Analyzing Price Deviations in DeFi Oracles

Ankit Gangwal[1(✉)], Rahul Valluri[1], and Mauro Conti[2]

[1] International Institute of Information Technology, Hyderabad, India
gangwal@iiit.ac.in, rahul.valluri@students.iiit.ac.in
[2] University of Padua, Padua, Italy
mauro.conti@unipd.it

Abstract. Decentralized Finance (DeFi) promises to transform the traditional financial systems into fair and transparent protocols that do not require trusted third parties. To circumvent the high volatility of crypto-assets, DeFi protocols advocate collateralizing their assets against conventional financial instruments. To do so, these protocols require access to external or off-chain data, such as asset exchange rates. DeFi protocols rely on oracles to access such information. Importing external data onto the chain via oracles consists of multiple data processing and aggregation stages. Thus, it is critical to minimize errors or deviations while the ground truth data moves through these stages. In this paper, we investigate the degree of price deviations at different levels between the data source and the final output rendered to an on-chain requester. In particular, we focus on Chainlink's oracle network for ETH-USD pricing. Our results show that despite checks and balances, the output rendered to the requester considerably deviates from data reported by the sources.

Keywords: Blockchain · DeFi · Oracle · Price deviations

1 Introduction

Blockchain-based systems, e.g., cryptocurrencies, eliminate the need for trusted intermediaries in transactions and make financial tools universally accessible. Decentralized Finance (DeFi) aims to further extend the concept of open financial systems. DeFi brings traditional as well as novel financial tools to the blockchain via smart contracts. By using blockchain as a building block, DeFi naturally inherits openness, decentralization, and censorship-resistance properties. Such properties are difficult to incorporate in the design of our traditional financial systems. Furthermore, DeFi advocates for interoperability and modular (Lego-like) design that enables component import from other DeFi products [23].

Like any other blockchain-based system, a DeFi protocol does not have direct access to external or off-chain data. In other words, they can only operate over the data that is already available or fed into the blockchain. Any interactions with the external world would require an intermediary. It is critical for DeFi

A. R. Beresford et al. (Eds.): CANS 2022, LNCS 13641, pp. 329–338, 2022.
https://doi.org/10.1007/978-3-031-20974-1_17

protocols to have near real-time access to market prices of both crypto and non-crypto assets. It is so because these assets are used as collateral in DeFi while their value is ever-changing or even volatile. To get access to real-world information, DeFi uses *oracles* [16]. An oracle is a data reporting infrastructure that acts as a bridge between DeFi and assets' off-chain price sources. With the increasing popularity and adaption of DeFi, the role of oracles is becoming more critical [13]. Nowadays, a wide variety of applications use oracles, e.g., synthetic assets [21], automated market maker [25], derivatives [6], non-fungible tokens [8].

The introduction of third-party data feeders, i.e., oracles, in DeFi is orthogonal to the fundamental concepts of blockchain decentralization. By using the services of an oracle, a DeFi protocol puts trust in a third-party intermediary. However, such a loss of decentralization is inevitable for a DeFi protocol that essentially need off-chain data for executing its core functionalities. This dilemma is known as *the oracle problem* [11]. A poorly managed or maliciously reporting oracle can put investors' funds to risk. Oracle-related vulnerabilities [12,22] have enabled a number of DeFi hacks [20]. A recent study [14] highlights that about two-third of DeFi hacks were results of oracle exploitation. By September 2021, the total value locked in DeFi projects was over 100 billion USD [23]. Therefore, it is important to understand and minimize the risks posed by oracles.

DeFi platforms naturally tend to use multiple oracles to minimize centralization and dependence on a single oracle. Different oracles can have dissimilar frequency of data updates. To further complicate the scenario, prices values reported by different oracles may differ from each other due to their distinct data sources. So, an aggregator node is needed to resolve disputes and aggregate price-values from multiple oracles. Building a price-value reporting infrastructure from scratch can be time consuming. Thus, to speed up the development of their core functionalities, several DeFi platforms choose to use pricing services from other platforms, such as Chainlink[1]. Chainlink is an important player in the DeFi ecosystem that offers a variety of services; price feeding is one such service. Chainlink runs a network of decentralized oracles and renders off-chain price-values to on-chain services. A multitude of DeFi projects, including several leading projects (e.g., Aave, Ampleforth), integrate[2] with Chainlink. As far as the projects with a use case of an oracle are concerned, we found that the majority of top-30 [10] DeFi projects by market capitalization rely on Chainlink.

