A Dataset-Driven Validation of Structural Constraints in Network Traffic–Based Detection of Financial Crimes

Muhammad Nuraddeen Ado[1,2], Dr. Shafi'i M Abdulhamid[1,3,4] and Prof. Idris Ismaila[1,4]

[1]Department of Cyber Security, ACETEL, National Open University of Nigeria, Abuja, Nigeria

[2]Department. Of Information Sciences, Federal University, Dutsin-Ma; Katsina State, Nigeria

[3]Department of Cyber Security, Community College Qatar, Doha, Qatar

[4]Department of Cyber Security, Federal University of Technology, Minna, Niger State, Nigeria

DOI: https://doi.org/10.51244/IJRSI.2026.1313CS002

Received: 16 February 2026; Accepted: 22 February 2026; Published: 03 March 2026

ABSTRACT

Modern financial ecosystems are increasingly exposed to money laundering, fraud, phishing, and transaction obfuscation within encrypted and decentralized network environments. While prior research identifies detection challenges conceptually, limited work empirically validates how structural constraints manifest across real-world network datasets. This study presents a dataset-driven validation of structural constraints in network traffic–based detection of financial crime using three progressively scaled real-time datasets: NetTran3, NetTran4, and NetTran5. Quantitative analysis shows that encrypted traffic obscures 22.03% of transactions in NetTran3, rises to 39.29% in NetTran4, and declines to 18.24% in NetTran5, reflecting evolving encryption dynamics. Tumbling services—used for transaction obfuscation—are observed in 34.42%, 34.01%, and 40.22% of transactions, respectively, indicating increasing transactional fragmentation. Additional constraints include scalability pressure, cross-chain complexity, adversarial manipulation, and data imbalance. Rather than benchmarking detection algorithms, this study systematically evaluates how these infrastructural characteristics intensify detection difficulty as dataset size and complexity increase. The findings provide empirical grounding for literature-identified constraints and establish a structured foundation for advancing robust, scalable, and privacy-aware financial crime detection systems.

Keywords: Financial Crime Detection, Network Traffic Analysis, Structural Constraints , Dataset-Driven Validation, Encrypted Traffic.

INTRODUCTION

Financial crime detection within modern digital ecosystems is increasingly constrained by structural characteristics of network environments. The proliferation of encrypted traffic, transaction obfuscation mechanisms such as tumbling services, cross-chain financial flows, scalability pressures, adversarial manipulation, and data imbalance has significantly complicated network traffic–based detection efforts. While prior studies conceptually discuss these challenges, limited research empirically validates how such structural constraints manifest across real-world datasets of varying scale and complexity. This study addresses that gap by conducting a dataset-driven validation of structural constraints using three progressively scaled real-time network datasets—NetTran3, NetTran4, and NetTran5—to examine how infrastructural properties evolve and influence detection feasibility.

Modern financial systems are increasingly exposed to sophisticated financial crimes, including money laundering, fraud, phishing schemes, and insider trading. These illicit activities frequently exploit anomalies in network traffic patterns, making the detection of such deviations crucial. The rise of digital financial technologies and online platforms has amplified these vulnerabilities, creating intricate transactional ecosystems that overwhelm traditional fraud detection mechanisms. Moreover, the absence of centralized oversight in digital

systems exacerbates the challenge, underscoring the need for robust anomaly detection techniques to prevent exploitation by malicious actors [1], [2].

To address these vulnerabilities, advanced methodologies like machine learning and graph- based analyses have emerged as vital tools in detecting financial anomalies. These techniques excel in identifying suspicious patterns, such as abrupt changes in transaction volumes, irregular use of financial instruments, or deviations in account behaviors. Forensic methods incorporating models like Random Forest and Graph Convolutional Networks (GCNs) have proven particularly effective in detecting and classifying anomalous transactions. Furthermore, visualization tools like address graphs and flow graphs enhance the analytical process, offering intuitive pathways to trace and map illicit fund transfers [3], [4].

