Quantifying fake volumes on cryptocurrency exchanges

Eitan Galam

May 18

Abstract

Blockchain technology is by design meant to bring transparency and self-regulation. Technological limitations have favored the growth of centralized (offchain) exchanges over decentralized (blockchain) exchanges. Most centralized cryptocurrency exchanges are neither regulated neither transparent in the sense that users' sole source of information is data provided by the exchange itself.

The claim has been made that a significant part of the volume allegedly traded on these exchanges is actually fake.

Increasing volume numbers, increases the exchange's ranking and popularity among both traders and post-ICO firms looking to list their token. Trading volumes are a major indicator for liquidity, slippage assumptions, risk management models and high volumes is a strong positive factor for a token.

Therefore, reliable volume numbers is essential for making informed decision regarding where to list tokens and which exchange to trade on.

In a widely republished article[1] Sylvain Ribes claimed that up to 93% of the volume reported on some major exchanges is actually fabricated. Ribes' post opened the way to a more in-depth and widely spread analysis. In this article, we have extensively analyzed all the trades provided by more than 50 exchanges and quantified the likelihood and extent of manipulation in reporting and fake trading (known as wash trading).

About the Author

The author is a mathematician, a quantitative trader and senior executive in the Investment Industry.
Mr Galam studied Applied Mathematics and Finance at Harvard University where he had the privilege to study under Prof Merton[1] and where he specialized in the use of advanced Artificial Intelligence methods in trading strategies. Before that he studied fundamental computer sciences at the Ecole Normale Superieur of Paris then headed by Prof. Jacques Stern[2].

Mr Galam spent ten years in the quantitative trading industry working as a senior quant trader and hedge fund executive in both high frequency and statistical arbitrage proprietary desks.

Since 2012, Mr Galam is a Managing Partner at Mayan Capital a quantitative Asset Manager based in New York. Lately, he co-founded M Square Solutions a company dedicated to market-making services to Blockchain firms. He advises financial institutions and executives on investment and risk management strategies, and lately, has been advising major blockchain firms in their listing strategy.

[1] Robert C. Merton - Nobel prize of Economics and one of the brains behind the famous Black & Scholes Formula [2]

[2] Prof. Jacques Stern is known as the father of the French cryptography. He received the 2006 CNRS Gold Medal. His notable work includes the cryptanalysis of numerous encryption and signature schemes, the design of the PointchevalStern signature algorithm, the NaccacheStern cryptosystem and NaccacheStern knapsack cryptosystem, and the block ciphers CS-Cipher, DFC, and xmx. Recently a spectacular total break of the SFLASH cryptosystem was discovered by Jacques Stern.

Contents

1 Background

2 Why does it matter anyway?

2.1 For traders

Both short term and longer-term traders see trading volumes as a key indicator.

A recent survey by Encrybit[3] found that 36% of traders see lack of liquidity as a main concern.

Short term holders (including market makers and high/middle frequency arbitragers) see in trading volumes an indication of the volume of trades they will be able to execute on a daily basis which has a direct impact on the total profit they will be able to generate.
Moreover, execution slippage is a key factor in any trading strategies as it impacts greatly the ability of the trader to enter positions which would *theoretically* be profitable. At the same time, trading volumes is often used as a easy-to-get proxy for market liquidity which imply a *de-facto* ceiling to the profit which can be generated everyday on a single pair.
Short term traders will not get involved in any market whose expected profits do not reach a given treashold, which naturally exclude under-traded markets.

Longer term traders are more interested in the intrinsic value of a token with the expectation of being able to sell the position after some time for a profit. While slippage might not be as key as it is for short term traders, volume is a strong indicator of a healthy token. Low trading volumes generally implies higher volatility and higher tail risks. Moreover, the final objective of any investors is to be able to close the position and limited volumes would impact the investor's exit opportunity all the more in scenarios of panic in which the investor might want to stop his loss but might not be able to do so due to shrinking liquidity.

In the crypto trading industry, the volume issue matters even more given the fear of "scams" and the general belief that more well-known tokens and exchanges are less likely to be fraudulent.

2.2 For token issuers

One of the promises made by token issuers is to have their token listed in a reasonable time to provide liquidity for early-stage investors willing to lock their gains. Listing their newly-issue token is usually one of the first projects of a post-ICO firm. The process usually take 1 to 6 months, depending on the exchange they are willing to list-on and the quality of the project.

