Harvard John A. Paulson
School of Engineering
and Applied Sciences

I See You!

Robust Measurement of Adversarial Behavior

Multi-Agent Security Workshop @ NeurIPS 2023

Lars Ankile, Matheus V. X. Ferreira, David Parkes

AI and Multi-Agent Systems Are Evolving Rapidly

- Algorithms and AI's are everywhere – especially where fast decisions are rewarded, like financial markets
- As AIs get more numerous and sophisticated, it gets next to impossible to keep up
- FINRA has moved to using more complex methods as AIs tricked the standard, "hard-coded" rules [1]

FINRA

Artificial Intelligence (AI) in the Securities Industry[1]

JUNE 2020

Contents

A REPORT FROM THE FINANCIAL INDUSTRY REGULATORY AUTHORITY

Introduction

Artificial Intelligence (AI) technology is transforming the financial services industry across the globe. Financial institutions are allocating significant resources to exploring, developing, and deploying AI-based applications to offer innovative new products, increase revenues, cut costs, and improve customer service.[2] First developed

Can we develop non-manipulable measures of the level of manipulative behavior in a multi-agent system?

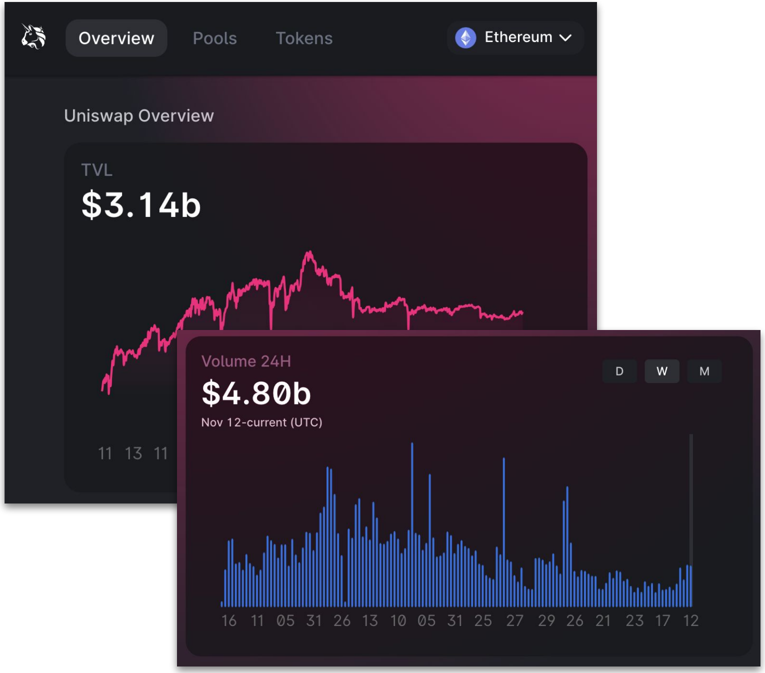
[1] FINRA, Jun 2022. URL https://www.finra.org/sites/default/files/2020-06/ai-report-061020.pdf.

The Blockchain as a Case-Study

- Permissionless and regulation free
- Easy to be anonymous and creating new identities (addresses) is virtually free
- Decentralized Exchanges process Billions of dollars of trading volume[1]
- The right to manipulate the market is institutionalized in an auction

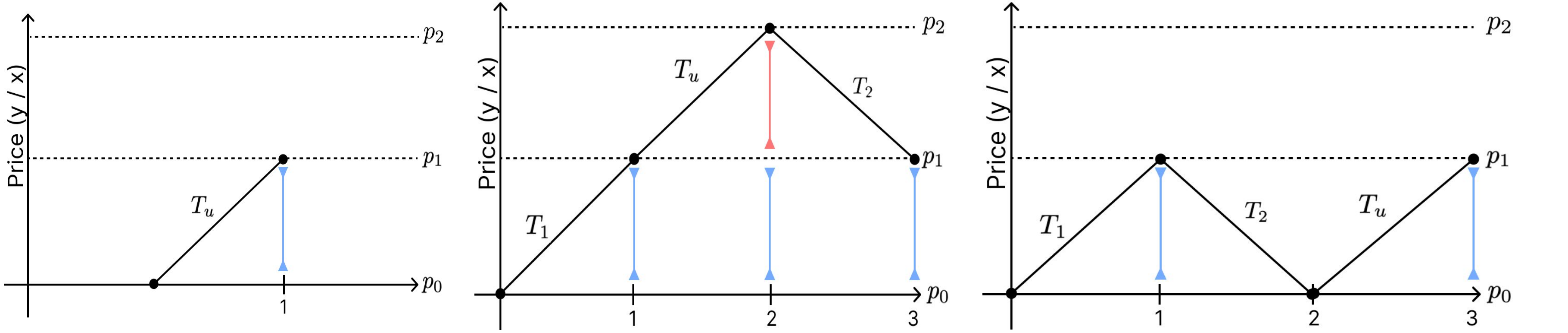
  → Big incentives for adversarial behavior
- Being distributed, there is a lot of data