Despite these technological advancements, significant challenges persist in anomaly detection for financial fraud. The vast volume of financial transactions processed daily presents scalability issues, often overwhelming traditional forensic tools with computational demands. Additionally, sophisticated evasion techniques, such as layering and the mixing of funds across multiple accounts, obscure transaction trails, complicating efforts to identify anomalies. Regulatory constraints further add complexity, as inconsistent frameworks across jurisdictions hinder a standardized approach to monitoring and reporting financial crimes [5], [6].

These challenges necessitate a shift towards more adaptive and scalable detection systems capable of addressing the evolving nature of financial fraud. Therefore this study explores the validity of identified challenges to the advances in the detection of financial crimes through network traffic's anomaly detection.

To investigate and validate the challenges that hinder advancements in network traffic-based anomaly detection techniques for detecting financial crimes, focusing on real-time data scenarios, this paper seeks to analyze the factors that hinder the effectiveness and progress of network traffic-based anomaly detection in financial crime detection. Specifically, it aims to assess the influence of anonymity, pseudonymity, and decentralization on detection techniques, examine the scalability of models under high transaction volumes, and analyze the effects of obfuscation strategies like mixing and tumbling services. The study also seeks to quantify the prevalence of false positives and negatives, test model resilience against adversarial attacks, address the impact of imbalanced datasets on reliability, and explore the complexities introduced by cross-chain transactions.

Unlike model-centric studies that benchmark classification accuracy, this work validates the infrastructural constraints that limit detection advancement.

Problem Statement

The rapid digitization of financial systems and the proliferation of decentralized and encrypted network infrastructures have significantly increased the complexity of detecting financial crimes such as money laundering, fraud, and transaction obfuscation. While existing literature identifies key detection challenges—including encrypted traffic, tumbling services, scalability pressure, cross-chain transactions, adversarial manipulation, and data imbalance—these constraints are often discussed conceptually without systematic empirical validation using real-world network datasets.

As network traffic grows in scale and complexity, it remains unclear how these structural constraints evolve and how they collectively influence the feasibility of network traffic–based financial crime detection. The absence of dataset-driven validation creates a gap between theoretical identification of challenges and their measurable manifestation in operational environments. Consequently, there is a need for a structured empirical investigation that validates how literature-identified constraints materialize across progressively scaled real-time datasets.

Aim and Specific Objectives of the Study

Aim

The aim of this study is to empirically validate structural constraints affecting network traffic–based detection of financial crime through a dataset-driven analysis using progressively scaled real-time network datasets.

Specific Objectives

i. To identify and synthesize key structural constraints affecting network traffic–based detection of financial crime through a structured review of existing literature.

ii. To empirically analyze the manifestation and evolution of these constraints—across progressively scaled real-time datasets (NetTran3, NetTran4, and NetTran5).

iii. To evaluate how these structural characteristics influence detection feasibility and provide an empirically grounded foundation for advancing robust, scalable, and privacy-aware financial crime detection systems.

Scope of the Study

This study does not propose, benchmark, or optimize specific anomaly detection or machine learning algorithms. Instead, its objective is to systematically validate structural and infrastructural challenges confronting network-based financial crime detection. Following a structured literature review to identify core constraints—such as encryption, tumbling services, scalability pressure, adversarial behavior, cross-chain complexity, and class imbalance—the study empirically examines how these challenges manifest across progressively scaled real-time datasets (NetTran3, NetTran4, and NetTran5). The contribution is therefore analytical and validation-oriented rather than algorithmic, focusing on dataset-level structural properties that shape detection feasibility.

