Crypto Vulnerability Web Intelligence Analysis with Machine Learning and Social Network Analysis

James Usher

Abstract

This paper investigates the identification and classification of cryptocurrency vulnerabilities using web intelligence, machine learning (ML), and social network analysis (SNA). By leveraging data from the Rekt crypto news website, the aim is to enhance the accuracy of vulnerability severity classification. The comprehensive analytics pipeline employs natural language processing (NLP) and ML to analyse crypto vulnerabilities, identify hacker groups through unsupervised ML, and predict suspicious transactions from on-chain trade activity. The integration of SNA allows us to map the relationships within networks of actors involved in crypto exploits, facilitating the detection and disruption of coordinated attacks. This research provides a robust framework for proactive cryptocurrency vulnerability management, contributing valuable insights to improve the security and resilience of digital asset systems.

1. Introduction

Web intelligence encompasses the collection and analysis of data from online sources to gain insights into crypto vulnerabilities. Machine learning algorithms are employed to automate the detection and classification of these vulnerabilities, leveraging large datasets and advanced analytics to identify patterns and anomalies. Social network analysis provides a framework for understanding the relationships and interactions within networks of actors,

such as hacker groups, which can be crucial for identifying and mitigating coordinated attacks. One of the primary sources utilised in this research is the *Rekt news* website, an anonymous platform dedicated to analysing and reporting major hacks and exploits within the decentralised finance (DeFi) and cryptocurrency space. *Rekt.news* offers detailed accounts of significant security breaches, providing valuable data for researchers and practitioners. The platform's leaderboard ranks various hacks and exploits, offering insights into the most significant vulnerabilities. This resource is instrumental in understanding the landscape of crypto vulnerabilities and the usefulness of different security measures. The objective of this study is to identify and analyse these vulnerabilities using advanced web intelligence techniques. By sourcing and gathering Crypto Vulnerability Smart Contract (CVSC) data from credible sources, such as *Rekt.news,* the aim to establish a robust mechanism that serves multiple purposes. Firstly, a mechanism for rating these vulnerabilities is proposed. Through applied analytics and Natural Language Processing (NLP), we can assess the severity of each smart contract vulnerability as taken from *Rekt.news*. This enables a more nuanced understanding of the risks involved, facilitating better-informed decision-making for investors and stakeholders alike. Secondly, the capture and analysis of CVSC data provide a foundation for future ML models. By aligning hacker groups and tracking Ethereum networks, we can assign groupings based on ML analysis of hack activities. This helps in understanding the behaviour and strategies of various hacker groups. Furthermore, we intend to apply known CVSC vulnerabilities to hacks based on the frameworks established by the Cloud Security Alliance (CSA) working groups. This approach not only standardises the methodology but also leverages existing industry knowledge to enhance the analysis. The main contributions of this research include the development of an advanced analytics pipeline for crypto vulnerability severity analysis. This pipeline integrates multiple sources of data, including web intelligence from *Rekt.news,* to enhance the accuracy and efficiency of vulnerability identification and classification. The pipeline also incorporates NLP techniques to analyse textual data related to vulnerabilities, enabling the extraction of relevant features and the assessment of their severity. Finally, by identifying rogue network nodes and capturing potential suspicious activities, the aim is to support Suspicious Activity Reporting

(SARs) which is a regulatory requirement for Investment Banks and Crypto currency trading firms . This contributes to broader efforts in monitoring and mitigating illicit activities within the cryptocurrency domain. By applying these methods, the research aims to improve awareness and management of crypto vulnerabilities. The enhanced analytics pipeline not only identifies existing vulnerabilities but also predicts potential future threats, providing a proactive approach to crypto security. This is achieved through the integration of ML models that classify vulnerabilities based on their characteristics and severity, and SNA techniques that map out the networks of actors involved in exploits, helping to identify and disrupt coordinated attacks. The paper is structured such that Section 2 of this paper reviews related works, highlighting previous studies and advancements in the field of crypto vulnerability analysis. This includes an examination of automated vulnerability analysis research as well as the contributions of various researchers in understanding, identifying and mitigating these vulnerabilities. Section 3 describes the methodology employed in this research. This section details the data collection process, the development of the analytics pipeline, the integration of NLP and ML techniques, and the application of SNA. Section 4 presents the results of the study, showcasing the effectiveness of the developed pipeline in identifying, classifying, and mitigating crypto vulnerabilities. This section includes an analysis of the performance metrics of the ML models, the insights gained from SNA, and the overall impact of the enhanced analytics on crypto vulnerability management. Section 5 concludes the paper, summarising the key findings and contributions of the research. This section also discusses potential directions for future research.