Listing its token on a *top-tier* exchange is by itself a sign of strength from

the issuer's team and is a major marketing event after the ICO. Moreover, blockchain entrepreneurs tend to have a bullish view on the overall crypto market and listing on major exchanges gives them added exposure to investors which in turn will push their token upward.

In this perspective, trading volumes is the main indicator for what makes an exchange *top-tier*. Higher ranking exchanges are tougher and more expensive to list on, but are believed to bring more upside.

Should the volumes listed on dedicated website proven to be fake, exchange ranking should be re-assessed which would dramatically change the listing strategy of token issuers.

As stated above, volume is used as a proxy for market liquidity as the two are usually highly correlated. Lower volume on an exchange means than liquidity will actually be lower than believed based on official volume numbers. Low liquidity for a token increases the risk of one trade crushing the market down. Such a move might create panic among the token's holders which might in turn lead to a selloff which would drive the price of the token to the lowest for no good reason.

As token price and stability are major factors in evaluating the success of a listing strategy, mis-calculating trading volumes can lead to disasters for a token and for the project behind it.

3 Historic

We first encountered the issue while doing some market making on a second-tier exchange (official daily volume 250m USD).

While constantly being positioned on the first limit on both the bid and the ask side of the market, we realized that we could not trade more than a few percent of the daily volume. After looking into it, we noticed some trades being printed in the middle of the spread without seeing any order added to the book.

Even if hidden-order were placed on the market without us being able to see them, the range of the trades implied spreads of 20 bps (0.02 %) for a newly issued token which is the kind of spread of a mid-cap stock on the NYSE.

After some quick modelisation, it seemed clear that this practice was widely spread among trading pairs on this exchange. As a market maker, we had to review our trading expectations both in terms of expected volumes and benefits.

4 Introduction

In this paper, we have thoroughly analyzed the whole reporting process from the trading tape to the end user. Our goal was to provide concrete numbers and to pinpoint the sources of manipulation. This is done through 4 sub-studies:

1. Checking reporting websites
2. Checking daily volumes numbers provided by the exchanges' API
3. Checking fake trades printed on the tape by exchanges
4. Quantifying self-trading activities[3]

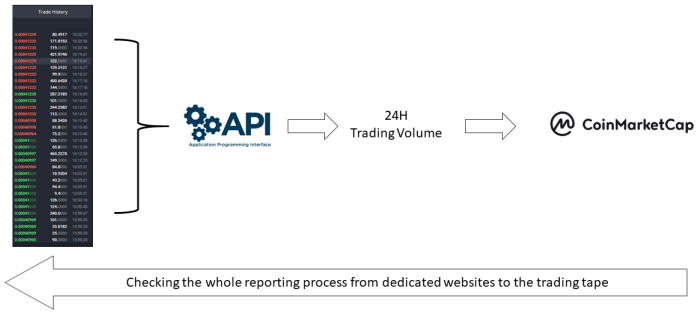


Figure 1: Volumes reporting from the trading tape to the end user

The first source of information for trading volumes are dedicated websites. In the first section of this study, we have implemented the methodology used by major websites and check our numbers against theirs, in order to make sure that websites do not knowingly publish fake numbers.

It should be noted that, while paying a website to report inflated numbers might sound like the easiest and most straight-forward way to increase an exchange ranking, it would be very unlikely to happen since it would require making sure all reporting websites agree with the scam.

Dedicated venues usually use daily volumes as provided by the exchange through their APIs. In the second part of our study, we have checked the data provided by the exchanges themselves to verify whether exchanges report fake numbers through their API.

Would an exchange decide to publish fake numbers, it would have to choose between two options:

[3]This study will be done in a separate dedicated study

1. Publish daily numbers which do not actually reflect the trades printed on the tape

2. Print fake trades on the tape, which would in turn increase daily volumes

In this study, we have systematically and extensively checked these two assumptions by analyzing all trading activity reported by 50 centralized exchanges ranging from the most active exchanges to second and third-tier exchanges.

In each section, we have analyzed results both in average and on a pair by pair basis.

