



Publisher-Centric Machine Learning-Based Solution for Click Fraud

Gihan Saminda Pathirage

MS23003378

A THESIS

SUBMITTED TO

SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Dr. Kalpani Manathunga

Approved for MSc. Research Project:

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

A handwritten signature in black ink that reads "Gihan Saminda". The signature is written in a cursive style, with the first name "Gihan" and the last name "Saminda" clearly legible.

Sign: Gihan Saminda Pathirage

Date: 13/09/2024.....

ABSTRACT

Publisher-Centric Machine Learning-Based Solution For Click Fraud

Gihan Saminda Pathirage

MSc. in Information Technology

Supervisor: Dr. Kalpani Manathunga

December 2024

Invalid traffic and click fraud present significant challenges in online advertising, impacting advertising metrics and causing substantial financial losses across the digital advertising ecosystem. While advertisers have access to various protective solutions and receive protection from advertising networks, publishers face limited options for detecting and preventing fraudulent activities on their websites. This gap in publisher-side protection creates a critical area for investigation and development of practical solutions. This research presents an effective publisher-side solution: the Ad Click Fraud Protector (ACFP), an open-source WordPress plugin that detects and prevents click fraud and invalid traffic. The research methodology involved studying browser fingerprinting approaches by collecting browser fingerprints from legitimate users and bots, distinguished through firewall rules and honeypots.

Experimental analysis identified six key browser fingerprinting attributes that effectively distinguish between legitimate and fraudulent traffic. These findings informed the development of the ACFP plugin, which incorporates additional security measures for enhanced protection. Testing of the plugin on two AdSense publisher accounts demonstrated its effectiveness in reducing invalid clicks, minimizing invalid traffic, and decreasing revenue deductions due to invalid clicks. The results show that publishers can effectively protect their ad accounts from penalties and deductions through browser fingerprint-based traffic filtering. This research provides publishers with an accessible, open-source solution for combating click fraud while contributing to the theoretical understanding of browser fingerprinting effectiveness in fraud detection. Additionally, it establishes a framework for future development in publisher-side protection systems.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Dr. Kalpani Manathunga, for her invaluable guidance and support throughout this research journey. Her mentorship has been instrumental in shaping this work, and her encouragement kept me motivated during challenging times. Her commitment to addressing my questions and providing detailed guidance throughout the entire process has been truly remarkable.

I extend my sincere appreciation to Dr. Prasanna Haddala, Dr. Dilshan De Silva, Dr. Dharshana Kasthurirathna, and Dr. Nathali Silva for their insightful feedback during my proposal and progress presentations. I am also grateful to Mr. Jagath Wickramaratne for his enlightening lecture series on research methodology and to all the academic staff at SLIIT who supported me during my academic journey. I would also like to acknowledge my fellow students at SLIIT for their friendship and support throughout this endeavor.

A special note of gratitude goes to my family for their unconditional love, patience, and understanding throughout this challenging journey. Additionally, I wish to acknowledge the various online communities, particularly on Reddit, for providing valuable resources and engaging in constructive discussions that enriched my research.

This would not have been possible without the collective support of everyone mentioned above. Their contributions have been instrumental in bringing this work to fruition.

TABLE OF CONTENTS

DECLARATION.....	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
List of Figures	ix
List of Tables.....	xii
Abbreviations.....	xiii
Chapter 1 Introduction	1
1.1 Background and Context.....	1
1.1.1 Evolution of Digital Advertising.....	1
1.2 Problem Statement.....	3
1.3 Aims of the Research	5
1.4 Significance of the Study	5
1.5 Organization of Thesis	6
Chapter 2 Literature Review	7
2.1 Understanding and Combating Click Fraud.....	7
2.1.1 The Landscape of Digital Advertising Fraud: Scope and Impact	7
2.1.2 Evolution of Detection Strategies: From Traditional Methods to AI-Driven Solutions.....	8
2.1.3 Advanced Bot Detection: Combating Sophisticated Automated Threats.....	9
2.2 Future Frontiers in Click Fraud Prevention	11
2.2.1 Emerging Challenges in the Digital Advertising Ecosystem	11
2.2.2 Innovation and Adaptation: Next-Generation Solutions	11
2.2.3 Balancing Security and Privacy: Ethical Considerations	12
Chapter 3 Methodology	14
3.1 Research Design.....	14
3.1.1 Introduction.....	14
3.2 Procedures.....	14
3.3 Tools and Techniques	15
3.4 The Sample/Participants	15
3.4.1 Website Sample.....	16
3.4.2 Survey and Interview Participants	16
3.5 Data Collection Methods	16
3.5.1 Quantitative Data Collection.....	17
3.5.1 Qualitative Data Collection.....	25