2. Literature Review

A vulnerability in a system or device allows unauthorised access, privilege elevation, or denial of service. The increasing complexity of crypto systems has necessitated robust methods to identify and classify these vulnerabilities. Both industry and academia have engaged in this battle, contributing significantly to the understanding of crypto vulnerabilities and their impacts. Central to these challenges is the identification and classification of vulnerabilities

within blockchain networks, which are critical to ensuring the security and integrity of decentralised financial systems. Recent advancements in machine learning , social network analysis, and web intelligence have provided innovative approaches to detecting and mitigating these vulnerabilities. This literature review fuses findings from seven pivotal studies that explore these methodologies in depth. Lorenz et al. (2020) address the pervasive issue of money laundering in the Bitcoin blockchain, particularly under conditions of label scarcity. Traditional supervised ML models struggle in such environments due to the insufficient labelled data necessary for training. To overcome this, the authors propose an active learning (AL) framework that strategically selects the most informative samples for manual annotation by experts. Their approach demonstrated that using merely 5% of the labels, the AL-enhanced model could achieve performance comparable to a fully supervised baseline, significantly reducing the annotation burden without compromising accuracy. Complementing this, Lin et al. (2024) introduce DenseFlow, an innovative framework designed to identify and trace money laundering activities within Ethereum transaction graphs. DenseFlow leverages anomaly detection by identifying dense subgraphs that signify illicit patterns and employs the maximum flow algorithm to map the movement of funds from source accounts through laundering pathways. Their study highlights the limitations of unsupervised anomaly detection methods when applied to real-world blockchain data, underscoring the necessity for more nuanced ML strategies that incorporate both graph-based features and more concentrated machine learning principles.

Roumeliotis et al. (2024) explore the application of large language models (LLMs) and natural language processing (NLP) techniques in analysing cryptocurrency-related sentiments on social media, particularly Twitter. By fine-tuning models such as GPT-4, BERT, and FinBERT, the study demonstrates the value of these models in accurately capturing and predicting the sentiment dynamics that influence cryptocurrency markets. Their findings suggest that sentiment analysis is a potent tool for forecasting market trends, as evidenced by the correlation between positive sentiments and price surges, and negative sentiments with market downturns. Similarly, Beck et al. (2024) develop an online data mining system that

correlates cryptocurrency news with tweets, aiming to predict the popularity of news articles based on their social media engagement. Utilising sentence-level embeddings and clustering algorithms, their study identifies distinct topics within Crypto Twitter and correlates sentiment shifts with significant real-world events, such as the FTX bankruptcy. This approach not only aids in real-time monitoring of public sentiment but also enhances the predictive accuracy of market movements based on social media signals.

AutoMESC, introduced by Soud et al. (2025), represents a significant stride in automating the identification and classification of Ethereum smart contract vulnerabilities and their fixes. The framework integrates multiple security analysis tools to label vulnerabilities, thereby creating a comprehensive and continuously updated dataset. This automated approach mitigates the limitations of existing datasets, which often suffer from incompleteness, lack of diversity, and outdated information. AutoMESC's ability to handle both Solidity and Vyper contracts and its incorporation of Common Weakness Enumeration (CWE) classifications provide a robust foundation for training ML models aimed at vulnerability detection and mitigation. Also related, Hasnaoui et al. (2024) propose an AI-driven opcode-based classification system for smart contracts. By analysing opcode patterns, their methodology distinguishes between primary smart contracts, oracle contracts, and bridge contracts, each serving different functions within the blockchain ecosystem. The use of ensemble ML models, including Random Forests and Convolutional Neural Networks (CNNs), demonstrates high accuracy in classification tasks. This study emphasises the importance of opcode-level analysis in understanding smart contract functionalities and preemptively identifying potential security risks.

Kang et al. (2024) delve into the intricacies of Crypto Twitter, utilising social network analysis to uncover the prevalence of bot activities and their impact on sentiment dynamics. By constructing tweet and user interaction networks, the study identifies coordinated bot networks that spread spam, manipulate sentiments, and amplify specific narratives within the cryptocurrency community. The findings reveal that bot-driven activities significantly distort public sentiment, thereby influencing market perceptions and decisions. The study

also revealed discovered sentiment indicators that point to real-life incidents in the crypto world, such as signals of distrust associated with the FTX firm in November of 2022 before fraudulent behaviour was discovered and the firm became bankrupt. This underscores the necessity and use for advanced SNA techniques to detect and mitigate the influence of malicious bots in cryptocurrency-related social media environments.