5 Media Reporting

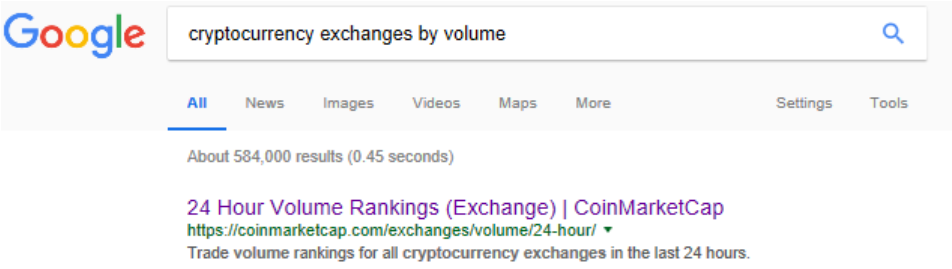
5.1 Methodology

As per coinMarketCap's FAQ page, the primary source of information used by the website is the 24h volumes provided by the exchanges. In this first section, we analyzed all the pairs listed on all the exchanges considered and systematically checked the numbers provided by the API against those of the website.

As clearly stated in the website, numbers are not updated in real time. We have taken the assumption that any difference below 10% should be considered as being explained by the lag of updating the numbers on the website.

In this section, we tried to verify/invalidate the following claims:

- Numbers do not come from the exchanges' API as stated in the FAQ
- API numbers are willingly modified



5.2 Results

As shown below and as expected, volumes reported by coinmarketcap are fully in line with the numbers provided by the exchanges' API.

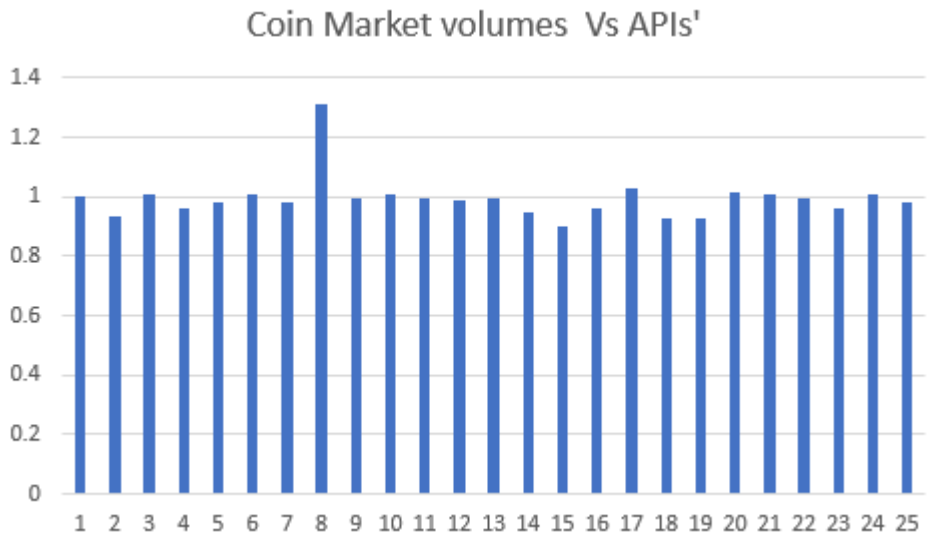


Figure 2: Ratios of trading volumes as reported by coinmarketcap website divided by the volumes provided by the exchanges' API for 25 exchanges

As expected, coinmarketcap's reporting seems to be done in accordance with its methodology[4]. As stated above, we did not have any serious reason to think that dedicated websites temper with volumes data but once this hypothesis verified we know that any manipulation comes from the exchanges themselves.

[4]Exchange number 8 showing unexpected results might come from this exchange not having been updated. One error over 25 points does not sound like a reason to fear manipulation

6 Exchange Reporting

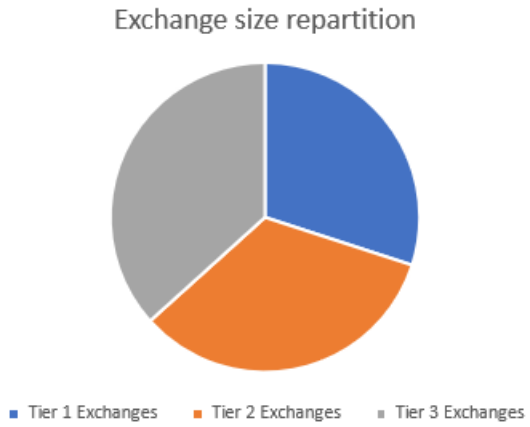
6.1 Methodology

A second layer of potential manipulation is the daily volumes as reported by the exchanges themselves.
In this section, we have recorded all trading activity reported on every pairs traded, and we have compared the 24h trading volumes reported by the API against the sum of all trades recorded on the trading tape over the same period.