  → Enabling experimentation



[1] https://info.uniswap.org/, accessed Nov 18, 2023

The Main Attack Class is Known as a Sandwich Attack



(a) Standalone execution

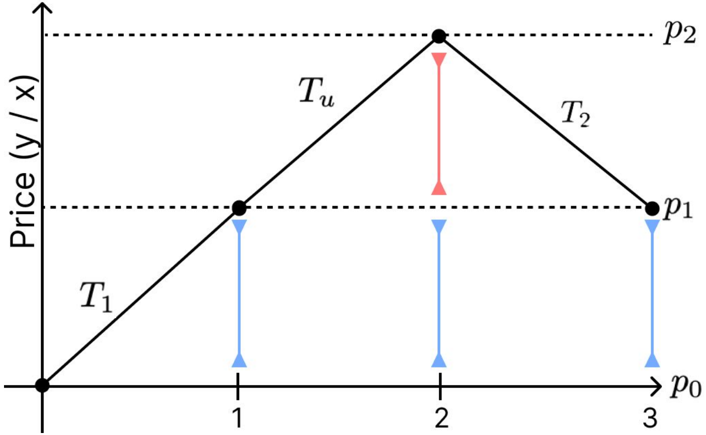
(b) Sequencer inserts transactions to create Sandwich

(c) Alternate ordering of the same set of transactions

The observation in (c) motivates the intuition behind our proposed metric

Most Current Methods Rely on Rules or Heuristics

- The standard approach “hardcodes” the rules of a sandwich attack, e.g.:
  - T_1 and T_2 comes from the same sender
  - T_1 and T_2 are in opposite directions but same size
- Simple strategies break the rules:
  - Create a new identity and send T_2 from that
  - Split T_2 into two halves

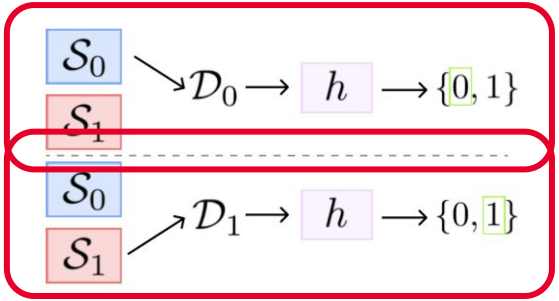


→ These tactics could be addressed, but AIs are likely to win cat-and-mouse game

See Appendix A in the paper for more detail on existing methods and obfuscation tactics

The DeFi Multi-Agent System as a Communication Game

- The communication game: Exchanges start with a state X_t
  - ***Traders*** submit transactions ($\textsc{Buy}(q,p)$/ $\textsc{Sell}(q,p)$) to a communication network
  - ***Sequencers*** connect to the network and observe sets of transactions $T = \{T_1, \ldots, T_n\}$ and outputs the order in which they will execute
  - The ***Exchange*** receives the transaction sequence and execute them in order $(T_{\sigma_1}, \ldots, T_{\sigma_n})$


$\mathcal{S}_0$
$\mathcal{S}_1$
$\mathcal{D}_0 \longrightarrow h \longrightarrow \{0,1\}$
$\mathcal{S}_0$
$\mathcal{S}_1$
$\mathcal{D}_1 \longrightarrow h \longrightarrow \{0,1\}$

- Malicious behavior includes: Message **injection**, **deletion**, and **reordering**