Key Contributions of the Study

The key contributions of this study are as follows:

a. Structured Identification of Detection Constraints

This paper synthesizes prior literature to systematically identify core structural constraints affecting network traffic–based financial crime detection, including encryption prevalence, tumbling services, scalability pressure, adversarial behavior, cross-chain complexity, and data imbalance.

b. **Dataset-Driven Empirical Validation**

Empirically validates how these constraints manifest across three progressively scaled real-time datasets (NetTran3, NetTran4, and NetTran5), providing quantitative evidence of their structural evolution.

c. **Progressive Environment Analysis**

Demonstrates how increasing dataset scale and transactional complexity intensify infrastructural limitations confronting detection systems.

d. **Constraint-Oriented Analytical Framework**

Establishes a structured evaluation framework for examining environmental limitations prior to algorithmic deployment, thereby informing the design of more adaptive and scalable detection systems.

e. **Foundational Grounding for Future Detection Advances**

Provides empirical grounding that supports the development of robust, privacy-aware, and scalable financial crime detection solutions aligned with real-world network conditions.

The remainder of the paper is structured as follows: Section 2 presents the review of relevant literature. Section 3 describes the methodology used for the validation of structural challenges, followed by the presentation and analysis of results in Section 4. Section 5 concludes the paper.

LITERATURE REVIEW

Analyzing network traffic is crucial for detecting financial fraud by identifying anomalies in transaction patterns. Machine learning techniques, such as Random Forest and Neural Networks, effectively detect transaction irregularities while adapting to the complexities of modern financial networks, addressing the increasing demand for cybersecurity in FinTech ecosystems [7].

Building on these foundations, the evolution of machine learning techniques has significantly enhanced the performance of anomaly detection systems. For instance, [8] employed the Louvain algorithm within graph databases, achieving a 99.77% accuracy in detecting fraudulent transactions. Similarly, ensemble learning methods such as Gradient Boosting and CatBoost have proven effective in reducing false positives, a critical barrier in fraud detection. Together, these advancements enable real-time monitoring and offer scalable solutions, addressing the pressing need for systems capable of managing large-scale and dynamic transactional environments.

Extending these capabilities, graph-based techniques have become instrumental in capturing the relational and temporal behaviors of transactional data. The study by [9] introduced innovative methods like reduced egonets and random walks for anomaly detection in transaction graphs. Despite these advancements, several challenges persist, necessitating further innovation. The scalability of fraud detection systems remains a key concern due to the vast volume of financial transactions requiring analysis. Moreover, the lack of labeled datasets limits the effectiveness of supervised models, as noted by [10] in their examination of big data's role in fraud detection. Emerging approaches, including federated learning and privacy-preserving technologies, offer promising solutions by enabling secure, distributed, and collaborative analyses that address both scalability and data privacy concerns.

Looking ahead, the integration of advanced technologies is bridging critical gaps in anomaly detection. Deep learning frameworks, such as RNNs and CNNs, have demonstrated their ability to capture temporal and spatial features within transactional data. Meanwhile, federated learning frameworks, as discussed by [11], ensure data privacy while facilitating collaborative model training. These advancements highlight the need for interdisciplinary research to develop scalable, privacy-compliant systems, strengthening financial networks against fraud..

Challenges In Network Traffic-Based Detection

Based on the reviewed literature, the challenges to advances in Network Traffic-Based Detection of Financial Crimes include:

Anonymity and Pseudonymity of Secure Transactions

[12] highlighted that the structural and temporal behaviors of financial transactions, such as those in Ethereum, create barriers for conventional detection methods, particularly when users leverage the decentralized nature of these platforms to obscure their activities.

Decentralized Nature of Financial Institutions

Dumitrescu et al. (2022) noted that decentralized platforms allow bad actors to exploit fragmented oversight, complicating efforts to identify suspicious patterns in financial transactions. This lack of centralized control undermines collaborative fraud detection, requiring innovative approaches to manage decentralized data networks.

Scalability Issues

Scalability is a pressing challenge as financial networks continue to grow. [10] emphasized that the increasing volume of transactions overwhelms existing detection systems, which struggle to process large-scale data efficiently. These limitations hinder real-time anomaly detection, especially in high-frequency trading and cross-border transactions .

Use of Mixing and Tumbling Services

These services obfuscate transaction origins and destinations, making it difficult to trace money laundering activities. [7] indicated that such practices are common in financial systems posing a significant barrier to network traffic analysis and fraud detection efforts .

