

**LEVERAGING HONEYPOTS AND MACHINE LEARNING
TO DETECT MALICIOUS ETHEREUM ACCOUNTS**

By
Yevadrian Andreas
22251025

MASTER'S DEGREE
in
MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 06 February 2025

**LEVERAGING HONEYPOTS AND MACHINE LEARNING
TO DETECT MALICIOUS ETHEREUM ACCOUNTS**

By
Yevadrian Andreas
22251025

MASTER'S DEGREE
in
MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

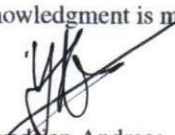


SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 06 February 2025

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work, and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgment is made in the thesis.


Yevadrian Andreas

Student

20/02/2025

Date

Approved by:


Dr. Eka Budiarto, S.T., M.Sc.

Thesis Advisor

20/02/2025

Date


Kalpin Erlangga Silaen, S.Si., M.Kom.

Thesis Co-Advisor

18/20/2025

Date

Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc

Dean of Faculty of Engineering and Information
Technology

Date

ABSTRACT

LEVERAGING HONEYPOTS AND MACHINE LEARNING TO DETECT MALICIOUS ETHEREUM ACCOUNTS

By

Yevadrian Andreas

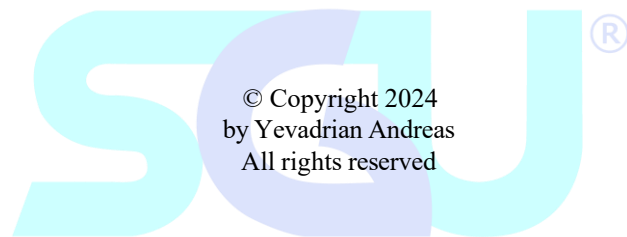
Dr. Eka Budiarto, S.T., M.Sc., Advisor

Kalpin Erlangga Silaen, S.Si., M.Kom., Co-Advisor

SWISS GERMAN UNIVERSITY

The rise of Web3 has brought decentralization and transparency, with Ethereum emerging as a key platform for decentralized applications. However, its popularity has also made it a prime target for attacks. This study aims to enhance Ethereum security by improving node honeypots, account risk scoring, and account classification. Over 28 days, Ethereum node honeypots recorded 88,224 JSON-RPC requests, revealing attacker patterns such as reconnaissance through account enumeration. The study investigates three key research questions. First, findings show that honeypots returning transaction hashes, especially random ones, receive the highest engagement, as attackers verify these hashes on block explorers before attempting Ether theft. This highlights the importance of response design in studying attacker behavior. Second, the study evaluates the effectiveness of the RiskProp method for Ethereum account risk scoring and finds it inadequate for newly observed or low-activity accounts due to its reliance on prior transaction history. Lastly, the study explores improvements to Ethereum account classification, demonstrating that machine learning models such as Random Forest and XGBoost outperform RiskProp in accuracy and generalization. Key classification features include transaction counts, active days, and Ether balance. Based on these findings, the study recommends enhancing honeypots with high-balance accounts to attract more sophisticated attackers, exploring adaptive risk-scoring methods, and incorporating deep learning techniques for improved classification. These insights contribute to advancing Ethereum security research and strengthening risk assessment methodologies. The code and dataset used in this study will be publicly available at GitHub.

Keywords: Ethereum, Blockchain, Honeypot, Cyber Security, Machine Learning,



DEDICATION

This thesis is dedicated to God and my family, who have been my greatest source of strength throughout this journey. I also dedicate this work to my advisors and co-advisors, whose invaluable guidance has been instrumental in completing my thesis. Additionally, I dedicate this work to my friends and lecturers who have supported me throughout my postgraduate studies. Finally, this work is dedicated to my country, Indonesia, and the global academic community.



ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr. Eka Budiarto, S.T., M.Sc. and Kalpin Erlangga Silaen, S.Si., M.Kom., my thesis advisor and co-advisor, for their time, invaluable advice, and guidance throughout this research project and the completion of this thesis. I am also deeply thankful to my family for their unwavering moral support throughout this journey. Finally, I extend my thanks to my lecturers and friends in MIT SGU for their encouragement during the course of this thesis.