→ Goal is to detect which sequencers behave maliciously

We Propose a Surveillance Metric on Price Trajectories

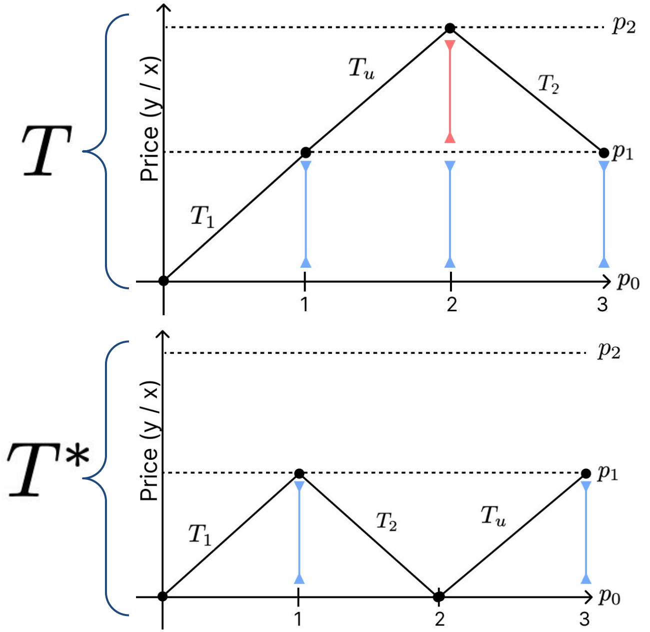
The p-surveillance metric for $p \geq 1$

$$S_p(T) = \left(\sum_{i=1}^{n} |p(T_{\leq i}) - p(\emptyset)|^p\right)^{\frac{1}{p}}$$

normalized surveillance metric

$$\bar{S}_p(T) := \frac{S_p(T)}{S_p(T^*)} - 1.$$

$$T^* \in \arg\min_{T'} \; S_p(T')$$



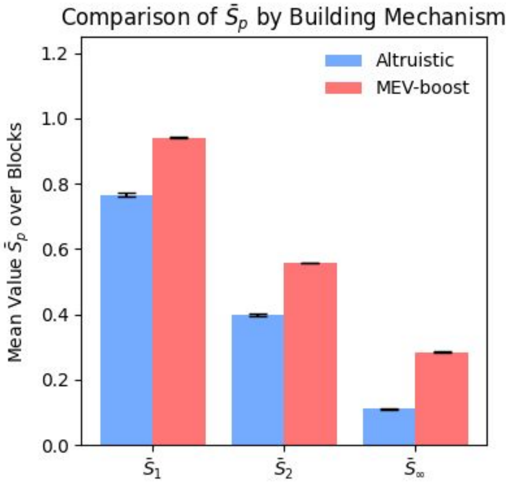
In practice, finding the optimal order is NP-hard [2], so we make an approximation, detailed in Appendix D

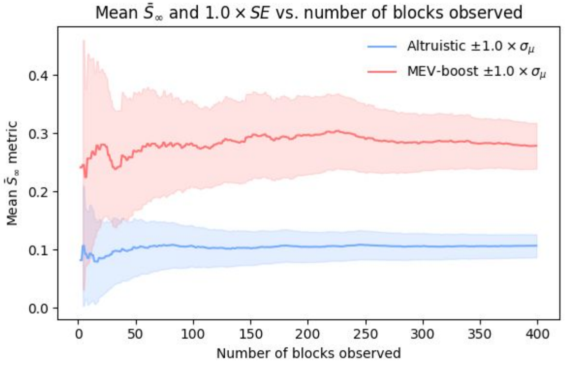
[2] Li, Yuhao, et al. "MEV Makes Everyone Happy under Greedy Sequencing Rule." *arXiv preprint arXiv:2309.12640* (2023).

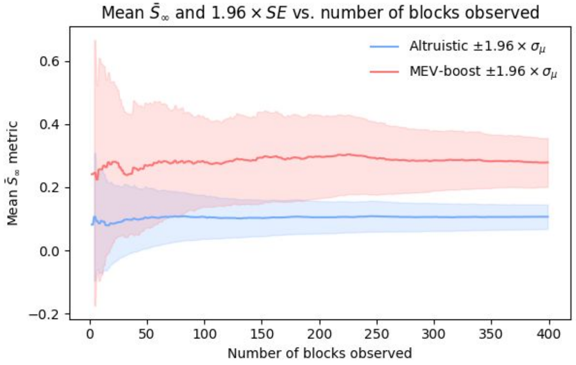
The Surveillance Metric Applied to Blockchain Data

When comparing bundles created by auction with standard bundles
→ we observe a significant difference in the surveillance metric…