False Positives and False Negatives in Detection

[8] reported that despite advancements in machine learning, models still face challenges in achieving high precision and recall, especially in detecting subtle fraudulent activities.

Adversarial Attacks on Detection Models

[10] observed that attackers often exploit vulnerabilities in machine learning algorithms, introducing adversarial samples that deceive detection systems. These attacks degrade the accuracy of anomaly detection models and highlight the need for robust defenses against evolving threats .

Imbalanced Datasets

[12] noted that the rarity of fraudulent transactions relative to legitimate ones skews datasets, leading to biased models that struggle to detect outliers effectively.

Cross-Chain Transactions

[10] highlighted that the movement of assets across chains makes it harder to trace the flow of funds and detect fraudulent activities. These transactions exploit gaps in monitoring mechanisms, requiring sophisticated tracking systems to bridge the visibility gaps.

METHODOLOGY

This section presents the dataset-driven structural constraint validation framework adopted in this study. The objective is not to develop or benchmark anomaly detection models, but to empirically examine how literature-identified structural constraints manifest across progressively scaled real-world network datasets. This study does not evaluate detection algorithm performance but instead quantifies how structural characteristics of network datasets intensify detection difficulty. Following the structured review of prior research, key constraints—including encryption prevalence, tumbling services, scalability pressure, adversarial manipulation, cross-chain transaction complexity, and data imbalance—were operationalized as measurable dataset-level indicators. These indicators were then systematically evaluated across NetTran3, NetTran4, and NetTran5 to analyze how infrastructural characteristics evolve with increasing scale and transactional complexity.

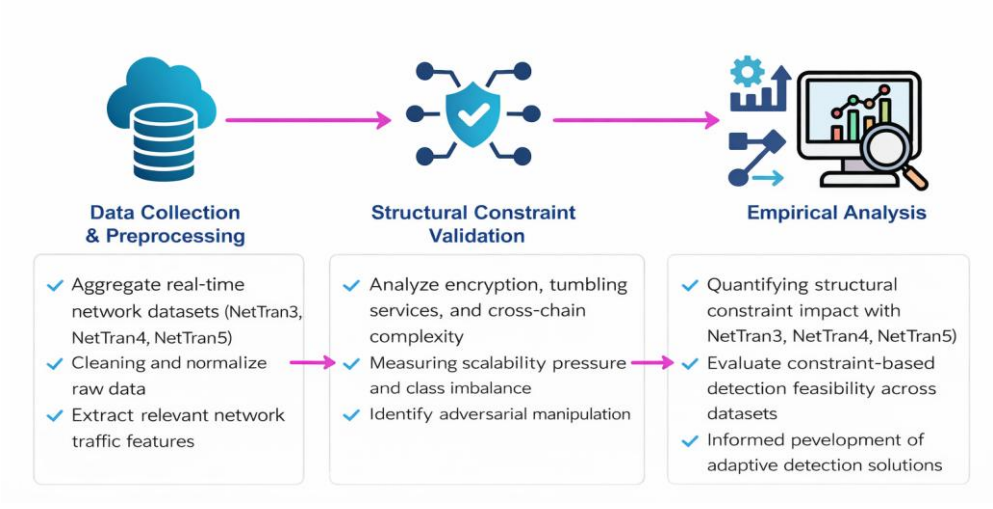

Figure 1: Dataset-Driven Structural constraint Validation Framework

Methodological Workflow

The evaluation framework adopted in this study follows a structured four-stage analytical workflow:

Stage 1: Extraction of Structural Challenges from Literature

Key constraints affecting network-based financial crime detection (e.g., encryption prevalence, tumbling services, scalability pressure, adversarial manipulation, cross-chain complexity, and data imbalance) were systematically identified through prior studies.

Stage 2: Mapping to Measurable Network Indicators

Each identified challenge was operationalized into quantifiable indicators within network traffic data (e.g., proportion of encrypted payloads, frequency of tumbling-related transaction patterns, transaction density growth rates).

Stage 3: Quantitative Assessment Across Datasets

The defined indicators were measured independently across three progressively scaled real-time datasets: NetTran3, NetTran4, and NetTran5.