Motivations: Irrespective of the type (i.e., first-party or third-party), any price-value reporting infrastructure in DeFi involves multiple entities (e.g., data source, oracle, aggregator) between the ground truth and a data requester. Thus, there exist different sources of errors in such a reporting infrastructure. Both over-reported and under-reported data have their own adverse effects. Over-reporting may induce customer churn while under-reporting may lead to platform exploitation. One of the key responsibilities of an aggregator in the infrastructure is to minimize the price deviations before rendering the data to the requester. To this end, an aggregator incorporates different measures (e.g., statistical mean, mode, median, time-weighted average) [13,16]. Hence, it is crucial

[1] https://chain.link/.
[2] https://chainlinkecosystem.com/ecosystem.

to understand the effectiveness of such measures. In this study, we focus on Chainlink's oracle network [7] for ETH-USD [9], which is one of the most common price-value pairs in the DeFi as well as the blockchain ecosystem.

Contributions: The major contributions of this paper are as follows: (1) we empirically evaluate the effectiveness of the aggregator node in the process of minimizing price deviations; (2) we investigate the degree of price deviations at different levels between the ground truth and the final output rendered to a requester; and (3) we make our collected data available[3] in the spirit of reproducible research.

Organization: The remainder of this paper is organized as follows. Section 2 presents a summary of the related works. Section 3 elucidates the price deviation issue in DeFi oracles. Section 4 elaborates our evaluation methodology and assessment of the price deviations in Chainlink's oracle network. Finally, Sect. 5 concludes the paper and outlines the directions for future works.

2 Related Works

In this section, we discuss the state of the art directly related to our work. DeFi platforms execute in the on-chain world, and depending on their target applications, such platforms typically require asset data from the off-chain world (e.g., the Internet). An oracle is the bridge between these two worlds. The work in [11] discusses the oracle problem and explains how an oracle can undermine the decentralization of blockchains. The authors in [2,4,26] survey the existing oracle implementations and define different types of interactions that happen between the on-chain and off-chain elements. The works in [18,19] discuss best practices and common design patterns for oracles. Eskandari et al. [13] further classify the existing oracles using a modular framework. The works in [1,12] propose decentralization of blockchain oracles. Similarly, Breidenbach et al. [7] discuss a multilayered approach for oracle's decentralization. Berger et al. [5] present an architecture for secure oracle usage. Lo et al. [17] present an approach to evaluate the reliability of oracles. Their evaluation approach uses the probability of failure rate to rank an oracle. The work in [3] explores the possibilities of using automated market maker projects, such as Uniswap[4], as price-value feeding oracles. Williams et al. [24] use game theory to understand the conditions for generating incentives over valid oracle queries for both cooperative and non-cooperative participants. Kaleem et al. [15] analyze usage trends, oracle pricing, and service quality associated with ChainLink on the Ethereum network. Liu et al. [16] discuss the technical architectures of mainstream DeFi platforms and investigate the deviations between external market prices and price-values retrieved from commonly used oracles. Different from the state of the art, our work explores the degree of price deviations at different levels between the ground truth and the final values provided to a data requester.

[3] https://github.com/CiaoAnkit/DeFiOraclesPriceDeviations.
[4] https://uniswap.org/.

3 Price Deviations in DeFi Oracles

A typical infrastructure that retrieves data from the off-chain world to the on-chain world involves several active entities, where each type of entity performs a specific set of actions. Figure 1 depicts an overview of retrieving the off-chain ground truth data for an on-chain data requester. While some implementations may combine distinct entities in practice (e.g., sources with feeders), we show each entity type separately. To remain in line with the goals of this study, we refrain from diving into the internal mechanics (for instance, polling frequency, selection mechanism, aggregation techniques) of different entities.