Contents

	Page
STATEMENT BY THE AUTHOR	1
ABSTRACT	2
DEDICATION	4
ACKNOWLEDGEMENT	5
TABLE OF CONTENTS	8
LIST OF FIGURES	9
LIST OF TABLES	10
1 INTRODUCTION	11
1.1 Background.....	11
1.2 Problem Statement.....	12
1.3 Research Objective	12
1.4 Research Questions.....	13
1.5 Scope and Limitation	14
1.6 Significance of Study	14
1.7 Thesis Structure	14
2 LITERATURE REVIEW	16
2.1 Web Technology	16
2.1.1 Evolution of Web Technology.....	16
2.1.2 Differences Between Web 3.0 and Web3	17
2.2 Blockchain	17
2.2.1 Key Characteristics of Blockchain	17
2.2.2 Applications of Blockchain	19
2.2.3 Types of Blockchain.....	20
2.3 Ethereum.....	21
2.3.1 Ethereum Currency	22
2.3.2 Ethereum Layer.....	22

2.3.3	Ethereum Account	23
2.3.4	Ethereum Transaction	24
2.3.5	Ethereum Node	25
2.3.6	Attacks on Ethereum	26
2.3.7	Studies on Ethereum Account Risk Scoring	27
2.3.7.1	Lin et al. (2024)	28
2.3.7.2	Fu et al. (2023)	29
2.3.7.3	Review on Ethereum Account Risk Scoring Studies	30
2.4	Honeypot	30
2.4.1	Honeypot and Cyber Deception	30
2.4.2	Applications of Honeypot	31
2.4.3	Types of Honeypot	31
2.4.4	Ethereum Honeypot	32
2.4.5	Studies on Ethereum Node Honeypot	32
2.4.5.1	Cheng et al. (2019)	32
2.4.5.2	Hara et al. (2020)	33
2.4.5.3	Chin et al. (2021)	33
2.4.5.4	Wang et al. (2022)	34
2.4.5.5	Subhan et al. (2023)	34
2.4.5.6	Sasaki et al. (2024)	34
2.4.5.7	Uchibori et al. (2024)	35
2.4.5.8	Review on Ethereum Node Honeypot Studies	35
2.5	Machine Learning	36
2.5.1	Types of Machine Learning	36
2.5.1.1	Random Forest	37
2.5.1.2	Extreme Gradient Boosting	38
2.5.2	Machine Learning Applications	38
2.5.2.1	Machine Learning in Ethereum	38
2.5.3	Machine Learning Performance Measurement	39
2.5.4	Machine Learning with Python	40
2.5.4.1	Role of Python in Machine Learning	40
2.5.4.2	Scikit-learn Library	40
2.5.4.3	Additional Python Libraries for Machine Learning	40
2.5.4.4	Hyperparameter Tuning with GridSearchCV	41
2.5.4.5	Performance Measurement with Classification Report	42
3	RESEARCH METHODS	43
3.1	Honeypot Design	43
3.2	Honeypot Deployment	45

3.3	Honeypot Data Analysis	45
3.4	Risk Scoring Data Preprocessing	46
3.5	Risk Scoring Calculation	46
3.6	Risk Scoring Results Analysis	47
3.7	Custom Model Data Preprocessing	48
3.8	Custom Model Training and Testing.....	48
3.9	Custom Model Results Analysis	49
4	RESULTS AND DISCUSSIONS	50
4.1	Honeypot Data Analysis	50
4.1.1	Collected Data.....	50
4.1.2	Attacks by IP Addresses.....	50
4.1.3	Attacks by Request Method	50
4.1.4	Attacks by Timestamp.....	53
4.1.5	Attacks by Proxy Mode.....	54
4.1.6	Effective Honeypot Responses	56
4.2	Risk Scoring Results Analysis	56
4.2.1	Honeypot Attackers Risk Scoring	56
4.2.2	Risk Scoring Improvement.....	56
4.3	Custom Model Results Analysis	57
4.3.1	Optimal Hyperparameter.....	57
4.3.2	Model Performance.....	57
4.3.3	Feature Importance.....	58
5	CONCLUSION AND RECOMMENDATIONS	60
5.1	Conclusion.....	60
5.2	Recommendations.....	60
5.2.1	Ethereum Node Honeypot.....	61
5.2.2	Ethereum Account Risk Scoring	61
5.2.3	Ethereum Account Classification	61
	REFERENCES	62
	CURRICULUM VITAE	68

List of Figures

1.1	Top Blockchains by Total Value Locked	11
1.2	Fishbone Diagram 1	13
1.3	Fishbone Diagram 2	13
3.1	Research Framework	43
3.2	Honeypot Design	45
4.1	Requests Trend of Proxy Mode A	53
4.2	Requests Trend of Proxy Mode B	53
4.3	Requests Trend of Proxy Mode C	54
4.4	Requests Trend of Combined Dataset	54
4.5	Requests Trend by All IP Addresses	55
4.6	Requests Trend by Attackers IP Addresses	55
4.7	Random Forest Feature Importance	58
4.8	XGBoost Feature Importance	59

List of Tables

2.1	Ethereum Node Honeypot Related Studies	36
3.1	Proxy Specification	44
4.1	Top IP Addresses	51
4.2	All Request Methods	52
4.3	Requests by Proxy Mode	55
4.4	Dataset Performance for Licit Account Classification	57
4.5	Dataset Performance for Illicit Account Classification	57
4.6	Model Performance for Licit Account Classification	57
4.7	Model Performance for Illicit Account Classification	58
4.8	Average Counts per Account	59