Stage 4: Comparative Structural Analysis Across Scales

A cross-dataset comparison was conducted to examine how structural constraints evolve as network size and complexity increase, thereby validating their impact on detection difficulty.

Data Capture

Real-time network traffic data was captured using Wireshark, ensuring a diverse range of transaction types and scenarios. The three tables (Table 1, Table 2 and Table 3) for NetTran3, NetTran4, and NetTran5 provide descriptive statistics for the datasets.

Dataset Characteristics:

NetTran3: [13], [16]

Table 1: Descriptive Statistics of NetTran3

	No.	**Time**	**Length**
count	36986.00	36986.00	36986.00
mean	18493.50	995.58	375.61
std	10677.08	557.94	497.82
min	1.00	0.00	42.00
max	36986.00	1999.35	1434.00

Table 1 summarizes 36,986 transactions, analyzing No. (index), Time, and Length variables. Time has a mean of 995.58, a standard deviation of 557.94, and right-skewness. Length, with a mean of 375.61 and high variability, shows a skewed distribution, clustering below the mean with broad-ranging values.

NetTran4 Dataset:

NetDSet4: [14], [16]

Table 2: Descriptive Statistics of NetTran4

	No.	**Time**	**Length**
count	114290.00	114290.00	114290.00
mean	57145.50	2551.82	477.94
std	32992.83	1614.66	552.58
min	1.00	0.00	42.00
max	114290.00	5344.25	1514.00

Table 2 summarizes 114,290 transactions, analyzing No., Time, and Length variables. Time shows a mean of 2551.82, broad range (0–5344.25), and positive skewness. Length, with a mean of 477.94, also exhibits right-skewness and high variability. Both variables display broader distributions and significant spread compared to smaller datasets.

NetTran5:

NetDSet5: [15], [16]

Table 3: Descriptive Statistics of NetTran5

	No.	**Time**	**Length**
count	225688.00	225688.00	225688.00
mean	112844.50	3742.24	473.08
std	65150.66	2368.17	574.80
min	1.00	0.00	42.00
max	225688.00	8172.38	1514.00

Table 3 summarizes 225,688 transactions, analyzing No., Time, and Length variables. Time shows a mean of 3742.24 and significant variability with right-skewness, ranging up to 8172.38. Length, with a mean of 473.08, also displays a broad, right-skewed distribution, highlighting variability and the presence of high outliers in both metrics.

As the datasets progress from NetTran3 (36,986 entries) to NetTran4 (114,290 entries) and NetTran5 (225,688 entries), both size and traffic variability increase significantly. The Time and Length variables consistently display right-skewed distributions, broader ranges, and rising mean values, indicating increasingly diverse traffic characteristics. Time shows higher maximum values and standard deviations, reflecting bursts in traffic duration, while Length exhibits greater dispersion, likely due to varying transaction behaviors. These trends highlight the growing complexity in network traffic data, underscoring the challenges machine learning models face in detecting financial crime patterns in such dynamic and variable environments.

Experiments On Key Challenges

The following section presents empirical validation results for each identified constraint. To validate the key findings above, the three sets of network traffic datasets were analyzed:

Result

The result of the analysis for the findings from the literature review above - Anonymity and Pseudonymity of Financial Transactions, Decentralized Nature of Financial, Scalability Issues, Use of Mixing and Tumbling Services, False Positives and False Negatives in Detection, Adversarial Attacks on Detection Models, Imbalanced Datasets and Cross-Chain Transactions are provided below:

Anonymity and Pseudonymity of Financial Transactions:

Anonymity and pseudonymity are defining features of many modern financial networks— especially within decentralized systems like cryptocurrencies—and present a significant obstacle to detecting and prosecuting fraudulent activities. These systems often allow users to conduct transactions without revealing identifiable information, making it difficult for law enforcement, regulators, and analysts to trace illicit behavior.