3.6 Dataset Imbalance (Disparity Between Good and Bad)	25
3.7 Quantitative Data Analysis	26
3.7.1 Canvas Fingerprint Analysis	26
3.7.2 Textual Browser Attributes Analysis.....	27
3.7.3 Machine Learning Analysis	27
3.8 Model Development Process	28
3.9 Qualitative Data Analysis.....	28
3.10 Testing and Validation.....	29
3.10.1 Evaluation Metrics	29
3.10.2 Load Testing.....	30
3.10.3 Behavioral Testing	30
3.10.4 Integration Testing	30
3.10.5 Real-World Validation.....	30
3.11 Ethical Considerations	31
Chapter 4 Implementation.....	32
4.1 System Architecture Overview	32
4.2 Server-Side Component	32
4.3 WordPress Plugin Component	33
4.4 System Workflow.....	35
4.5 Initialization Phase.....	35
4.6 Verification Phase	35
4.7 Display Management	35
4.8 Ongoing Monitoring	35
4.9 Security Measures.....	35
4.10 Integration Capabilities	35
Chapter 5 Results	37
5.1 Survey and Initial Findings.....	37
5.2 Preliminary Analysis Results	37
5.2.1 Canvas Fingerprint Summary Statistics	37
5.2.2 Textual Browser Fingerprint Attributes Summary Statistics.....	38
5.2.3 Correlation Analysis.....	39
5.2.4 Preliminary Feature Importance.....	40
5.3 Model Performance Results.....	41
5.3.1 Canvas Fingerprint Model	42
5.3.2 Textual Browser Fingerprint Attributes.....	44
5.4 Plugin Testing Results.....	47
5.5 Real-World Implementation Results	49

Chapter 6 Discussion	51
6.1 Overview of Findings	51
6.2 Canvas Fingerprint Analysis Process and Findings	51
6.3 Discussion of Selected Features.....	54
6.3.1 PDF Viewer Status	54
6.3.2 Font Loading Duration and Font Count	55
6.3.1 Font Preferences.....	57
6.3.1 Open Database Availability.....	59
6.3.1 Plugins.....	60
6.4 Discussion of Other Browser Attributes	61
6.4.1 DOM Blockers	61
6.4.2 Audio Duration.....	62
6.4.3 Screen Frame Dimensions	64
6.4.4 Operating System CPU	65
6.4.5 Languages	66
6.4.6 Color Depth.....	67
6.4.7 Device Memory.....	68
6.4.8 Screen Resolution	69
6.4.9 Hardware Concurrency	70
6.4.10 Time Zone	71
6.4.11 Session Storage	72
6.4.12 Local Storage	73
6.4.13 IndexedDB Support.....	74
6.4.14 CPU Class	75
6.4.15 Platform.....	76
6.4.16 Canvas.....	78
6.4.17 Touch Support	79
6.4.18 Vendor	80
6.4.19 Vendor Flavors	82
6.4.20 Cookies Enabled	83
6.4.21 Color Gamut.....	84
6.4.22 Inverted Colors.....	86
6.4.23 Forced Colors.....	87
6.4.24 Monochrome	88
6.4.25 Contrast	89
6.4.26 Reduced Motion.....	89
6.4.27 HDR	90

6.4.28 Math	92
6.4.29 Video Card	92
6.4.30 Architecture	94
6.5 Additional Observations.....	96
Chapter 7 Conclusion.....	99
7.1 Introduction.....	99
7.2 Revisiting Research Questions.....	99
7.3 Implications.....	101
7.4 Contributions.....	102
7.4.1 Two-Layered Protection System for Publishers.....	102
7.4.2 Identification of Key Browser Fingerprinting Attributes	102
7.4.3 Novel Browser Fingerprint Dataset	103
7.5 Limitations	103
7.6 Future Research Directions.....	105
References.....	107
Appendices.....	115
Appendix 1: Survey Questions	115
Appendix 2: Textual Browser Fingerprint Attribute Statistics.....	123