Ground truth is the data that DeFi oracles aim to fetch on-chain. *Data sources* reside closer to the ground truth. Sensors, humans, and APIs are some common types of data sources. These sources measure, congregate, and record a representation of the ground truth. Depending on the configuration, a data source may only be capable of storing the observed data momentarily. *Data feeders* interface with the sources to collect observed data. It is worth mentioning that a data feeder may connect with one or more sources to avoid failures, increase reliability, minimize the impact of a faulty/malicious source, etc. On the other side, data feeders provide their accumulated data to an on-chain *data aggregator*. In fact, data feeders are the bridge between off-chain and on-chain world and thus, are called oracles in the DeFi terminology. An aggregator generally selects and contracts with multiple oracles for a reliable data stream. Next, the aggregator aggregates the data (using statistical measures, majority voting, etc.) from multiple oracles and resolves any dispute before presenting the data to *requester*.

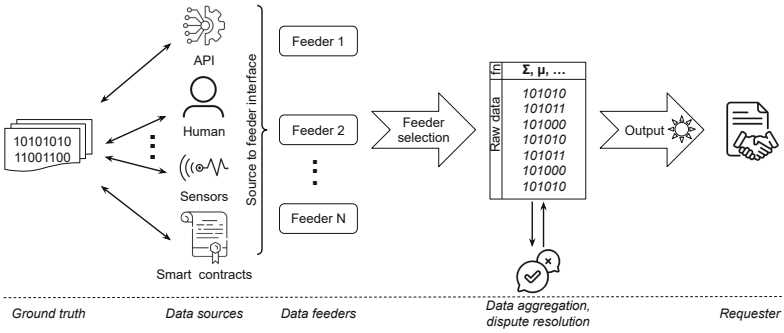

Fig. 1. An abstract view of fetching off-chain data for an on-chain requester.

When the ground truth data is in transit towards the requester, we can observe the data-in-transit at three different spots. The first observable data is reported by the data sources (cf. ① in Fig. 2), the second observable data comes from oracles (cf. ② in Fig. 2), and the third observable data is the final output of the aggregator (cf. ③ in Fig. 2). Ideally, the ground truth data should experience negligible alterations in the transit. However, each data processor (i.e., sources,

oracles, and aggregator) processes the data coming from previous stages according to its own logic. More importantly, oracles and aggregators process data coming from multiple sources and multiple oracles, respectively. Consequently, the data-in-transit suffers variations.

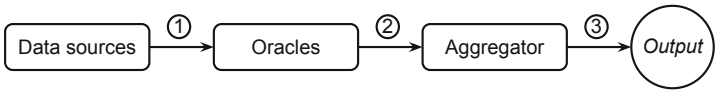

Fig. 2. The typical hierarchy of data reporting stack in DeFi.

Each data processor should intuitively incorporate mechanisms that minimize the price deviations. Nevertheless, it is ultimately the responsibility of an aggregator to choose (many) reliable oracles that use trustworthy data sources. In this paper, we study the price deviations at different stages between the data sources and the output of the aggregator. In particular, we analyze and compare the price-value data reported at stages ①, ②, and ③ indicated in Fig. 2. Furthermore, we want to understand how effective is the aggregator's aggregation of oracle data, i.e., if and by how much the output of the aggregator deviates from the data reported by the sources.

4 Evaluation

In this section, we present our analysis of the price deviations in Chainlink's oracle network. Section 4.1 elaborates our data collection phase. We present our findings from the analysis of the collected data in Sect. 4.2.

4.1 Data Collection

To collect price-value data for our analysis, we first identified the nodes at various levels in Chainlink's oracle network for ETH-USD price feeds. At the time of our study, it consisted[5] of one aggregator node and thirty-one oracle nodes. We also identified the data source nodes used by these oracle nodes. Two (i.e., AmberData[6] and Kaiko[7]) out of the nine data source nodes either required subscription or were not responding consistently. Thus, we continued with AlphaVantage[8], BraveNewCoin[9], CoinCap[10], CoinGecko[11], CoinMarketCap[12], Coin-

[5] https://data.chain.link/ethereum/mainnet/crypto-usd/eth-usd.
[6] https://web3api.io/api/v2/market/prices/eth/latest.
[7] http://us.market-api.kaiko.io/v1/data/trades.v1/spot_direct_exchange_rate/eth/usd.
[8] https://alphavantage.co/query.
[9] http://bravenewcoin.p.rapidapi.com/market-cap.
[10] https://api.coincap.io/v2/rates/ethereum.
[11] https://api.coingecko.com/api/v3/coins/ethereum/market/_chart/range.
[12] https://pro-api.coinmarketcap.com/v2/cryptocurrency/quotes/latest.