Quantitative dataset analysis demonstrates this constraint as follows: in the analysis of the NetTran3 dataset. Captured over a 40-minute period using Wireshark, NetTran3 contains 36,986 entries, of which 8,148 (or 22.03%) are classified as "protected payloads" in the "Info" column. These entries represent encrypted traffic transmitted through secure protocols such as TLSv1.2 and QUIC, which are designed to protect user privacy by obscuring the content and endpoints of transactions. While essential for securing legitimate communications, these encryption layers also shield illicit financial activities, such as money laundering and terrorism financing, from forensic scrutiny. One major implication of this encryption is that only the owner of a specific wallet or address has access to the complete transaction history. Fraud analysts are therefore forced to rely on network traffic patterns—such as timing, volume, and destination addresses—rather than the actual content of transactions. This limited visibility significantly reduces the accuracy of detection models, as fraudulent activities can easily mimic the behavior of legitimate ones. The frequent occurrence of "protected payloads" across financial traffic data reflects a growing trend: encryption and pseudonymity are not just protective measures, but potential enablers of financial crime.

A broader comparative analysis of three real- time network traffic datasets—NetTran3, NetTran4, and NetTran5—offers deeper insights into this issue. The proportion of protected packets varies across datasets and timeframes:

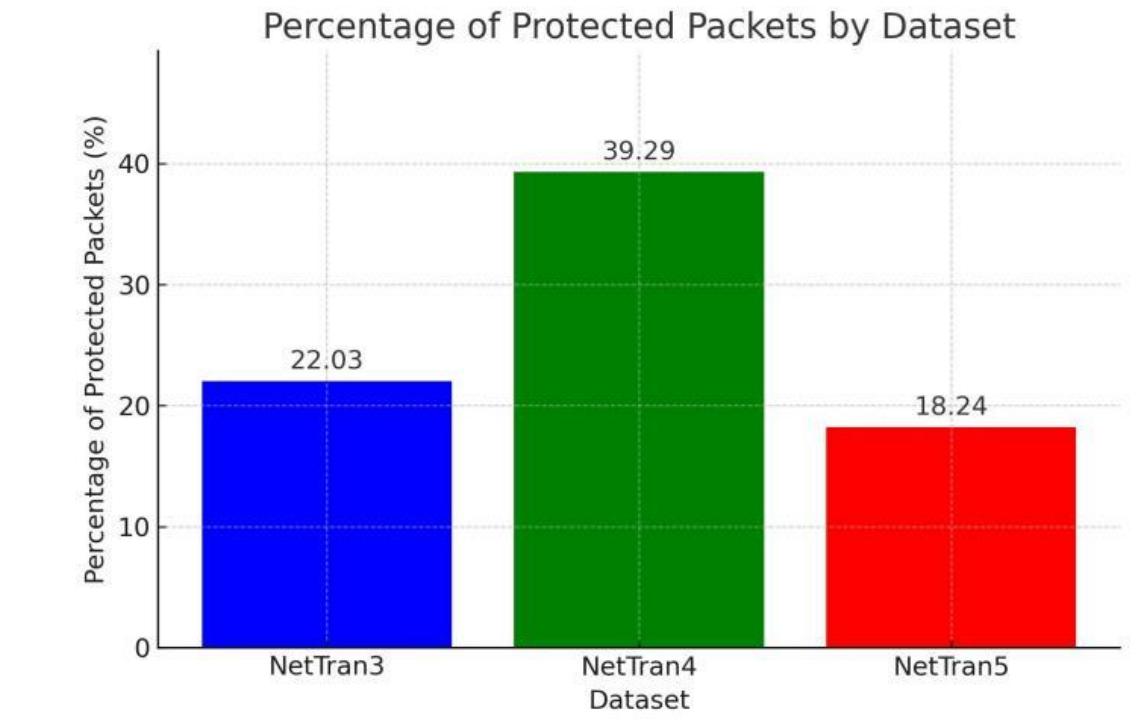

Figure 2: Percentage of Protected Packets

NetTran4, which had the longest intermediate capture duration, recorded the highest proportion of protected packets (39.29%), indicating a possible spike in encrypted or sensitive transactions during that period. Conversely, NetTran5, with the longest capture time, had a lower encrypted traffic share (18.24%), suggesting a more varied mix of secure and non- secure communication. These shifts suggest that factors beyond time—such as user behavior, transaction types, and prevailing network conditions—strongly influence the presence and volume of encrypted traffic.