List of Figures

Figure 1.1: Digital Advertising Ecosystem Overview	1
Figure 1.2: IVT Incident	4
Figure 3.1: Procedure.....	14
Figure 3.2: Fingerprint Collection Workflow	19
Figure 3.3 Distribution in Final Dataset.....	21
Figure 3.4: Site 1 Cloudflare Traffic Analysis	21
Figure 3.5: Site 2 Cloudflare Traffic Analysis	22
Figure 3.6: Site 3 Cloudflare Traffic Analysis	22
Figure 3.7: Site 4 Cloudflare Traffic Analysis	23
Figure 3.8: Site 5 Cloudflare Traffic Analysis	23
Figure 3.9: Cloudflare Rule Configuration.....	24
Figure 3.10: Cloudflare Rule Performance Metrics.....	24
Figure 3.11: Model Development Pipeline	28
Figure 4.1: System Architecture Diagram.....	32
Figure 4.2: Model Prediction Flow	33
Figure 4.3: Plugin Advanced Settings Interface.....	33
Figure 4.4: Banned Users Interface.....	34
Figure 4.5: Proxy & Location Settings Interface	34
Figure 4.6: General Settings Interface	34
Figure 4.7 Algorithm.....	36
Figure 5.1: Correlation Matrix Heatmap.....	40
Figure 5.2: Textual Fingerprint Feature Importance (Pre-model).....	41
Figure 5.3: Canvas Fingerprint Feature Importance (Pre-model).....	41
Figure 5.4: Canvas Fingerprint Final Feature Rankings	42
Figure 5.5: Canvas Fingerprint Model Confusion Matrix	43
Figure 5.6: Canvas Fingerprint Model Learning Curve.....	43
Figure 5.7: Canvas Fingerprint Model ROC Curve.....	44
Figure 5.8: Textual Browser Fingerprint Model Confusion Matrix.....	45
Figure 5.9: Textual Browser Fingerprint Model ROC Curve	45
Figure 5.10: Textual Browser Fingerprint Model Final Precision-Recall Curve.....	46
Figure 5.11: Textual Browser Fingerprint Model Final Feature Importances.....	46
Figure 5.12: Plugin Test Results Sample	47
Figure 5.13: Plugin Test Results Sample	47
Figure 5.14: Plugin Test Results Sample	48
Figure 5.15: Plugin Test Environment Setup	48
Figure 5.16: Invalid Traffic Deduction Before and After Implementation	50
Figure 6.1: Good Canvas Fingerprint Sample	52
Figure 6.2: Good Canvas Fingerprint: Dark Brown Channel Analysis	52
Figure 6.3: Good Canvas Fingerprint: Blue Channel Analysis.....	53
Figure 6.4: Bad Canvas Fingerprint Sample.....	53
Figure 6.5: Bad Canvas Fingerprint: Dark Brown Channel Analysis	53
Figure 6.6: Bad Canvas Fingerprint: Blue Channel Analysis	54
Figure 6.7: Distribution of pdfViewerEnabled	55
Figure 6.8: Distribution of pdfViewerEnabled_duration	55
Figure 6.9: Distribution of fonts_count.....	56
Figure 6.10: Distribution of fonts_duration	57

