trading structures, what proportion of tokens wash trading activity on IDEX and EtherDelta:

We now examine the extent to which the tokens traded on each DEX are affected by wash trading. We compute the share of a token’s total trading volume that can be attributed to wash trading. Based on this, we show the share of all tokens that exhibit at least a certain wash trading percentage in Figure 7. On both exchanges, at least 10% of the traded tokens have a wash trading share of at least 20%. On EtherDelta, almost 10% of the tokens have been entirely wash traded. On IDEX, this only applies to less than 1% of all traded tokens. While the figure illustrates the minimum share of wash trading activity, the total share of tokens that have ever been subject to at least one set of wash trades, irrespective of their volume, is 31.54% for IDEX, and 42.12% for EtherDelta.
6.3 Wash Trade Activity in a Token’s Lifespan

To determine in which phase of a token’s trading lifespan wash trading occurs, we first determine the relative time position of each wash trading activity in relation to the first and last (legitimate) trading activity in the respective token. We then compute the median of these wash trading activities per token. Figure 8 shows the results for both exchanges. It can be seen that wash trading is more frequent at the beginning and at the end. The peaks at the beginning of the tokens’ lifespans may be attributed to manipulators wanting to generate trading volume so that real traders become aware of a new token. The fact that wash trading also occurs frequently at the end could be due to a token showing only little trading activity and manipulators wanting to increase the trading activity again. Since the trading volume of 10% of the tokens traded on EtherDelta consists almost exclusively of wash trades (see Figure 7), the median wash trading time in the lifespan of these tokens may be close to 0.5. This may explain the peak of tokens with increased wash trading activities in the middle of the lifespan.

6.4 Wash Trade Activity Over Time

We now look at the achieved wash trading volume over time, first from an absolute perspective and then relative to the rest of the (legitimate) trading volume. Figure 9 shows the wash trading volume by month, over a period of about 3 years. IDEX activity started slightly later than on EtherDelta, reaching its peak in June 2018 with over $15 million in monthly wash trading volume. On EtherDelta, activities are focused on the period from mid-2017 to mid-2018, with more than $12 million in wash trading volume in January 2018. Activities on EtherDelta and IDEX flatten off sharply from mid-2018 and mid-2019 respectively, so that virtually no wash trading volume is visible.

Figure 10 shows the relative share of wash trading volume on a weekly basis. Notably, IDEX exhibits several weeks with over 50% and even close to 100% of the total trading volume being wash traded at the end of 2017. From 2018 onward, a share of up to 20% can be observed on both exchanges, which then further declines in 2019 and 2020.

Both graphs show peaks in late 2017 and early 2018. This falls into a phase in which the public awareness of cryptocurrencies and tokens was particularly high. At this time, many token prices reached record highs, only to be followed by a crash in early 2018.

6.5 Summary

Finally, we show detailed wash trading statistics for IDEX and EtherDelta in Table 2. Note that all statistics referring to wash trades include both self-trades and more complex wash trades. On EtherDelta, the majority of wash trades are self-trades, whereas on IDEX, the self-trades only account for a small share. However, the
total wash trading volume exceeds $75 million on both exchanges, with a total of $159 million. On IDEX, illegitimate accounts traded an equivalent of over $12 million with themselves, on EtherDelta almost $65 million. The fraction of all trades that are wash trades is about 4% on IDEX and 1.3% on EtherDelta. On both exchanges, more than 30% of the tokens that are tradable have been subject to wash trading activity. On EtherDelta, the wash trades were executed by over 5,000 different accounts, leading to an average of about 13 wash trades per manipulative trader account. On the other hand, only 659 trader accounts were responsible for over 200,000 wash trades on IDEX. Finally, considering that both DEX charge trading fees of 0.3%, the total fees paid to the exchanges to perform wash trading, lead to an equivalent of 478,134 U.S. Dollars.

