

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

Al and Multi-Agent Systems Are Evolving Rapidly

- Algorithms and Al's are everywhere especially where fast decisions are rewarded, like financial markets
- As Als 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?



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 volume¹
- The right to manipulate the market is institutionalized in an auction
 - \rightarrow Big incentives for adversarial behavior
- Being distributed, there is a lot of data
 → Enabling experimentation





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 T₂ into two halves



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



The DeFi Multi-Agent System as a Communication Game

- The communication game: Exchanges start with a state X_t
 - **Traders** submit transactions (|BUY(q, p)/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
$$p \geq 1$$

 $S_p(T) = \left(\sum_{i=1}^{n} |p(T_{\leq i}) - p(\emptyset)|^p\right)^{\frac{1}{p}}$
 $T = \frac{S_p(T)}{S_p(T^*)} - 1$
 $T^* \in \arg \min_{T'} S_p(T')$
 $T = \frac{T_1}{S_p(T')}$
 $T = \frac{T_1}{T_1}$
 $T = \frac{T_2}{T_1}$
 $T = \frac{T_2}{T_1}$
 $T = \frac{T_2}{T_1}$
 $T = \frac{T_2}{T_1}$
 $T = \frac{T_2}{T_1}$

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



 \mathbf{T}

The Surveillance Metric Applied to Blockchain Data

When comparing bundles created by auction with standard bundles \rightarrow we observe a significant difference in the surveillance metric...

Comparison of \overline{S}_p by Building Mechanism 1.2 1.0 0.8 0.6 0.4 0.4 0.2 \tilde{S}_1 \tilde{S}_2 \tilde{S}_{∞} ... and a relatively small number of observed bundles is needed to reach a reasonable level of confidence





More Detailed Analysis

Ŝ.





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: <u>larsankile@g.harvard.edu</u> (Applying for PhD positions this fall!)