Our objective was to check the following assumptions:

- Global volume numbers reported by some exchanges are inflated
- Volumes on specific pairs are inflated

After filtering exchanges which do not report daily volumes or not on a sufficient number of pairs, we end up with 30 exchanges which are evenly distributed in terms of volume.



6.2 Results

Looking at global volumes, we found that while most exchanges' reports seemed fair in average, **a whole 20% of exchanges reported in average between 13% and 100% more than what the trade tape actually shows**.

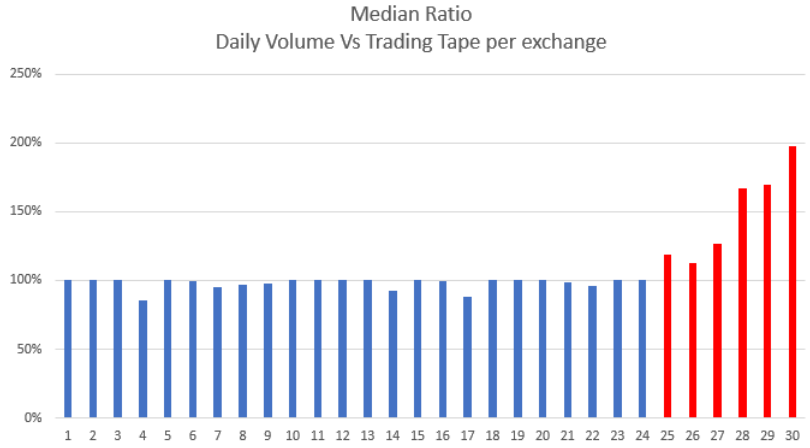


Figure 3: Median Ratio of daily volume reported by the API versus the total amount of tokens printed on the tape

If we look closer at the results and check on each pair the ratio of reported daily volume Vs. all trades on the tape, we find that most exchanges actually inflate daily volumes on some specific coins. Even though, these specific manipulations do not impact significantly the total volume reported by the exchange (globally), it does matter a lot for people looking at the concerned tokens. **This study shows that another 40% of exchanges play with numbers** (i.e report at least 20% more trades than what it shows on the tape) on a significant number of coins. **This means that 60% of the exchanges considered tamper with the daily volumes they report through their API.**

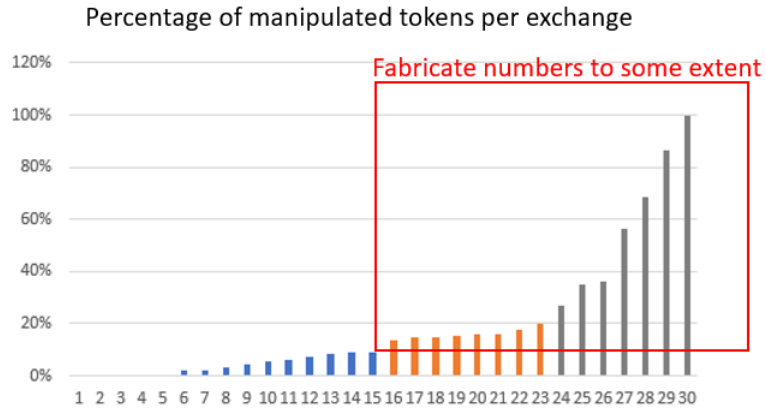


Figure 4: Percentage of pairs reporting more than 20% more volume than what was actually printed on the tape, per exchange

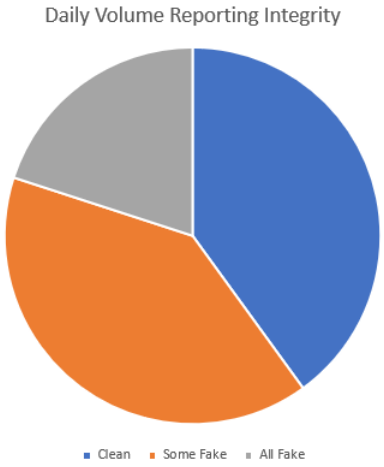


Figure 5: 60% of the exchanges considered seem to temper with daily volumes reporting on more than 10% of the pairs listed

If we look at these numbers per token, we find that some tokens tend to be systematically manipulated over all pairs and exchanges. In the following figure we have analyzed all tokens and computed the ratios of the number of pairs which shows manipulation over the total number of pairs traded.