### Table 2: Wash Trades Summary for IDEX and EtherDelta.

|                     | IDEX  | EtherDelta |
|---------------------|-------|------------|
| # Self-Trades       | 28,193| 58,791     |
| # Wash Trades       | 213,029| 69,711    |
| Self-Trades Share (Of All Trades) | 0.53% | 1.09%     |
| Wash Trades Share (Of All Trades) | 3.99% | 1.30%     |
| Total Self-Traded Volume ETH | 23,716.08 | 136,548.22 |
| Total Wash Volume ETH | 273,027 | 155,155    |
| Total Self-Traded Volume USD | 12,587,517 | 64,801,288 |
| Total Wash Volume USD | 83,531,254 | 75,846,518 |
| Wash Trade Fees Received USD | 250,594 | 227,540    |
| # Self-Traded Tokens | 239   | 2,758      |
| # Wash Tokens       | 380   | 2,759      |
| Wash Token Share    | 31.54%| 42.12%     |
| # Self-Trader Accounts | 332   | 5,501      |
| # Wash Trader Accounts | 659   | 5,533      |
| # Analyzed SCCs     | 199   | 48         |
| # SCCs with Wash Trading | 193   | 47         |
| Mean # Tokens Washed per SCC | 2.46  | 1.62       |

![Weekly wash volume share](image)

**Figure 10:** Weekly share of wash trade volume on both IDEX and EtherDelta. At the end of 2017, the majority of trades on IDEX were wash trades.

7 DISCUSSION

Our wash trade detection method has identified various wash trading structures and manipulated trading volumes worth $159 million. We have used several parameters for the method, which we have chosen very conservatively. This means our findings constitute a lower bound of the actual extent of wash trading activities. If we were to set these parameters in a more relaxed way, e.g., considering SCCs with at least 10 repetitions instead of 100, or a higher margin at which we define that a trader effectively did not change their position, the number of detected wash trades would be higher. At the same time it would be possible that some trades are falsely recognized as wash trades. An example of this would be individual traders who speculate in the short term, quickly buying back their sold tokens, while coincidentally trading with the same other trader. Especially in the case of self-trades, however, there is no room for interpretation.

The question of why so much wash trading is done at all can be answered as follows: Considering the thousands of different tokens that exist, it can be difficult for token founders to get enough attention to be listed on known exchanges. Therefore, some token founders may be thinking about giving their project a jump start with fake trading volume, which is one reason also suggested by a recent report [5]. However, the wash trading activities on IDEX and EtherDelta have declined significantly since mid-2018. This is likely due to the emergence of new AMM-based exchanges such as Uniswap and others. We do believe that actors willing to boost a token using wash trades will also attempt to do so on the now more popular AMM DEX. The question remains how wash trading can be performed on these DEX, since users do not trade with each other but against a liquidity pool.

#### 7.1 Potential Countermeasures

A standard countermeasure that is implemented by many exchanges in traditional markets is self-trade prevention functionality. A single account is prevented from filling its own buy or sell orders. Even though IDEX’s off-chain trade matching engine claims to have self-trade prevention functionality 3, we have observed successful self-trades up until the end of our dataset (May 2020). EtherDelta does not seem to implement such functionality. However, as shown in the analysis of the wash trading structures, various other trading topologies exist that can circumvent such a self-trade prevention functionality.

While defining a limit to the number of trades that can be performed with the same trading partners might seem like a viable approach at first, it may disrupt legitimate trades, if the number of traders in a token market is relatively small. Another option to make it harder for wash traders to operate with multiple accounts is the introduction of Know Your Customer (KYC) procedures, that require traders to identify themselves with official identity information. This is standard in the traditional domain, and on most centralized exchanges, but not widely adopted in the realm of decentralized finance. IDEX has introduced KYC in July 2019 6, which coincides with lower wash trading volumes.

---

3https://docs.idex.io/#self-trade-prevention

6https://medium.com/idex/idex-kyc-transition-period-and-updated-asset-availability-for-us-markets-set-to-begin-d45e945f842d
8 CONCLUSION AND FUTURE WORK

This paper is the first to analyze the phenomenon of wash trading on decentralized cryptocurrency exchanges. We have presented a method to detect wash trades, and have examined the decentralized exchanges IDEX and EtherDelta on the Ethereum blockchain. On an empirical basis, we have identified wash trading activities in excess of 159 million U.S. Dollars and discovered common wash trading structures. These structures consist mainly of one or two accounts, but there also exist more complex patterns. Surprisingly, self-trades occur frequently, which could easily be prevented. On both exchanges, more than 30% of all tokens have been subject to wash trading activity, and 10% of the tokens on EtherDelta have been almost exclusively washed trade. These figures represent a lower bound, but yet underpin the need for countermeasures that also work in decentralized systems.