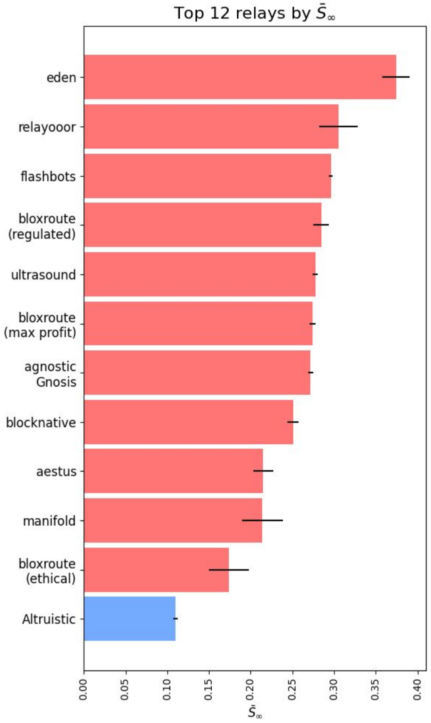
… and a relatively small number of observed bundles is needed to reach a reasonable level of confidence

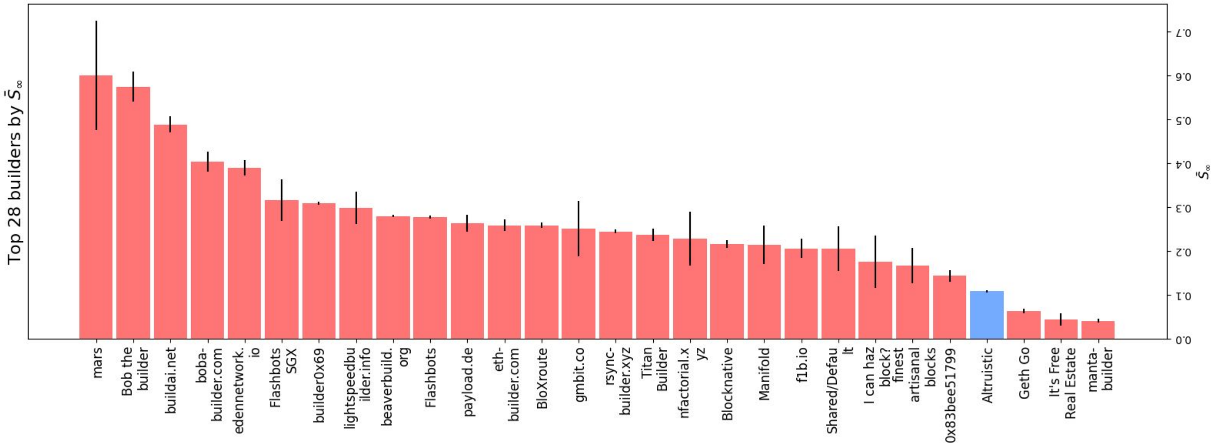

Comparison of $\bar{S}_p$ by Building Mechanism


Mean $\bar{S}_\infty$ and $1.0 \times SE$ vs. number of blocks observed


Mean $\bar{S}_\infty$ and $1.96 \times SE$ vs. number of blocks observed

More Detailed Analysis





See Appendix F for more analysis of the metric and data

Conclusion and Future Directions

- Empirical
  - Controlled experiment
  - Quantification of relationship between sequencer utility and metric
- Theoretical
  - Sufficient conditions for metric to be non-decreasing in adversary's utility
  - How to best define utility

Ankile, Lars, Matheus XV Ferreira, and David Parkes. "I See You! Robust Measurement of Adversarial Behavior." *Multi-Agent Security Workshop@ NeurIPS'23*. 2023.

Engage with code and data on the project GitHub:
https://github.com/ankile/defi-measurement

Get in touch:
larsankile@g.harvard.edu
(Applying for PhD positions this fall!)

matheus@seas.harvard.edu

References

[1] FINRA, “”Artificial Intelligence (AI) in the Securities Industry,” Jun 2022. URL https://www.finra.org/sites/default/files/2020-06/ai-report-061020.pdf.

[2] Li, Yuhao, et al. "MEV Makes Everyone Happy under Greedy Sequencing Rule." arXiv preprint arXiv:2309.12640 (2023).