Paprika[13], CryptoCompare[14]. The data for the aggregator node and each of the oracle nodes was fetched[15] using a public contract address[16] while the data for source nodes was retrieved via their respective open APIs mentioned above.

We used automated python scripts to collect live data from April 17, 2022, 10:00 AM to May 23, 2022, 07:45 AM. We configured our scripts to execute once every fifteen minutes. While processing the collected data, we replaced any missing value with the mean of the corresponding values of other nodes in the same class. For cases when all the nodes in a given class had missing values, we discarded all the values collected in that round completely. Our final database has a total of 3399 rows of values and 40 columns (i.e., one aggregator node, thirty-one oracle nodes, seven data sources, and one column for timestamps).

4.2 Analysis of Price Deviations

To analyze the price deviations, we process our database of collected values row-by-row. For each row, we compute the average of oracle values (μ^O), maximum of oracle values (max^O), minimum of oracle values (min^O), average of source values (μ^S), maximum of source values (max^S), and minimum of source values (min^S). The aggregator node's values (v^A) are used as-is. Since the oracle nodes depend on the values from source nodes, μ^O should remain between max^S and min^S. Similarly, the aggregator node depends on the values from oracle nodes, v^A should remain between max^O and min^O. However, the true price deviations can be determined by comparing v^A against max^S and min^S. In what follows, we present our findings on the price deviations at different levels.

Figure 3 depicts the instances when μ^O exceeded max^S or fell below min^S. The positive values show μ^O-max^S when μ^O exceeded max^S while the negative values show μ^O-min^S when μ^O fell below min^S. A zero value denotes that μ^O was between max^S and min^S. Our analysis shows that μ^O exceeded max^S or fell below min^S on 35.36% of occasions, and the range of deviations was from –32.15 USD to +9.76 USD per ETH.

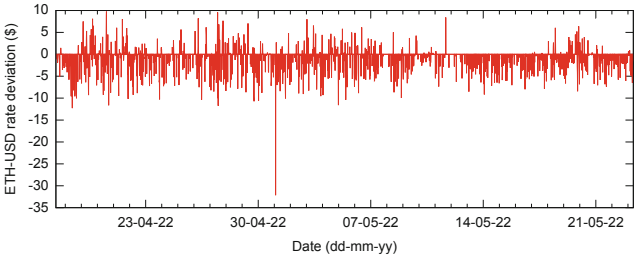

Fig. 3. The instances when μ^O exceeded max^S or fell below min^S. For such cases, positive values show μ^O-max^S while negative values show μ^O-min^S.

[13] https://api.coinpaprika.com/v1/tickers/eth-ethereum.
[14] https://min-api.cryptocompare.com/data/price.
[15] https://api.reputation.link/contract/.
[16] 0x37bc7498f4ff12c19678ee8fe19d713b87f6a9e6.

Next, Fig. 4 shows the instances when v^A exceeded max^O or fell below min^O. The positive values show v^A-max^O when v^A exceeded max^O while the negative values show v^A-min^O when v^A fell below min^O. A zero value denotes that v^A was between max^O and min^O. We find that v^A exceeded max^O or fell below min^O on only 3.09% of occasions. It appears that the aggregator's aggregation mechanism significantly controls the high number of deviations observed in oracles' output against data sources (cf. Fig. 3). However, the range of deviations in v^A further broadened and was from –37.22 USD to +17.36 USD per ETH.

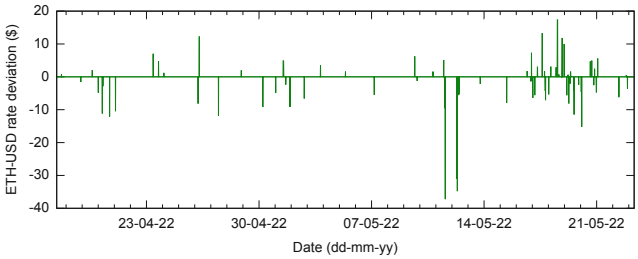

Fig. 4. The instances when v^A exceeded max^O or fell below min^O. For such cases, positive values show v^A-max^O while negative values show v^A-min^O.