matheus@seas.harvard.edu



References

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

[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





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

🗐 N

🔮 F

2020 IEEE Symposium	n on Security and Privacy				
Flash E	Boys 2.0:				
Frontrunning in Decent	ralized Exchanges, Miner				
Extractable Value, and	d Consensus Instability				
Bills Data Data California	New York Marcel 11 Marcel and Marcel				
Consell Tech Consell Tech Co	week Took UNDC CMU				
Ide Bases	2021 IEEE Symposium of	n Security and Priva	ev (SP)		
Cornell Tech					
	High-Frequen	cy Tradi	ng on		
	Decentralized On	-Chain F	Exchanges		
Abstract—Blockchaims, and specifically smart contracts, is promised to create fair and transparent trading ecosystems. Undertunation, see show that this revention has not been rest. ¹	Deconstantized on ontain Exchanges				
document and quantify the widespread and rhing deployment arbitrage bets in blockchain systems, specifically in decentrally	Liyi Zhou *, Kaihua Qin *, Christof Ferreira	a Torres [†] , Duc V	Le ⁴ and Arthur Gervais *		
Street, these hots exploit inefficiencies in DEXes, paying h transaction from and outlening network latence to Frontran.	imperant Golfge London, United Ringson University of Lonembourg, Lonembourg				
anticipate and exploit, ordinary users' DEX trades. We shally the hreadth of DEX arbitrage bots in a subset	Fundae Onivers	aty, childed Suites			
study bets' profit-making strategies, with a focus on blockcha specific elements. We observe bots engage in what we call prior	Abstract-Decentralized exchanges (DEXs) allow	Censorship-resi			
ger anclose (PGAs), competitively bidding up transaction fees order to obtain priority ordering, i.e., early block position a	parties to participate in financial markets while retain- ing full custody of their funds. However, the trans-	reliance on an t all attemnted a			
and complex new continuous time, partial information, gas theoretic model that we formalize and study. We release	5 parency of blockchain-based DEX in combination with the latency for transactions to be processed, makes	(P2P) network.			
interactive web portal, frontrumme, to provide the commun- with real-time data on FGAs.	g market-manputation beasities. For instance, adversaries could perform front-running — the practice of exploit- ion (amplitude and exploit- ion formation the set of the set	deterministicall	r	n	ARTIFACT
tion ordering pases a systemic risk to consensedury securi We explain that such fees are just one form of a general pl	in the price of an asset for financial gain. In this work we formalize, analytically exposit and	of the trader's a	An Empirical Study	e Raiders of the Dark Forest: of Frontrunning on the	EVALUATED
nomenon in DEXes and beyond—what we call minor contacts reliae (MEV)—that poses concrete, measurable, consensus-lag contacts this. We show consistent that MPV means are the	empirically evaluate an augmented variant of front- running: sandwick attacks, which involve front- and	This paper. back-running ² ,	Ethereun	Blockchain	G attaction
thread to Efforcement today. Our work highlights the large, complex risks created	back-running victim transactions on a blockchain-based DEX. We quantify the probability of an adversarial	chain DEX. To to formalize and			PASSED
tranaction-ordering dependencies in searct contracts and i ways in which traditional forms of financial-market exploited are admitted in and reservative block-bain consumics.	trader being able to undertake the attack, based on the relative positioning of a transaction within a blockchain	sandwich, a per P2P network fo	Christof Ferreira Torres	Ramiro Camino	
	a daily revenue of over several thousand USD when performing and with thicks on constraints DEX	squeeze it by pl	Shi, University of Laternoourg Laternoourg Rat	imbourg insulate of science and rech	www.
L INTRODUCTION	Universe, an exchange with over 5M USD daily trading volume by June 2020. In addition to a single-adversary	If the target to	SnT, Universi	y of Luxembourg	
billion in trade volume per day. The vari majority of I	game, we simulate the outcome of sandwich attacks under multiple competing adversaries, to account for	before which b	Abstract	rener is among the first commutensive	surveys on the ex-
of customer assets and sattle trades. At bost loosely regular controllined enclosures have encodenced scandals mention for	the real-world trading environment.	We restrict	Ethereum prospered the inception of a plethora of smart con-	tent and impact of this phenomenon. A Barrow ICO 1111 was supcertible to such	Iready in 2017, the
high-profile thefts [28] to malicasance such as price in nipulation [18]. One pepular alternative is what is called	Decades of asset trading on traditional exchanges have	(AMM) DEXs limit order book	tract applications, ranging from gambling games to decentral- ized finance. However, Ethereum is also considered a highly	other vulnerabilities - but no real attack	was observed in the
	ulation techniques, such as front-running [6], pump and	istic nature whi in our analysis	adversarial environment, where vulnerable smart contracts will eventually be exploited. Recently, Ethereum's pool of	blockchain were brought to knowledge by	two independently