The Blockchain Ecosystem is Riddled with Jargon

- **Blockchain:** A decentralized and distributed digital ledger that records transactions across multiple computers in a secure and immutable manner
- **Block:** A collection of transactions in a blockchain, digitally linked to preceding and succeeding blocks, creating a chronological chain
- **Sequencer:** An entity or mechanism in a blockchain network responsible for ordering transactions before they are added to the blockchain
- **DEX:** Decentralized Exchange, a type of cryptocurrency exchange without a central authority, enabling direct peer-to-peer cryptocurrency transactions
- **MEV:** Miner Extractable Value, the profit a miner can make through their ability to arbitrarily include, exclude, or reorder transactions within a block
- **MEV-Boost:** A mechanism that allows block builders to bid for the right to propose the blocks, aiming to decentralize the process of extracting MEV

Ethereum is the largest smart contract-enabled blockchain

Flashbots, the original creators of the MEV-boost mechanism is one of many companies operating in the space

Most Current Methods Rely on Rules or Heuristics

2020 IEEE Symposium on Security and Privacy

Flash Boys 2.0:
Frontrunning in Decentralized Exchanges, Miner
Extractable Value, and Consensus Instability

Philip Daian, Cornell Tech, phil@cs.cornell.edu
Steven Goldfeder, Cornell Tech, goldfeder@cornell.edu
Tyler Kell, Cornell Tech, sk3
Yunqi Li, UIUC
Xueyuan Zhao, CMU
Iddo Bentov, Cornell Tech, ib327@cornell.edu
Lore
lorenz.bre

Abstract—Blockchains, and specifically smart contracts, ha
promised to create fair and transparent trading ecosystems.
Unfortunately, we show that this promise has not been met.
document and quantify the widespread and rising deployment
arbitrage bots in blockchain systems, specifically in *decentrali*
exchanges (or "DEXes"). Like high-frequency traders on W
Street, these bots exploit inefficiencies in DEXes, paying hi
transaction fees and optimizing network latency to frontrun, i
anticipate and exploit, ordinary users' DEX trades.
We study the breadth of DEX arbitrage bots in a subset
transactions that yield quantifiable revenue to these bots. We a
study bots' profit-making strategies, with a focus on blockcha
specific elements. We observe bots engage in what we call *prior*
gas auctions (PGAs), competitively bidding up transaction fees
order to obtain priority ordering, i.e., early block position a
execution, for their transactions. PGAs present an interesti
and complex new continuous-time, partial-information, gam
theoretic model that we formalize and study. We release
interactive web portal, frontrun.me, to provide the commun
with real-time data on PGAs.
We additionally show that high fees paid for priority trans
tion ordering poses a systemic risk to *consensus-layer* securi
We explain that such fees are just one form of a general ph
nomenon in DEXes and beyond—what we call *miner extracta*
value (MEV)—that poses concrete, measurable, consensus-lay
security risks. We show empirically that MEV poses a realis
threat to Ethereum today.
Our work highlights the large, complex risks created
transaction-ordering dependencies in smart contracts and
ways in which traditional forms of financial-market exploitati
are adapting to and penetrating blockchain economies.

I. Introduction

Cryptocurrency exchanges today handle more than $
billion in trade volume per day. The vast majority of t
volume occurs in *centralized* exchanges, which hold custo
of customer assets and settle trades. At best loosely regulat
centralized exchanges have experienced scandals ranging fr
high-profile thefts [28] to malfeasance such as price m
nipulation [18]. One popular alternative is what is called

DOI 10.1109/SP40000.2020.00040

2021 IEEE Symposium on Security and Privacy (SP)

High-Frequency Trading on
Decentralized On-Chain Exchanges

Liyi Zhou *, Kaihua Qin *, Christof Ferreira Torres †, Duc V Le ‡ and Arthur Gervais *
*Imperial College London, United Kingdom
†University of Luxembourg, Luxembourg
‡Purdue University, United States

2021 IEEE Symposium on Security and Privacy (SP) | 978-1-7281-8934-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/SP40001.2021.00027

Abstract—Decentralized exchanges (DEXs) allow parties to participate in financial markets while retaining full custody of their funds. However, the transparency of blockchain-based DEX in combination with the latency for transactions to be processed, makes market-manipulation feasible. For instance, adversaries could perform *front-running* — the practice of exploiting (typically non-public) information that may change the price of an asset for financial gain.

