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

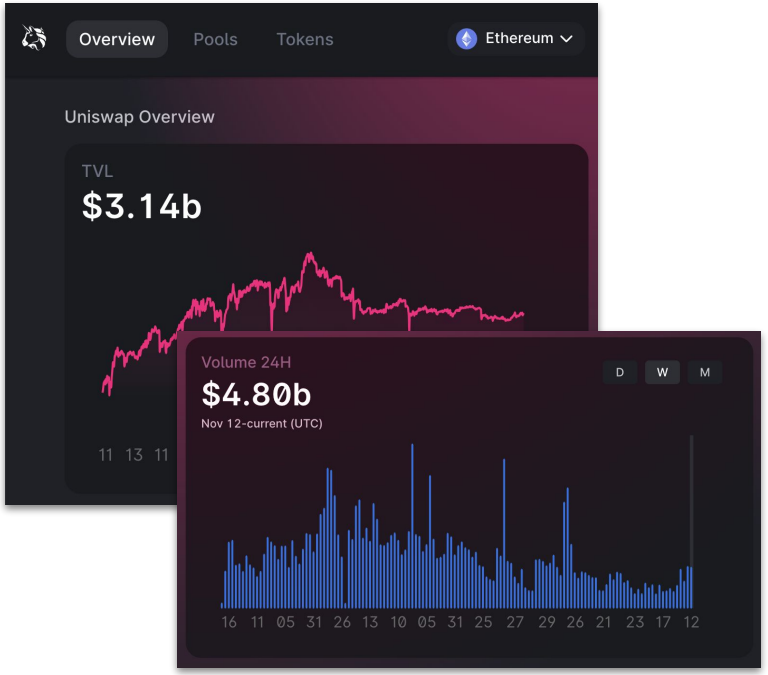
- Algorithms and AI's are everywhere – especially where fast decisions are rewarded, like financial markets
- As AIs get more numerous and sophisticated, it gets next to impossible to keep up
- FINRA has moved to using more complex methods as AIs tricked the standard, "hard-coded" rules [1]

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.

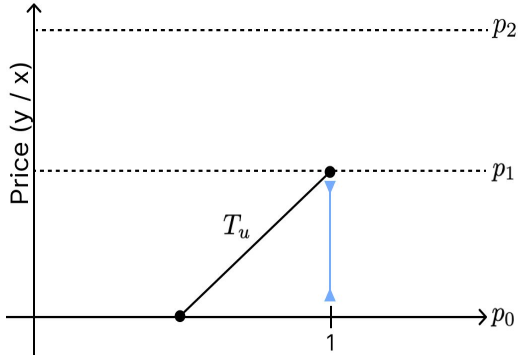
The Blockchain as a Case-Study

- 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

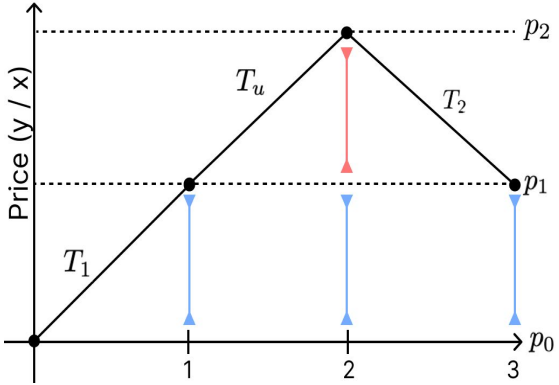


[1] https://info.uniswap.org/, accessed Nov 18, 2023

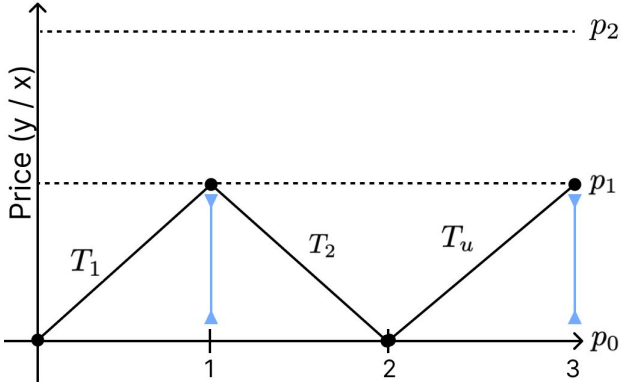
The Main Attack Class is Known as a Sandwich Attack