0 2020, Philip Alexander Dalar, Under license to IEEE. DCI 10.1109/SP40000.2020.00040	f dump schemes [54] and wash trading [3]. In the context of cryptocurrencies, research to date indicates that the	rithmically perf instant liquidity	pending transaction has become a far more aggressive envi- ronment. In the hope of making some profit, attackers con-	formed audience. In the first report [10], 1	he researchers tried
Autor and realized on the or a first contrast. Contrasts of	ecosystem requires a greater awareness of such malprac- tices (54), [21], [45], and better exchange design [12] to pre-	for market part	tinuously monitor the transaction pool and try to frontrun their victims' transactions by either displacing or suppressing	tion in a smart contract. Since this fur	iction was callable
	went misbehavior. Most existing legislation does not reg-	a total liquidity	them, or strategically inserting their transactions. This paper aims to shed some light into what is known as a dark forest	transactions in the transaction pool to a	dark forest full of
	exchanges - leaving ignorant traders open to exploitation	trading volume	and uncover these predators' actions. We present a method- obase to efficiently measure the three tenes of freetostration	and frontrunned by bots observing the sul	mitted transactions
	Decentralized exchanges (DEXs) allow traders to trade	November 2018 empirically eva	displacement, insertion, and suppression. We perform a large-	in the transaction pool. Even though the their efforts, their approach failed in th	y tried to obfuscate e end, and they be-
	inancial assets without giving up asset custody to a third party. Orders can be placed and matched in their	quantify optime world empirical	200K attacks with an accumulated profit of 18.41M USD for	came a victim of a frontrunning bot. A second group of researchers [24] report	few months later, a ied a successful re-
	entirety through immutable blockchain smart contracts, offering the possibility of censorship resistance, where	study the probe	lucrative and a prevalent issue.	covery using lessons learned from the j incident [10]. The success was due to ther	reviously reported n mining their trans-
	orders cannot be modified prior and after execution ¹ .	prospects for su	1 Introduction	actions privately without broadcasting the network. The researchers used a new fur	em to the rest of the actionality provided
	Disclosure: Arthur Gervais works on the Liquidity Network, a community-driven, open source layer-2 blockchain scaling solution.	20154 the SE		by SparkPool called the Taichi Network the transactions were not available to fin	k [15]. In this way, intransing bots but
	itself. Censorship-resistance is a key property of permissionless blockchains such as Bitcoin and Ethereum, where users remain	information, we or ³ The process of	brokers act as intermediaries between clients and the market,	relied entirely on having a reliable and I However, this approach enables central	sonest mining pool. ration and maxims
	custodian over their assets at any point in time.	and sell an asset.	edge about potential future buy/sell orders which can impact	users to entrust their transactions to Spark	Pool. Similar to how
	di 1991 Lini Phone Hadar Bourne to 1920	~	the market. In this context, frontrunning is executed by priori- tizing a broker's trading actions before executing the client's	mise apparently vulnerable hosts [8], a rec	ent experiment [21]
	DOI 10.110/05P40001.3021.00027 Authorized licensed use limited to: MIT Libraries. Downloaded on Nov	ember 18,2023 at 22:29;	orders such that the trader pockets a profit. Frontrunning is illegal in regulated financial markets. However, the recent	assessment on their nature and origin. Su	rprisingly, the fron-
		_	revolution enabled by decentralized finance (DeFi), where smart contracts and miners replace intermediaries (brokers) is	techniques or complex instructions, and	code examples on
			both, a blessing and a curse. Removing trusted intermediaries can streamline finance and substantially lower adjacent costs,	several ways to perform frontranning atta	cks. The first survey
		_	but misaligned incentives for miners leads to generalized from- transing, in which market participants behave exactly like	defining a taxonomy of frontrunning att three different variants on how these ca	icks [12] identified a be performed. To
			unethical brokers used to in the "old" financial world. Un- fortunately, this is already happening at a large scale. Our	understand these approaches – displacer suppression – a short refresh on gas and	nent, sesertion, and transaction fees in
		_	······································		
			USENIX Association	30th USENIX Security	Symposium 1343
		_			

IEV-Explore v1		Da	shboard
lashbots Transpare	ency Dashboard		
	🊯 EigenPhi 🛛 🔍	🚯 Ethereum 👔 BSC	Search by Address/Txn Hash/Block
355,74	A Home	Sandwich Overview	
Total extracted REY weekly ETH paid to proposers from all	 Sandwich Liquidation Flashloan 	Summary ()	107001
			\$2,176,144.89
	MEV Live-stream		\$11,169,584.47
	Tools	© 19 minutes ago	
	💓 Pools	Top 10 Exploited Contracts ()	
82.34 67.34	 €igenTx ▲ Research 	247.4k	 0.37cb7fad 61.3% Uniway V2; Ro 7.4% 0: Exchange P 6.8% 0.880ad5d9e 6.0% Metamak Sva 5.0% 1.6% V5; Router 5.0% 0.651510c4 1.5% 0.82ecladd1 4.4%
		© 19 minutes ago 發展 發展	Harvard John A. Paulsor