In this work we formalize, analytically exposit and empirically evaluate an augmented variant of front-running: *sandwich attacks*, which involve front- and back-running victim transactions on a blockchain-based DEX. We quantify the probability of an adversarial trader being able to undertake the attack, based on the relative positioning of a transaction within a blockchain block. We find that a single adversarial trader can earn a daily revenue of over several thousand USD when performing sandwich attacks on one particular DEX — Uniswap, an exchange with over 5M USD daily trading volume by June 2020. In addition to a single-adversary game, we simulate the outcome of sandwich attacks under multiple competing adversaries, to account for the real-world trading environment.

I. Introduction

Decades of asset trading on traditional exchanges have brought to fruition a veritable collection of market manipulation techniques, such as front-running [6], pump and dump schemes [54] and wash trading [3]. In the context of cryptocurrencies, research to date indicates that the ecosystem requires a greater awareness of such malpractices [54], [21], [45], and better exchange design [12] to prevent misbehavior. Most existing legislation does not regulate crypto-exchanges to the same degree as traditional exchanges — leaving ignorant traders open to exploitation by predatory practices, some of which is close to risk-free.

Decentralized exchanges (DEXs) allow traders to trade financial assets without giving up asset custody to a third party. Orders can be placed and matched in their entirety through immutable blockchain smart contracts, offering the possibility of censorship resistance, where orders cannot be modified prior and after execution[1].

Censorship-resis
reliance on an u
all attempted a
(P2P) network.
however, in co
deterministicall
undertake — a
of the trader's
This paper.
back-running[2],
chain DEX. To
to formalize and
sandwich, a pre
P2P network fo
squeeze it by pl
(i.e. front-run)
If the target tr
the price of an
before which b
order afterward
We restrict
(AMM) DEXs [
limit order book
istic nature whi
in our analysis.
rithmically perf
instant liquidity
for market part
of an AMM D
a total liquidity
75% market liq
trading volume
November 2018
empirically eva
quantify optima
world empirical
study the proba
relative position
prospects for su
world scenario i

Disclosure: Arthur Gervais works on the Liquidity Network, a community-driven, open source layer-2 blockchain scaling solution.

[1] DEX prevent anyone from censoring trades, even the exchange itself. Censorship-resistance is a key property of permissionless blockchains such as Bitcoin and Ethereum, where users remain custodian over their assets at any point in time.

[2] While the SEC
information, we on

[3] The process of
and sell an asset.

DOI 10.1109/SP40001.2021.00027

Frontrunner Jones and the Raiders of the Dark Forest:
An Empirical Study of Frontrunning on the
Ethereum Blockchain

ARTIFACT EVALUATED
usenix ASSOCIATION
PASSED

Christof Ferreira Torres
SnT, University of Luxembourg

Ramiro Camino
Luxembourg Institute of Science and Technology

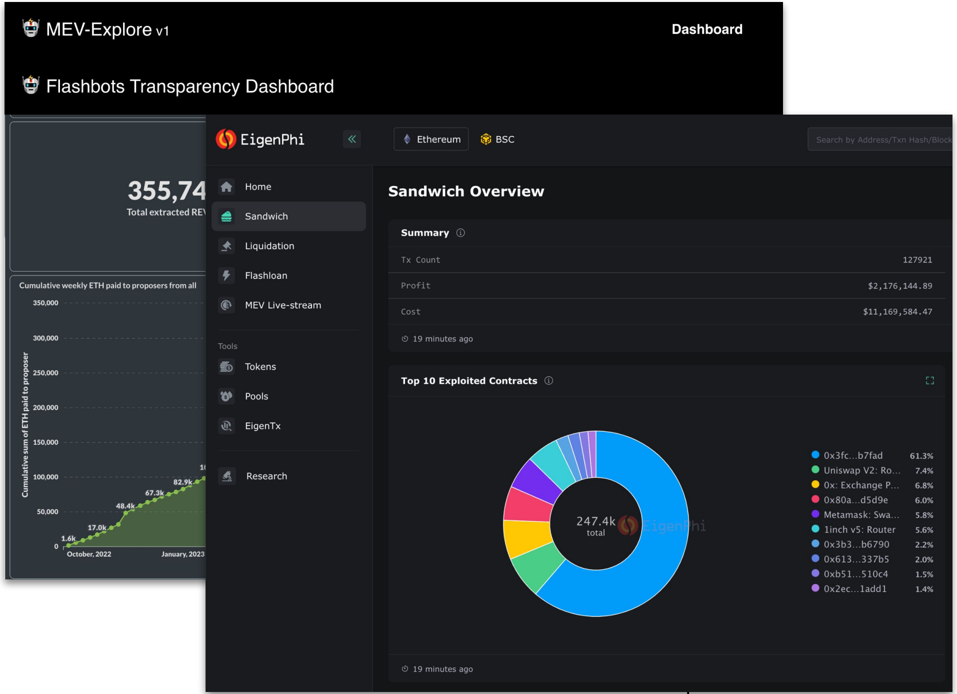
Radu State
SnT, University of Luxembourg