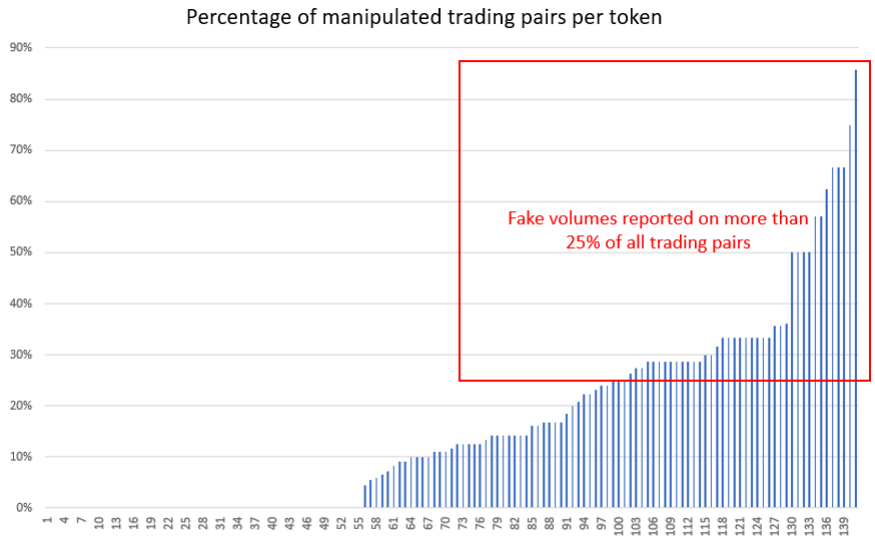


Figure 6: Percentage of manipulated pairs per token

The figure above, shows that exchanges consistently inflate trading volumes

on some specific tokens which indicates that **tempering with volume reporting is done in a strategic way by exchanges which are trying to *push* some tokens**.

In the graph below, we show the percentage of manipulated pairs across all exchanges (for instance COI/BTC, COI/USDT, COI/ETH on all exchanges for a token COI). After filtering out coins that are tradable on less than 8 pairs, we still find that some coins tend to be systematically manipulated across all pairs and exchanges.

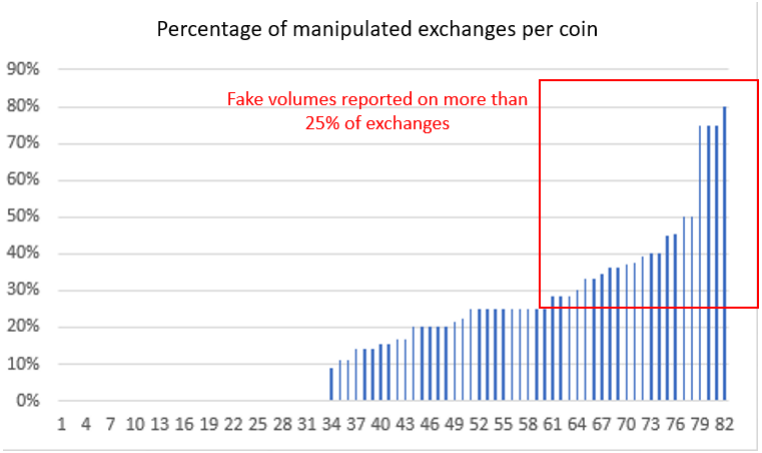


Figure 7: Percentage of manipulated exchanges per coin

These results indicate that exchanges inflate volume numbers on certain *key-coins*, either because these are the coins people are looking at or because they have special interests in showing strong volumes on these coins.

In order to verify if there were some consistency among exchanges to manipulate specific coins, we have computed the ratio of exchanges manipulating volumes per coin[5] (in the following figure, only coins listed on at least 4 exchanges are shown). The

[5]For instance, if a coin COI is listed on 6 exchanges and we found that 4 exchanges report at least 20% more trades than what actually happened we would mark that 66% of exchange manipulate COI