(a) Standalone execution



(b) Sequencer inserts transactions to create Sandwich

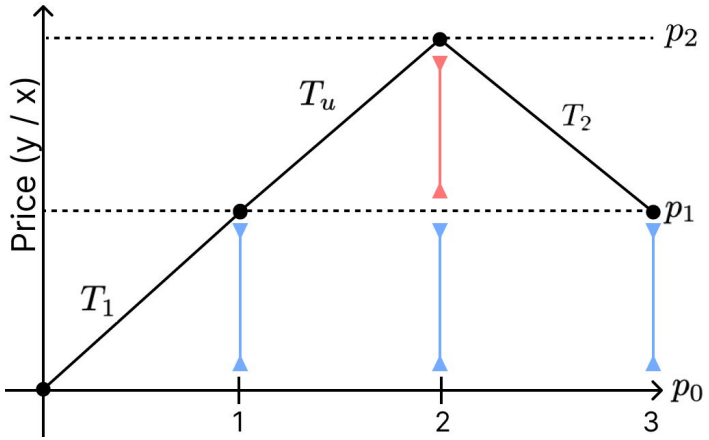


(c) Alternate ordering of the same set of transactions

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

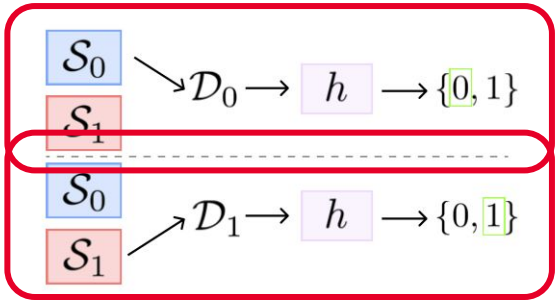


→ These tactics could be addressed, but AIs are likely to win cat-and-mouse game

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 $(T_{\sigma_1}, \ldots, T_{\sigma_n})$



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

  $\rightarrow$ Goal is to detect which sequencers behave maliciously

We Propose a Surveillance Metric on Price Trajectories

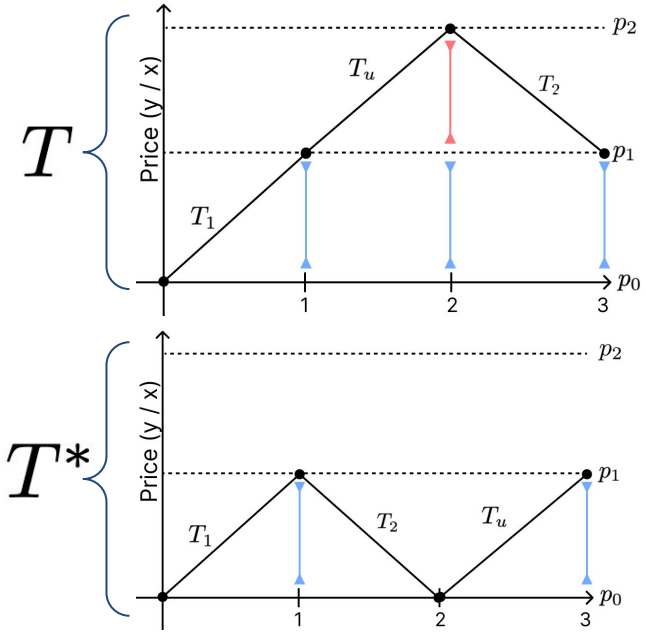
The p-surveillance metric for $p \geq 1$