Despite fluctuations, the consistent presence of protected packets across all datasets reinforces their relevance in financial crime investigations. While encryption is crucial for ensuring privacy in financial systems, it simultaneously serves as a shield that can obscure malicious activity. This dual nature of encryption complicates network- based anomaly detection, as it forces fraud detection systems to operate with partial visibility and limits their ability to distinguish between legitimate and fraudulent transactions. As a result, the effectiveness of fraud detection systems—especially those reliant on network traffic analysis—can be severely compromised in pseudonymous and encrypted environments. Addressing this issue requires innovative approaches that balance privacy protection with the ability to detect and deter financial crimes.

Decentralized Nature of Some Financial Systems:

Fraud detection in decentralized financial networks is challenging due to the lack of central oversight, making it difficult to establish monitoring responsibility or jurisdiction. The NetTran4 and NetTran5 datasets highlight this complexity, with 39.29% of NetTran4 entries marked as "protected payloads," showcasing significant encryption that conceals transaction details.

This encryption aligns with the decentralized design of financial systems, where only sender and receiver nodes access transaction specifics, minimizing centralized monitoring. Additionally, the datasets reveal diverse IP addresses and protocols like TCP and QUIC, indicative of peer-to-peer, multi-node architectures. Transactions span multiple jurisdictions without regulatory control, further complicating fraud detection and oversight. These decentralized, encrypted, and globally distributed features undermine traditional monitoring approaches, requiring more adaptive solutions for effective fraud prevention.

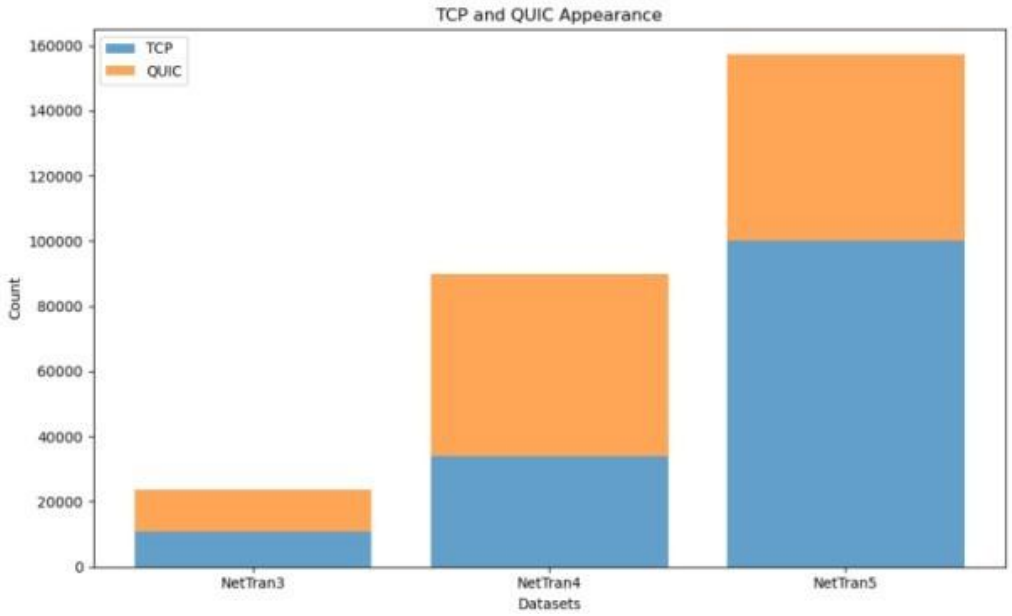

Figure 3: TCP and QUIC Protocols

Figure 3 above illustrates the prevalence of TCP and QUIC protocols across NetTran3, NetTran4, and NetTran5, emphasizing their role in decentralized financial networks. The increase in QUIC, optimized for encrypted peer-to-peer connections, alongside TCP, reflects the multi- node architecture where independent nodes validate transactions without central oversight. This decentralized, globally distributed structure complicates fraud detection, as jurisdiction-free, dynamic traffic patterns challenge traditional anomaly detection and enforcement methods.