To understand the overall impact of the aggregator's aggregation of oracles' output, we compare v^A against max^S and min^S. Figure 5 exhibits the instances when v^A exceeded max^S or fell below min^S. Again, the positive values show v^A-max^S when v^A exceeded max^S while the negative values show v^A-min^S when v^A fell below min^S. A zero value denotes that v^A was between max^S and min^S. Surprisingly, v^A exceeded max^S or fell below min^S on 38.36% of occasions, which is even more than the deviations observed in oracles' output against data sources (cf. Fig. 3). At the same time, the range of deviations was also astonishingly high, i.e., from –75.31 USD to +21.21 USD per ETH.

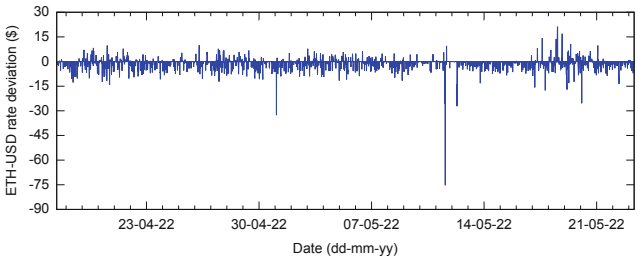

Fig. 5. The instances when v^A exceeded max^S or fell below min^S. For such cases, positive values show v^A-max^S while negative values show v^A-min^S.

As a further analysis, we compare v^A against μ^S. The plot in Fig. 6 illustrates v^A-μ^S. Here, 33.39% of values experienced a deviation of at least ±5 USD per ETH, and about 99% of values deviated within ±20 USD per ETH. Nevertheless, the overall deviation range was from –91.35 USD to +29.12 USD per ETH.

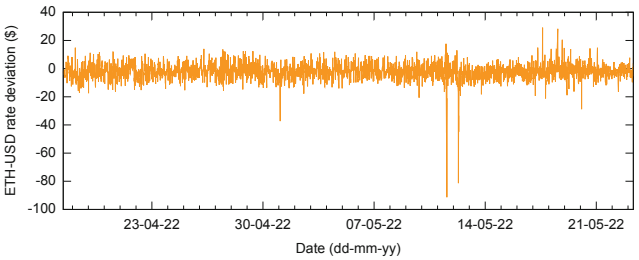

Fig. 6. The deviation of v^A from μ^S, the plot shows v^A-μ^S.

Based on our study, we infer mainly two reasons for the wide deviations between the output of the aggregator and the ground truth data reported by the sources. First, the aggregator is exposed to only oracle nodes. Second, each oracle node may use a distinct set of source nodes. Consequently, even a few faulty/malicious data sources/oracles can significantly affect the aggregator's output. Therefore, the aggregator must select reliable oracles and use robust mechanisms for data aggregation.

5 Conclusion and Future Works

DeFi platforms, like any other blockchain-based system, execute with on-chain data. Any interaction with the off-chain world requires an intermediary. DeFi oracles act as the bridge between these two worlds. However, oracles themselves rely on data sources. Thus, an aggregator is engaged to collect and sanitize data from multiple oracles before rendering it to an on-chain requester. This process of fetching the ground truth data onto the chain can induce variations in data while it is in transit. Understanding such price deviations is crucial because both over-reporting and under-reporting can lead to undesirable situations. In this paper, we empirically studied Chainlink's oracle network. In particular, we investigated the price deviations at different stages between the data sources and the output of the aggregator. Our results show that the output of the aggregator considerably deviates from the data reported by the data sources. In the future, we will extend our study to cover other publicly accessible oracle networks. We will also explore the possibilities to minimize these price deviations.