Future work can take multiple directions. Our volume matching algorithm could be extended to check all contiguous subsets, which would lead to approximately quadratic complexity, requiring a highly efficient implementation or distributed computing. In general, devising new wash trade detection algorithms can be of interest. These might also include temporal aspects, which we have not considered in this work. We also consider the investigation of recently popularized DEX such as Uniswap to be promising, which are not based on the order book model. Furthermore, other aspects can be investigated, such as the effect of wash trades on price development and the token user community. Finally, new decentralized finance concepts such as loans, derivatives and insurances may offer further potential for manipulation which is not yet well understood.

REFERENCES

[1] 2020. 7 U.S. Code CHAPTER 1—COMMODITY EXCHANGES. https://www.law.cornell.edu/uscode/text/7/chapter-1
[2] Hayden Adams. 2020. Uniswap Whitepaper. https://arxiv.org/pdf/1906.02979v1.pdf
[3] Manoj Apte and Girish Palshikar. 2005. Collusion Set Detection using Graph Clustering.
[4] Paddy Baker. 2019. OKEx Comes Clean On Wash Trading Problem. https://www.cryptobriefing.com/okex-wash-trading/
[5] Anna Baydakova. 2019. For $15K, He’ll Fake Your Exchange Volume – You’ll Get on CoinMarketCap. https://www.cointelegraph.com/article-for-15k-he-ll-fake-your-exchange-volume-you-ll-get-on-coinnmarketcap
[6] Longbing Cao, Yuming Ou, Philip S. Yu, and Gang Wei. 2010. Detecting Abnormal Coupled Sequences and Sequence Changes in Group-Based Manipulative Trading Behaviors. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Washington, DC, USA) (KDD ’10), Association for Computing Machinery, New York, NY, USA, 85–94. https://doi.org/10.1145/1835804.1835818
[7] Yi Cao, Yuhua Li, Sonya Coleman, Ammar Belatreche, and Thomas Martin McGinley. 2014. Detecting wash trade in the financial market. In 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFE). IEEE, 85–91.
[8] Y. Cao, Y. Li, S. Coleman, A. Belatreche, and T. M. McGinley. 2016. Detecting Wash Trade in Financial Market Using Digraphs and Dynamic Programming. IEEE Transactions on Neural Networks and Learning Systems 27, 11 (2016), 2351–2363.
[9] Commodity Futures Trading Commission (CFTC). 2020. CFTC Glossary — Wash Trading. https://www.cftc.gov/LearnAndProtect/EducationCenter/CFTCGlossary/glossary_wxzy.html#washtrading
[10] Commodity Futures Trading Commission (CFTC). 2020. Commodity Exchange Act & Regulations. https://www.cftc.gov/LawRegulation/CommodityExchangeAct/index.htm
[11] W. Chen, J. Wu, Z. Zheng, C. Chen, and Y. Zhou. 2019. Market Manipulation of Bitcoin: Evidence from Mining the Mt. Gox Transaction Network. In IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. 964–972.
[12] Wei-Chen, Tuo Zhang, Zhiguang Chen, Zhibin Zheng, and Yutong Lu. 2020. Traveling the token world: A graph analysis of Ethereum ERC20 token ecosystem. In Proceedings of The Web Conference 2020. 1411–1421.
[13] Lin William Cong, Xi Li, Ke Tang, and Yang Yang. 2019. Crypto Wash Trading. Available at SSRN 3530220 (2019).
[14] Philip Daian, Steven Goldfeder, Tyler Kel, Yunqi Li, Xueyu Zhao, Itdo Bentov, Lorenzo Breidenbach, and Aris Juels. 2019. Flash boys 2.0: Frontrunning, transaction reordering, and consensus instability in decentralized exchanges. arXiv preprnt arXiv:1904.05234 (2019).