The integration of ML models in predictive analytics is further exemplified by Beck et al. (2024), who employ sequence-to-sequence (seq2seq) neural networks and autoregressive models to forecast the popularity of cryptocurrency news articles on Twitter. Their approach utilises temporal features and content-contextual indicators to predict engagement metrics, achieving substantial improvements over baseline models. The real-time deployment of these models facilitates responsive adjustments to content dissemination strategies, enhancing the ability of stakeholders to anticipate and react to market trends based on social media signals. A comparative assessment of various ML models across these studies highlights distinct advantages of certain approaches. Random Forests and CNNs emerge as particularly effective in handling blockchain transactions and news interactions. The recursive nature of seq2seq models also proves beneficial in capturing long-term dependencies and sequential patterns in tweet engagements. Moreover, the incorporation of active learning paradigms, as demonstrated by Lorenz et al. (2020), significantly enhances models in scenarios with limited labelled data, a common challenge in cryptocurrency vulnerability datasets.

The reviewed literature underscores the pivotal role of machine learning, natural language processing, and social network analysis in identifying and classifying vulnerabilities within cryptocurrency ecosystems. Active learning frameworks, such as those proposed by Lorenz et al. (2020), and graph-based models like DenseFlow (Lin et al., 2024) provide robust methodologies for detecting illicit activities under constraints of label scarcity and complex transaction patterns. Meanwhile, sentiment analysis models and real-time predictive systems facilitate a deeper understanding of market dynamics influenced by social media

sentiments. Social network analysis further enhances the detection of malicious bot activities that distort public sentiment and market perceptions.

Future research should focus on integrating these diverse methodologies into unified frameworks that leverage the strengths of each approach. As the blockchain landscape continues to evolve, the synergy between machine learning, NLP, and social network analysis will be instrumental in preemptively identifying vulnerabilities and ensuring the resilience of decentralised financial systems.

3. Methodology

4. Results

5. Conclusion

Bibliography

1. "*rekt.news*" 2025, viewed 10 January 2025, <https://rekt.news/>.
2. “Blockchain-DLT-Attacks-and-Weaknesses-Enumeration” 2025 , viewed 10 January 2025,<https://docs.google.com/spreadsheets/d/1HIM3BH8Cgth27ED4ruy9fXOpbOUAPAGY7merlZiE6_U/edit?gid=1028635246#gid=1028635246
3. www.3 https://saifmohammad.com/WebPages/nrc-vad.html >.
4. Johannes Beck, Roberta Huang, David Lindner, Tian Guo, Zhang Ce, Dirk Helbing, and Nino Antulov-Fantulin. 2019. Sensing Social Media Signals for Cryptocurrency News. In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 1051–1054. https://doi-org.tudublin.idm.oclc.org/10.1145/3308560.3316706
5. Inwon Kang, Maruf Ahmed Mridul, Abraham Sanders, Yao Ma, Thilanka Munasinghe, Aparna Gupta, and Oshani Seneviratne. 2024. Deciphering Crypto Twitter. In Proceedings of the 16th ACM Web Science Conference (WEBSCI '24). Association for Computing Machinery, New York, NY, USA, 331–342. https://doi.org/10.1145/3614419.3644026
6. Hasnaoui, M. Zrikem and R. Elassali, "AI-Driven Opcode-Based Smart Contract Classification," 2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC), Marrakech, Morocco, 2024, pp. 1-6,
7. Lin, D., Wu, J., Yu, Y., Fu, Q., Zheng, Z., & Yang, C. (2024). DenseFlow: Spotting Cryptocurrency Money Laundering in Ethereum Transaction Graphs. Proceedings of the ACM on Web Conference 2024.
8. Joana Lorenz, Maria Inês Silva, David Aparício, João Tiago Ascensão, and Pe dro Bizarro. 2020. Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity. In ACM International Conference on AI in

Finance (ICAIF '20),October 15–16, 2020, New York, NY, USA. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3383455. 3422549

9. Roumeliotis, K.I.; Tselikas, N.D.; Nasiopoulos, D.K. LLMs and NLP Models in Cryptocurrency Sentiment Analysis: A Comparative Classification Study. *Big Data Cogn. Comput.* 2024.
10. M. Soud, I. Qasse, G. Liebel and M. Hamdaqa, "AutoMESC: Automatic Framework for Mining and Classifying Ethereum Smart Contract Vulnerabilities and Their Fixes," 2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Durres, Albania, 2023, pp. 410-417, doi: 10.1109/SEAA60479.2023.00068.