References

1. Adler, J., Berryhill, R., Veneris, A., Poulos, Z., Veira, N., Kastania, A.: Astraea: a decentralized blockchain oracle. In: IEEE International Conference on iThings, GreenCom, CPSCom, SmartData, pp. 1145–1152 (2018)

2. Al-Breiki, H., Rehman, M.H.U., Salah, K., Svetinovic, D.: Trustworthy blockchain oracles: review, comparison, and open research challenges. IEEE Access **8**, 85675–85685 (2020)
3. Angeris, G., Chitra, T.: Improved price oracles: constant function market makers. In: ACM Conference on Advances in Financial Technologies, pp. 80–91 (2020)
4. Beniiche, A.: A study of blockchain oracles. arXiv preprint:2004.07140 (2020)
5. Berger, B., Huber, S., Pfeifhofer, S.: OraclesLink: an architecture for secure oracle usage. In: IEEE International Conference on Blockchain Computing and Applications, pp. 66–72 (2020)
6. Biryukov, A., Khovratovich, D., Tikhomirov, S.: Findel: secure derivative contracts for ethereum. In: International Conference on Financial Cryptography and Data Security, pp. 453–467 (2017)
7. Breidenbach, L., et al.: Chainlink 2.0: next steps in the evolution of decentralized oracle networks. White Paper, pp. 1–136 (2021)
8. Chainlink: 16 Ways to Create Dynamic Non-Fungible Tokens (NFT) Using Chainlink Oracles (2020). https://blog.chain.link/create-dynamic-nfts-using-chainlink-oracles/
9. Chainlink: ETH-USD data feeds (2022). https://data.chain.link/ethereum/mainnet/crypto-usd/eth-usd
10. CoinMarketCap: Top DeFi Tokens by Market Capitalization (2022). https://coinmarketcap.com/view/defi/
11. Egberts, A.: The Oracle Problem - An Analysis of How Blockchain Oracles Undermine the Advantages of Decentralized Ledger Systems. Available at SSRN 3382343 (2017)
12. Ellis, S., Juels, A., Nazarov, S.: Chainlink: a decentralized oracle network. White Paper, pp. 1–38 (2017)
13. Eskandari, S., Salehi, M., Gu, W.C., Clark, J.: SoK: oracles from the ground truth to market manipulation. In: 3rd ACM Conference on Advances in Financial Technologies, pp. 127–141 (2021)
14. Finance Magnates: DeFi Startup Acala to Restructure Oracle Network - For the Better (2020). https://www.financemagnates.com/thought-leadership/defi-startup-acala-to-restructure-oracle-network-for-the-better/
15. Kaleem, M., Shi, W.: Demystifying pythia: a survey of chainlink oracles usage on ethereum. In: International Conference on Financial Cryptography and Data Security, pp. 115–123 (2021)
16. Liu, B., Szalachowski, P., Zhou, J.: A first look into DeFi oracles. In: IEEE International Conference on Decentralized Applications and Infrastructures, pp. 39–48 (2021)
17. Lo, S.K., Xu, X., Staples, M., Yao, L.: Reliability analysis for blockchain oracles. Comput. Electr. Eng. **83**, 1–10 (2020)
18. Mühlberger, R., et al.: Foundational oracle patterns: connecting blockchain to the off-chain world. In: International Conference on Business Process Management, pp. 35–51 (2020)
19. Pasdar, A., Dong, Z., Lee, Y.C.: Blockchain Oracle Design Patterns. arXiv preprint:2106.09349 (2021)
20. Peaster, W.M.: Biggest DeFi hacks in 2020 (2021). https://defiprime.com/hacks2020
21. Salehi, M., Clark, J., Mannan, M.: Red-black coins: dai without liquidations. In: International Conference on Financial Cryptography and Data Security, pp. 136–145 (2021)

22. samczsun: So You Want to Use a Price Oracle (2020). https://samczsun.com/so-you-want-to-use-a-price-oracle/
23. Werner, S.M., Perez, D., Gudgeon, L., Klages Mundt, A., Harz, D., Knottenbelt, W.J.: SoK: decentralized finance. arXiv preprint:2101.08778 (2021)
24. Williams, A.K., Peterson, J.: Decentralized common knowledge oracles. arXiv preprint:1912.01215 (2019)
25. Zaugust: CoFiX: A Computable Trading System (2020). https://github.com/Computable-Finance/Doc
26. Zhao, Y., Kang, X., Li, T., Chu, C.K., Wang, H.: Towards trustworthy DeFi oracles: past, present, and future. arXiv preprint:2201.02358 (2022)