[15] Markus Franke, Bettina Hoser, and Jan Schröder. 2008. On the Analysis of Irregular Stock Market Trading Behavior. In Data Analysis, Machine Learning and Applications, Christine Freisach, Hans Burkhardt, Lars Schmidt-Thieme, and Reinhold Decker (Eds.). Springer Berlin Heidelberg, 355–362.
[16] Neil Gandal, JT Hamrick, Tyler Moore, and Tali Oberman. 2018. Price manipulation in the Bitcoin ecosystem. Journal of Monetary Economics 95 (may 2018), 86–96. https://doi.org/10.1016/j.jmoneco.2017.12.004
[17] John M. Griffin and Amin Shams. 2020. Is Bitcoin Really Untethered? The Journal of Finance 75, 4 (2020), 1913–1964. https://onlinebrielibrary.wiley.com/doi/abs/10.1111/jofi.12903
[18] CME Group. 2013. https://www.cfcf.gov/sites/default/files/stellen/groups/public/rulesandproducts/documents/idocs/rul070913meccombotnymexcomandk1.pdf
[19] Josh Ramps and Bennett Kleinberg. 2018. To the moon: defining and detecting cryptocurrency pump-and-dumps. Crime Science 7, 1 (2018), 18.
[20] Tao Li, Donghua Shi, and Baolian Wang. 2019. Cryptocurrency pump-and-dump schemes. Available at SSRN 3263741 (2019).
[21] Sanders X Lin, Eric Budish, Lin William Cong, Zhigao He, Jonatan H Bergquist, Mohit Singh Panesar, Jack Kelly, Michelle Lauer, Ryan Prinster, Stephanie Zhang, et al. 2019. Deconstructing Decentralized Exchanges. Stanford Journal of Blockchain Law & Policy (2019).
[22] Patrick McCorry, Malte Möser, and Syed Taha Ali. 2018. Why preventing a cryptocurrency exchange heist isn’t good enough. In Cambridge International Workshop on Security Protocols. Springer, 225–233.
[23] Satoshi Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. (2008).
[24] Kaihua Qin, Liyi Zhou, Benjamin Livshits, and Arthur Gervais. 2020. Attack-defi theDeFi Ecosystem with Flash Loans for Fun and Profit. arXiv preprint arXiv:2003.03810 (2020).
[25] Kevin Wayne Robert Sedgewick. 2020. Algorithms, 4th Edition. https://algs4.cs.princeton.edu/home/. Accessed: 2020-08-10.
[26] Shahar Somin, Goren Gordon, and Yaniv Altshuler. 2018. Network analysis of erc20 tokens trading on ethereum blockchain. In International Conference on Complex Systems. Springer, 439–450.
[27] Friedhelm Victor. 2020. Address clustering heuristics for Ethereum. In International Conference on Financial Cryptography and Data Security. Springer, 617–633.
[28] Friedhelm Victor and Tanja Hagemann. 2019. Cryptocurrency Pump and Dump Schemes: Quantification and Detection. In 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, 244–251.
[29] Friedhelm Victor and Bianca Katharina Lüders. 2019. Measuring ethereum-based DEX exchanges: from decentralization to regulation. In Proceedings of The Web Conference 2019 - IEEE Conference on Computer Communications. 1113–1120.
[30] Shahar Somin, Goren Gordon, and Yaniv Altshuler. 2018. Network analysis of erc20 tokens trading on ethereum blockchain. In International Conference on Complex Systems. Springer, 439–450.
[31] Friedhelm Victor. 2020. Address clustering heuristics for Ethereum. In International Conference on Financial Cryptography and Data Security. Springer, 617–633.
[32] Friedhelm Victor and Tanja Hagemann. 2019. Cryptocurrency Pump and Dump Schemes: Quantification and Detection. In 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, 244–251.
[33] Friedhelm Victor and Bianca Katharina L üders. 2019. Measuring erc20 tokens trading on ethereum blockchain. In International Conference on Financial Cryptography and Data Security. Springer, 617–633.
[34] Jiahua Xu and Benjamin Livshits. 2019. The anatomy of a cryptocurrency pump-and-dump scheme. In 28th (USENIX) Security Symposium (USENIX) Security 19) 1669–1685.