Figure 6.11: Distribution of fontPreferences_duration	58
Figure 6.12: Distribution of fontPreferences_min	58
Figure 6.13: Distribution of fontPreferences_mono	58
Figure 6.14: Distribution of fontPreferences_sans.....	59
Figure 6.15: Distribution of fontPreferences_serif	59
Figure 6.16: Distribution of fontPreferences_system	59
Figure 6.17: Distribution of openDatabase	60
Figure 6.18: Distribution of plugins_count.....	61
Figure 6.19: Distribution of domBlockers_duration	62
Figure 6.20 Distribution of audio.....	64
Figure 6.21: Distribution of audio_duration	64
Figure 6.22: Distribution of screenFrame	65
Figure 6.23: Distribution of osCpu_duration	66
Figure 6.24: Distribution of languages_duration	67
Figure 6.25: Distribution of colorDepth	68
Figure 6.26: Distribution of deviceMemory	69
Figure 6.27: Distribution of deviceMemory_duration	69
Figure 6.28: Distribution of screenResolution_duration.....	70
Figure 6.29: Distribution of hardwareConcurrency	71
Figure 6.30: Distribution of hardwareConcurrency_duration.....	71
Figure 6.31: Distribution of timezone_duration.....	72
Figure 6.32: Distribution of sessionStorage_duration	73
Figure 6.33: Distribution of localStorage_duration	74
Figure 6.34: Distribution of indexedDB_duration.....	75
Figure 6.35: Distribution of cpuClass	76
Figure 6.36: Distribution of platform.....	77
Figure 6.37: Distribution of platform_duration	77
Figure 6.38: Distribution of plugins_duration	78
Figure 6.39: Distribution of canvas_winding	78
Figure 6.40: Distribution of canvas_duration	79
Figure 6.41: Distribution of touchSupport_duration.....	80
Figure 6.42: Distribution of touchSupport_maxTouchPoints	80
Figure 6.43: Distribution of touchSupport_touchEvent.....	80
Figure 6.44: Distribution of vendor	81
Figure 6.45: Distribution of vendor_duration	82
Figure 6.46: Distribution of vendorFlavor	83
Figure 6.47: Distribution of vF_duration.....	83
Figure 6.48: Distribution of cookiesEnabled	84
Figure 6.49: Distribution of colorGamut	85
Figure 6.50: Distribution of colorGamut_duration	85
Figure 6.51: Distribution of invertedColors.....	86
Figure 6.52: Distribution of forcedColors.....	87
Figure 6.53: Distribution of forcedColors_duration	88
Figure 6.54: Distribution of reducedMotion	90
Figure 6.55: Distribution of reducedMotion_duration.....	90
Figure 6.56: Distribution of hdr	91
Figure 6.57: Distribution of hdr_duration.....	91
Figure 6.58: Distribution of math_duration	92
Figure 6.59: Distribution of videoCard_vendor.....	93
Figure 6.60: Distribution of videoCard_duration.....	94
Figure 6.61: Distribution of architecture.....	95

Figure 6.62: Distribution of architecture_duration	95
Figure 6.63: Top Threat Regions for Site 1 (Monthly)	96
Figure 6.64: Top Threat Regions for Site 2 (Monthly)	97
Figure 6.65: Top Threat Regions for Site 3 (Monthly)	97
Figure 6.66: Top Threat Regions for Site 4 (Monthly)	98
Figure 6.67: Top Threat Regions for Site 5 (Monthly)	98

List of Tables

Table 3.1: Website Sample	16
Table 3.2: Collected Fingerprint Data Points	17
Table 3.3: Honeypot Details.....	19
Table 5.1: Classification Report for Canvas-Based Model.....	42
Table 5.2: Classification Report for Textual Attribute-Based Model.....	44
Table 5.3: Website 1 Results	49
Table 5.4: Website 2 Results	49

Abbreviations

Abbreviation	Full Form
ACFP	Ad Click Fraud Protector
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
API	Application Programming Interface
AWS	Amazon Web Services
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CPM	Cost Per Mille
CPC	Cost Per Click
CPA	Cost Per Action
CSS	Cascading Style Sheets
DDoS	Distributed Denial of Service
DMP	Data Management Platform
DSP	Demand-Side Platform
GDPR	General Data Protection Regulation
GIVT	General Invalid Traffic
HDR	High Dynamic Range
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IP	Internet Protocol
JSON	JavaScript Object Notation
JWT	JSON Web Token
LSH	Locality-Sensitive Hashing
LSTM	Long Short-Term Memory
ML	Machine Learning
PNG	Portable Network Graphics
PPC	Pay Per Click
RGB	Red Green Blue
ROI	Return on Investment
RTB	Real-Time Bidding
SIVT	Sophisticated Invalid Traffic
SME	Small and Medium-sized Enterprise
SSP	Supply-Side Platform
SVG	Scalable Vector Graphics
UI	User Interface
UTC	Coordinated Universal Time
VPN	Virtual Private Network
VPS	Virtual Private Server
WAF	Web Application Firewall
WebGL	Web Graphics Library