Abstract

Ethereum prospered the inception of a plethora of smart contract applications, ranging from gambling games to decentralized finance. However, Ethereum is also considered a highly adversarial environment, where vulnerable smart contracts will eventually be exploited. Recently, Ethereum's pool of pending transaction has become a far more aggressive environment. In the hope of making some profit, attackers continuously monitor the transaction pool and try to frontrun their victims' transactions by either displacing or suppressing them, or strategically inserting their transactions. This paper aims to shed some light into what is known as a *dark forest* and uncover these predators' actions. We present a methodology to efficiently measure the three types of frontrunning: *displacement*, *insertion*, and *suppression*. We perform a large-scale analysis on more than 11M blocks and identify almost 200K attacks with an accumulated profit of 18.41M USD for the attackers, providing evidence that frontrunning is both, lucrative and a prevalent issue.

1 Introduction

The concept of frontrunning is not new. In financial markets, brokers act as intermediaries between clients and the market, and thus brokers have an advantage in terms of insider knowledge about potential future buy/sell orders which can impact the market. In this context, frontrunning is executed by prioritizing a broker's trading actions before executing the client's orders such that the trader pockets a profit. Frontrunning is illegal in regulated financial markets. However, the recent revolution enabled by decentralized finance (DeFi), where smart contracts and miners replace intermediaries (brokers) is both, a blessing and a curse. Removing trusted intermediaries can streamline finance and substantially lower adjacent costs, but misaligned incentives for miners leads to generalized frontrunning, in which market participants behave exactly like unethical brokers used to in the "old" financial world. Unfortunately, this is already happening at a large scale. Our paper is among the first comprehensive surveys on the extent and impact of this phenomenon. Already in 2017, the Bancor ICO [11] was susceptible to such an attack – among other vulnerabilities – but no real attack was observed in the wild. Some concrete frontrunning attacks on the Ethereum blockchain were brought to knowledge by two independently reported attacks and their mitigation approaches to the informed audience. In the first report [10], the researchers tried to recover some liquidity tokens by calling a specific function in a smart contract. Since this function was callable by everyone, the authors – who also compared the pending transactions in the transaction pool to a *dark forest* full of predators – assumed that their function call could be observed and frontrunned by bots observing the submitted transactions in the transaction pool. Even though they tried to obfuscate their efforts, their approach failed in the end, and they became a victim of a frontrunning bot. A few months later, a second group of researchers [24] reported a successful recovery using lessons learned from the previously reported incident [10]. The success was due to them mining their transactions privately without broadcasting them to the rest of the network. The researchers used a new functionality provided by SparkPool called the Taichi Network [15]. In this way, the transactions were not available to frontrunning bots but relied entirely on having a reliable and honest mining pool. However, this approach enables centralization and requires users to entrust their transactions to SparkPool. Similar to how honeypots gather intelligence by luring attackers to compromise apparently vulnerable hosts [8], a recent experiment [21] detailed the interactions with two bots and reported relevant assessment on their nature and origin. Surprisingly, the frontrunning bots do not rely on advanced software development techniques or complex instructions, and code examples on developing such bots are readily available [22, 23]. There are several ways to perform frontrunning attacks. The first survey defining a taxonomy of frontrunning attacks [12] identified three different variants on how these can be performed. To understand these approaches – *displacement*, *insertion*, and *suppression* – a short refresh on gas and transaction fees in



Sandwich Overview

Summary

Tx Count	127921
Profit	$2,176,144.89
Cost	$11,169,584.47

19 minutes ago

Top 10 Exploited Contracts

0x3fc...b7fad	61.3%
Uniswap V2: Ro...	7.4%
0x: Exchange P...	6.8%
0x80a...d5d9e	6.0%
Metamask: Swa...	5.8%
1inch v5: Router	5.6%
0x3b3...b6790	2.2%
0x613...337b5	2.0%
0xb51...510c4	1.5%
0x2ec...1add1	1.4%

19 minutes ago

Harvard John A. Paulson
School of Engineering
and Applied Sciences