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

normalized surveillance metric

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

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



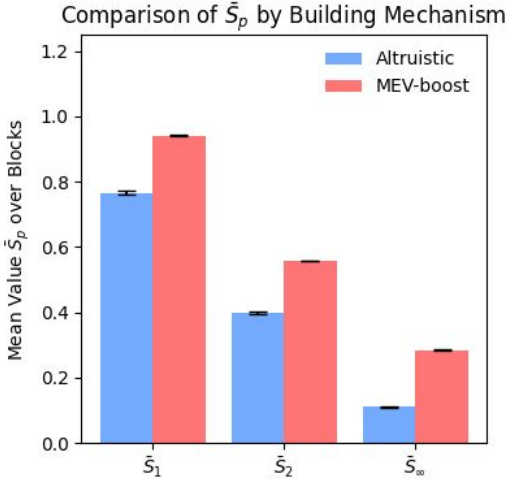
In practice, finding the optimal order is NP-hard [2], so we make an approximation, detailed in Appendix D

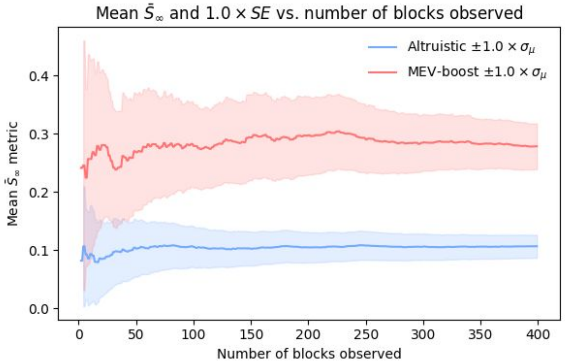
[2] Li, Yuhao, et al. "MEV Makes Everyone Happy under Greedy Sequencing Rule." *arXiv preprint arXiv:2309.12640* (2023).

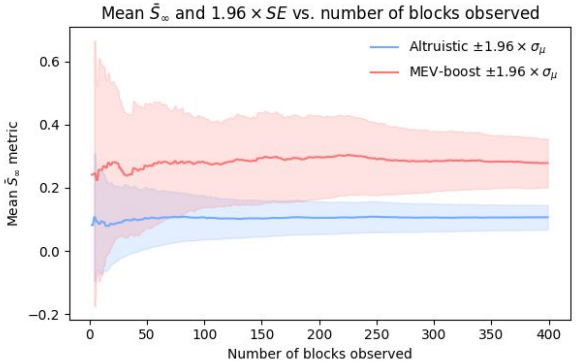
The Surveillance Metric Applied to Blockchain Data

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

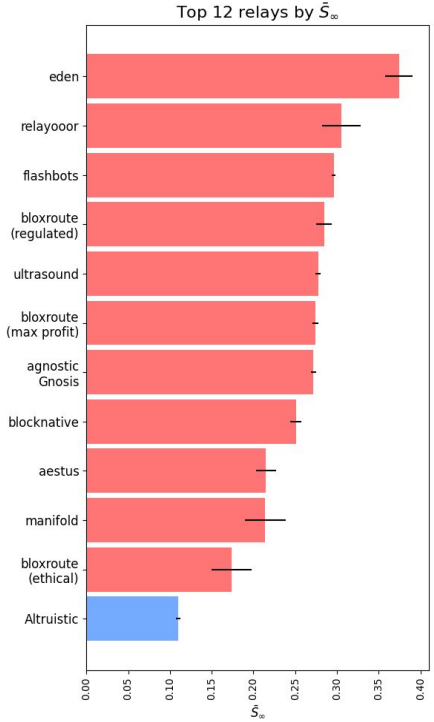
… and a relatively small number of observed bundles is needed to reach a reasonable level of confidence

