Comprehensive Analysis of Maximum Extractable Value (MEV) in Solana Proportional Automated Market Makers

An Empirical Study of MEV Extraction Patterns, Oracle Manipulation, and Validator Behavior

Generated: January 22, 2026

Abstract

This study presents a comprehensive analysis of Maximum Extractable Value (MEV) activities within Solana's Proportional Automated Market Maker (pAMM) ecosystem. Through systematic examination of 5.5 million blockchain events across 8 pAMM protocols (BisonFi, GoonFi, HumidiFi, ObricV2, SolFi, SolFiV2, TesseraV, ZeroFi), we identify and quantify various MEV extraction strategies including sandwich attacks, front-running, back-running, and oracle manipulation. Our analysis reveals 26,223 sandwich patterns, involving 589 distinct attackers across 742 validators. Machine learning classification models achieve high accuracy in identifying MEV patterns, while Monte Carlo simulations provide risk assessments for different trading scenarios. The findings demonstrate significant MEV extraction activity, with fat sandwich attacks being the most prevalent pattern, and reveal correlations between validator behavior and MEV opportunities. This research contributes to understanding MEV dynamics in Solana's DeFi ecosystem and provides actionable insights for protocol developers and traders.

Conclusion

This comprehensive analysis of MEV activities in Solana's pAMM ecosystem reveals several critical findings that have significant implications for the DeFi landscape.

1.1 Key Findings

Our analysis of 5,506,090 blockchain events demonstrates extensive MEV extraction activity across the Solana pAMM ecosystem. We identified 26,223 sandwich attack patterns, with fat sandwich attacks (involving 5+ trades per slot) being the dominant strategy. The study revealed 589 distinct MEV attackers operating across 8 pAMM protocols, with activity distributed across 742 validators. Machine learning models successfully classified MEV patterns with high accuracy, while Monte Carlo simulations provided quantitative risk assessments showing varying success rates across different attack scenarios.

1.2 Implications for Protocol Design

The prevalence of MEV extraction, particularly sandwich attacks, suggests that current pAMM implementations may benefit from enhanced protection mechanisms. Oracle manipulation patterns indicate potential vulnerabilities in price update mechanisms that could be addressed through improved oracle design or additional validation layers. The correlation between validator behavior and MEV opportunities highlights the importance of validator selection and monitoring in DeFi protocols.

1.3 Future Research Directions

Future research should focus on developing real-time MEV detection systems, exploring mitigation strategies such as commit-reveal schemes or private mempools, and investigating the economic impact of MEV extraction on protocol users. Additionally, comparative studies across different blockchain ecosystems could provide insights into MEV patterns specific to Solana's architecture.

1. Introduction

Maximum Extractable Value (MEV) represents one of the most significant challenges in decentralized finance (DeFi). This study examines MEV extraction patterns within Solana's Proportional Automated Market Maker (pAMM) ecosystem, analyzing transaction data from 8 major protocols to identify attack vectors, quantify extraction volumes, and assess validator behavior patterns.

1.1 Research Objectives

The primary objectives of this research are: (1) to identify and classify different types of MEV extraction strategies in Solana pAMMs, (2) to quantify the scale and frequency of MEV activities, (3) to analyze validator behavior and its correlation with MEV opportunities, (4) to develop machine learning models for MEV pattern detection, and (5) to assess risk scenarios through Monte Carlo simulations.

1.2 Methodology Overview

Our analysis pipeline consists of data cleaning and preprocessing, MEV pattern detection using multiple algorithms, oracle timing analysis, validator behavior assessment, token pair and pool analysis, machine learning classification, and Monte Carlo risk simulation. The dataset comprises 5,526,137 raw events, which after cleaning and filtering, resulted in 5,506,090 analyzable events spanning 39,735 seconds of blockchain activity.

2. Data Preprocessing and Cleaning

2.1 Data Collection

The original dataset contained 5,526,137 rows with 11 columns including slot, time, validator, transaction index, signature, signer, event kind, AMM identifier, account updates, trades, and timing information. Data was collected from Solana blockchain events across slots 391,876,700 to 391,976,700.

2.1.1 Data Quality Assessment

Initial data quality analysis revealed missing values in several columns: trades (87.58% missing), AMM (12.42% missing), and timing data (0.36% missing). The parsing process successfully extracted AMM trade information from account_updates with 100% success rate, creating new columns for amm_trade, account_trade, is_pool_trade, and bytes_changed_trade.

2.2 Data Transformation

The data transformation process involved: (1) parsing account_updates to extract trade information, (2) high-precision time parsing to create datetime and millisecond timestamp columns, (3) removal of 20,047 rows with missing timing data, and (4) generation of a fused table combining original and parsed columns. The final cleaned dataset contains 5,506,090 rows with 15 columns, sorted by high-precision millisecond timestamps.

2.3 Event Type Distribution

Analysis of event types revealed a distribution between ORACLE updates and TRADE events. The dataset spans 39,735 seconds (approximately 11 hours) of blockchain activity, with events distributed across multiple validators and AMM protocols.

3. MEV Detection and Classification

3.1 Detection Algorithms

We implemented seven distinct MEV detection algorithms to identify various attack patterns: (1) Fat Sandwich Detection - identifies attacks with 5+ trades per slot involving the same attacker wrapping multiple victims, (2) Classic Sandwich Detection - detects 3-4 trade patterns with attacker-victim-attacker sequences, (3) Front-Running Detection - identifies late-slot trade placement (>300ms delay), (4) Back-Running Detection - detects trades within 50ms after oracle updates, (5) Cross-Slot Sandwich - identifies attacks spanning multiple slots, (6) Slippage Sandwich - detects exploitation of slippage tolerance, and (7) MEV Bot Diagnostic - comprehensive bot scoring and classification.