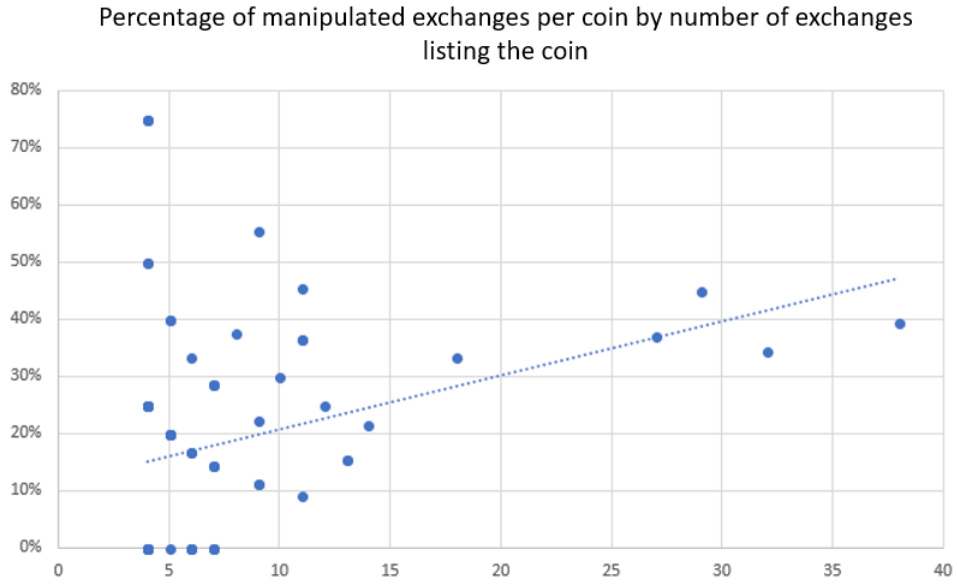


Figure 8: Percentage of manipulated exchanges per coin by number of exchanges which list the coin

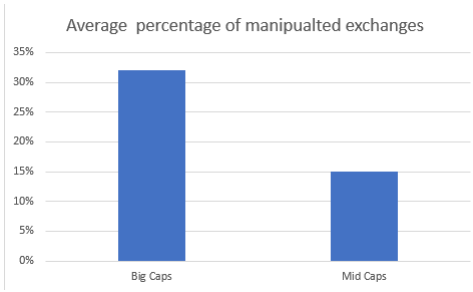


Figure 9: Percentage of manipulated exchanges per coin by number of exchanges which list the coin

These graphs indicate two major findings:

1. Some specific altcoins are systematically manipulated across exchanges (which could indicate a policy of token issuers of making deals with exchanges to have their volumes manipulated)
2. Major coins tend to be more manipulated by exchanges (which would indicate manipulation done by exchanges on most-visible coins)

The numbers clearly prove that a significant numbers of exchanges (and to some extent token issuers) have clear policies of fabricating

volumes on strategic pairs to improve their ranking and attract new traders and tokem issuers.

7 Wash trading

7.1 Methodology

In this section, we have analyzed the trading tape itself. Our objective was to assess if the trades reported on the tape are legitimate trading activities or some kind of wash trading.

Theoretically, there would be two options for an exchange or a token-issuer wishing to engage in wash trading:

1. Report trades that do not correspond to actual orders (which is doable only by the exchange itself)

2. Create orders and trade with themselves (which could be made by any player interested in inflating volumes)

In this paper we have focused on the first type of fake trades.[6]

In this study, we have quantified the number of trades being printed in the middle of the spread or without any order being lifted in the orders book's first limit.
In order to do that, we constantly followed the first limit of the book (bid/ask) and we marked all trades that were printed in the middle of the spread without the spread being changed.

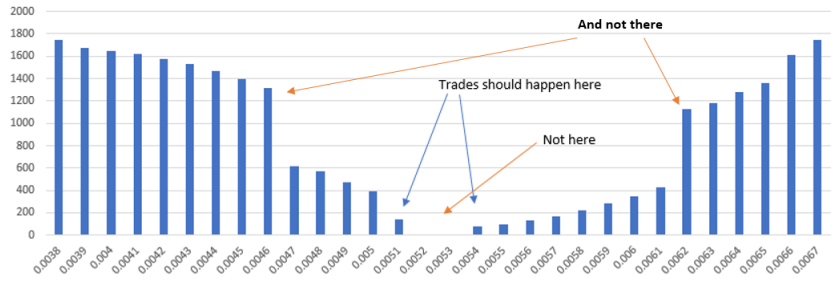


Figure 10: Trades should not be printed inside or outside the first limit

7.2 Results

In the following graph, we show the median "fake trades" percentages per exchange.