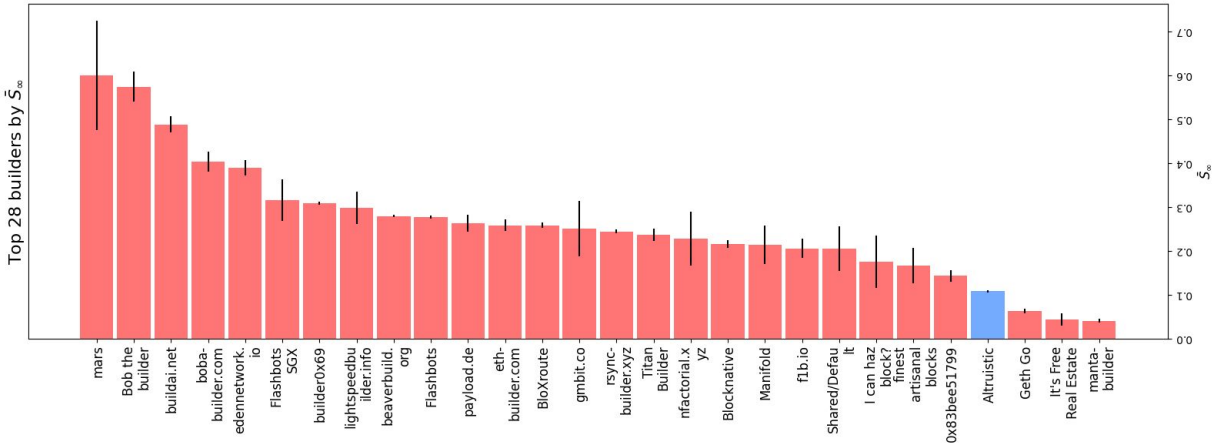






More Detailed Analysis





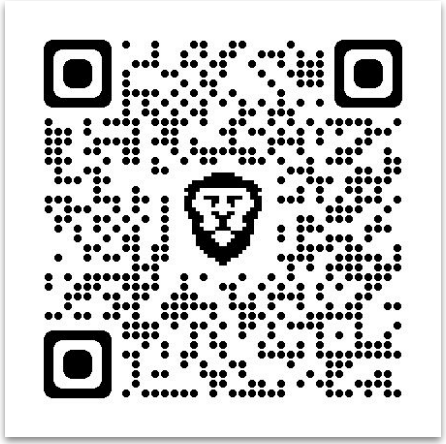
See Appendix F for more analysis of the metric and data

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).