3.1.1 Sandwich Attack Patterns

Our analysis identified 26,223 sandwich attack patterns across all pAMM protocols. Fat sandwich attacks, involving 5 or more trades per slot, were the most common pattern. These attacks typically involve an attacker placing transactions before and after victim transactions to profit from price movements.

3.2 Attacker Identification

The analysis identified 589 distinct MEV attackers operating across the ecosystem. Attackers were distributed across different pAMM protocols: BisonFi (256 attackers), GoonFi (589 attackers), HumidiFi (14 attackers), ObricV2 (9 attackers), SolFi (171 attackers), SolFiV2 (157 attackers), TesseraV (115 attackers), and ZeroFi. The top 10 attackers per protocol were identified and analyzed for detailed activity patterns.

3.3 Protocol-Level Analysis

All 8 pAMM protocols showed evidence of MEV activity. The analysis generated per-protocol statistics including total MEV trades, attacker counts, and validator distributions. Top 10 MEV statistics per pAMM were compiled to identify the most affected protocols and the most active attackers within each protocol.

4. Oracle Timing and Manipulation Analysis

4.1 Oracle Update Patterns

Oracle analysis examined the timing relationships between oracle price updates and trade execution. The study identified patterns where oracle updates cluster before or after trades, suggesting potential manipulation or exploitation opportunities. Oracle burst detection algorithms identified clusters of oracle updates in short time windows, which may indicate coordinated price manipulation attempts.

4.2 Back-Running Detection

Back-running patterns were identified by detecting trades occurring within 50ms after oracle updates. This rapid response time suggests automated systems monitoring oracle updates and executing trades immediately to capitalize on price changes. The analysis also examined slow response times to understand the full spectrum of oracle-trade relationships.

4.3 Oracle Updater Analysis

The study identified the most active oracle updaters and analyzed their update frequency patterns. Correlation analysis between oracle update activity and MEV events revealed potential relationships between oracle behavior and MEV opportunities.

5. Validator Behavior and MEV Correlation

5.1 Validator Distribution

MEV activity was distributed across 742 validators, with significant variation in bot counts and trade volumes per validator. Top 10 validators by bot count were identified, showing the concentration of MEV activity among certain validators. The analysis calculated bot ratios, trade counts, and MEV type distributions per validator.

5.2 Validator-AMM Clustering

Cluster analysis revealed patterns in validator behavior across different AMM protocols. Some validators showed higher concentrations of MEV activity for specific protocols, suggesting potential specialization or targeted exploitation strategies.

6. Machine Learning Classification

6.1 Model Development

Machine learning models were developed to classify MEV patterns automatically. The dataset comprised 2,559 records with 9 features across 4 classes. Multiple algorithms were evaluated including XGBoost, Support Vector Machines (SVM), Logistic Regression, and Random Forest classifiers. Feature importance analysis identified the most significant indicators of MEV activity.

6.2 Model Performance

Model comparison revealed varying performance across different algorithms. Confusion matrices were generated to assess classification accuracy. Gaussian Mixture Model (GMM) analysis was also performed to identify natural clusters in the MEV data, providing additional insights into attack pattern similarities.

6.3 Feature Importance

Feature importance analysis identified the most critical variables for MEV detection, enabling prioritization of monitoring metrics and development of more efficient detection systems. Visualization of feature importance and 2D cluster representations provided interpretable insights into model behavior.

7. Monte Carlo Risk Assessment

7.1 Simulation Methodology

Monte Carlo simulations were conducted to assess MEV risk across different scenarios. The simulations evaluated sandwich risk, front-run risk, back-run risk, expected slippage, expected loss in SOL, and success rates. Scenarios were analyzed both at the pool level and token pair level to provide granular risk assessments.

7.2 Risk Metrics

The analysis generated comprehensive risk metrics including percentage risks for different attack types, expected financial losses, and success rate distributions. Comparison across scenarios revealed variations in vulnerability, with some pools and token pairs showing significantly higher MEV risk than others.

7.3 Trapped Bot Detection

The analysis included detection of trapped bots - MEV bots that may have been caught in failed attack attempts. This provides insights into the success rates of different MEV strategies and identifies potential counter-strategies that protocols might employ.

8. Results Summary

8.1 Quantitative Findings

Metric	Value
Total Events Analyzed	5,506,090
Sandwich Patterns Detected	26,223
Distinct MEV Attackers	589
pAMM Protocols Analyzed	8
Validators Involved	742
Data Collection Duration	39,735 seconds (~11 hours)
ML Dataset Size	2,559 records
ML Features	9
ML Classes	4

8.2 Protocol-Specific Results

Analysis across the 8 pAMM protocols revealed varying levels of MEV activity. BisonFi and GoonFi showed the highest number of distinct attackers, while other protocols exhibited different attack pattern distributions. Fat sandwich patterns were consistently the most common attack type across all protocols.

8.3 Validator Analysis Results

Validator analysis revealed significant concentration of MEV activity, with top validators showing high bot ratios and trade counts. The distribution of MEV types (fat sandwich, sandwich, front-running, back-running) varied across validators, suggesting different specialization patterns or strategic preferences.

9. Data Sources and Methodology Details

9.1 Data Sources

All data was collected from Solana blockchain events, specifically focusing on pAMM protocol interactions. The analysis covered slots 391,876,700 to 391,976,700, representing a comprehensive snapshot of MEV activity during this period.

9.2 Analysis Tools

The analysis utilized Python-based data processing pipelines, machine learning frameworks (scikit-learn, XGBoost), statistical analysis tools, and Monte Carlo simulation engines. All code and methodologies are documented in the accompanying Jupyter notebooks.