[6] We leave the second type of fake trades for a dedicated in-depth analysis

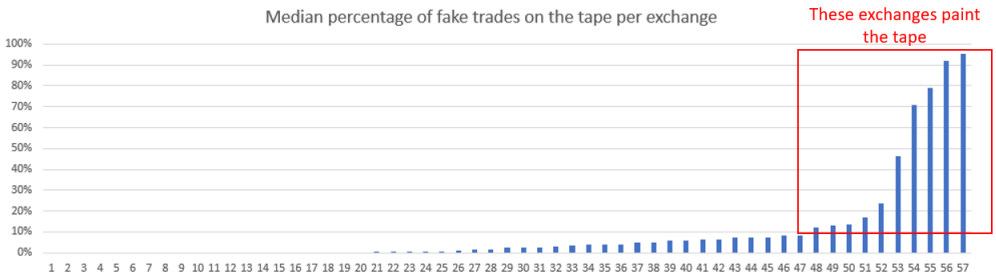


Figure 11: Some exchanges clearly have a policy of printing trades in the middle of the spread

It should be notice that most exchanges are steadily fully clean, which validates our methodology that such *in the spread* trades are exchange induced and not natural behavior.

Looking at each pair, we find that **17% of the exchanges print a significant amount of fake trades globally and that 28% of exchanges actively use wash trading on specific coins (up to a chocking 95%).**

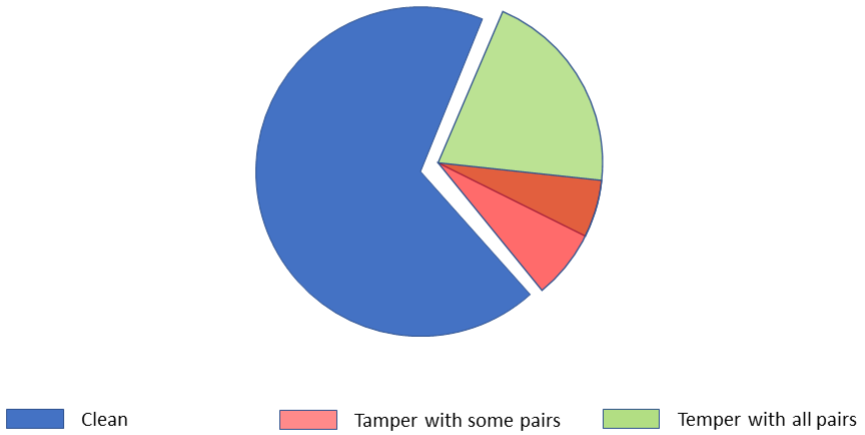


Figure 12: A third of all exchanges seem to print fake trades to significantly inflate volumes on at least some pairs

8 Conclusion

This study is the first publicly available paper quantifying the extent of volume manipulation on crypto-exchanges.

The findings clearly show that a significant portion of the total official trading volume on crypto is fake.[7] More importantly, we showed that some exchanges are intensively active in tempering with numbers which significantly alter their effective ranking. **The most important finding might be that volume manipulation is very high (up to 95%) on specific exchanges/pairs.**

Based on a client's specific needs (a token issuer or a trader), an exchange ranked in the top 20 in terms of volume can actually be less liquid than an unknown exchange.

This study reveals the absolute need to properly understand reported trading activities on exchanges before getting involved with specific tokens and even more importantly before listing new tokens. Listing strategy, should be thoroughly analyzed to optimize the real expected volume on a token and increase the probability of positive trend.

At an industry level, this study shows the need for increased transparency in the crypto-trading space. External regulation is far from being the preferred option for industry players and auto regulation seem tough to achieve in the short term. With decentralized exchanges still lagging way behind centralized ones in terms of liquidity, we expect the emergence of some kind of hybrid exchanges using the blockchain technology to add traceability and transparency to their trading activity or some independent label to certify exchanges' activities.

[7] In Israel, we have an expression that says *"Good Morning Elyahu"* which means that what we have just stated was known by everyone. The main novelty here is that the claim has been scientifically proven and quantified

References

[1] https://medium.com/@sylvainartplayribes/chasing-fake-volume-a-crypto-plague-ea1a3c1e0b5e

[2] https://en.wikipedia.org/wiki/Robert_C._Merton

[3] https://encrybit.io/insights/crypto-exchange-problems.html