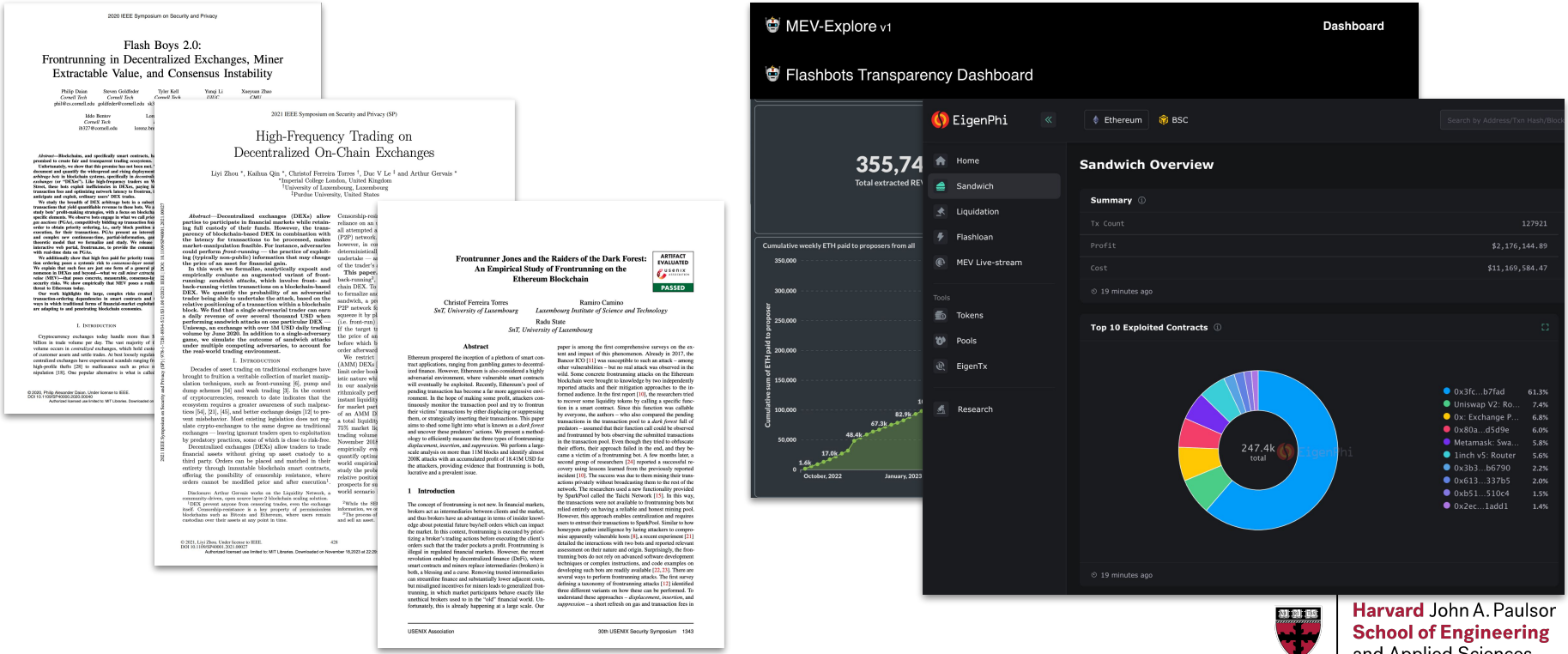
The Blockchain Ecosystem is Riddled with Jargon

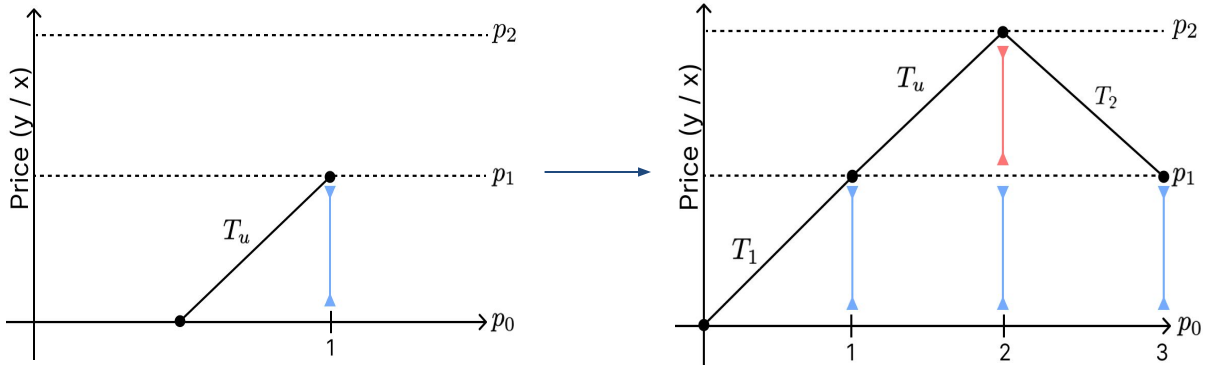
- **Blockchain:** A decentralized and distributed digital ledger that records transactions across multiple computers in a secure and immutable manner
- **Block:** A collection of transactions in a blockchain, digitally linked to preceding and succeeding blocks, creating a chronological chain
- **Sequencer:** 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




Price (y / x)
p_2
p_1
p_0
T_u
1
Price (y / x)
p_2
T_u
T_2
p_1
T_1
p_0
1
2
3