Scalability Issues:

The scalability of fraud detection systems is a critical concern, especially as blockchain networks grow. Fraud detection algorithms often require complex computations across a large number of transactions, which can significantly slow down the detection process. This challenge is exacerbated in popular networks like Bitcoin and Ethereum, where transaction volumes are high.

NetTran4 and **NetTran5** can indeed support the argument that scalability is a significant challenge in fraud detection within large blockchain networks. The datasets contain a substantial volume of entries, representing network transactions with a high frequency of encrypted data flows. This encryption is essential for security but also adds complexity to monitoring efforts, as decrypting and analyzing each transaction would require intensive computational resources. Furthermore, the diverse protocols and IP addresses involved suggest that these entries are likely part of a distributed network structure, akin to blockchain. As blockchain networks grow, so does the transaction volume, resulting in exponentially more data points to monitor, analyze, and validate in real time. Fraud detection algorithms, therefore, need to process this expanding dataset quickly, which demands robust computing power and advanced algorithms that can operate efficiently at scale.

Additionally, the presence of protected, encrypted data (about 39.29% in **NetTran4**) implies that each transaction has multiple cryptographic layers that require decryption or advanced pattern analysis to detect suspicious behavior. This cryptographic complexity is particularly burdensome in large networks like Bitcoin or Ethereum, where transaction throughput can be immense. As these datasets mirror characteristics of a real blockchain network, they demonstrate how high transaction volumes, alongside the need for rapid and secure analysis, can overwhelm traditional fraud detection methods. Scaling these methods effectively would require specialized infrastructure capable of handling high-speed, distributed data across diverse protocols, reflecting a fundamental challenge to ensuring timely fraud detection in expanding blockchain ecosystems.

Datasets like NetTran4 and NetTran5 illustrate the challenges posed by high transaction volumes and encryption, which complicates anomaly detection by requiring advanced pattern recognition or decryption. The diverse protocols and distributed architectures in these datasets highlight the difficulty of real-time monitoring as networks grow. For instance, NetTran4's 39.29% encrypted data showcases the computational strain of analyzing multi-layered encryption in financial ecosystems. These findings emphasize the need for scalable solutions and advanced algorithms to address the limitations of traditional fraud detection methods in complex, high-speed networks.

Use of Mixing and Tumbling Services:

Fraudsters often use mixing services (also known as tumbling services or layering) to obscure the source of their funds. These services break down cryptocurrencyinto smaller transactions and route them through different addresses, making it extremely difficult to track the origin of the funds. This obfuscation method poses a significant challenge for traditional fraud detection techniques.

To determine the possibility of patterns typical of mixing and tumbling services, the following steps were taken:

i. Sort the data by time to identify rapid transactions.

ii. Group by source addresses and check if a source is sending multiple transactions within a small time window (e.g., less than 5 seconds) to various destinations.

iii. Filter by transaction size to focus on smaller transactions, and observe variations in the lengths of these transactions.

Based on this, the python code was used specifically for:

i. Sorting and Grouping: The dataset is sorted by Time, and transactions are grouped by Source (the sender's address).

ii. Filtering: The script filters for smaller transactions (below average Length) and checks if the time difference between consecutive transactions is within the threshold (e.g., 5 seconds).

iii. Multiple Destinations: The code checks if a single source is sending funds to multiple destinations, a typical characteristic of tumbling services.

Visualization: The results are visualized with a scatter plot, which can help you see how these transactions behave over time.

Fraudsters use mixing services to obscure fund origins by breaking transactions into smaller amounts sent rapidly to multiple destinations. To detect this, data was sorted by time, grouped by source, and filtered for small transactions sent within short intervals. The process identified patterns typical of tumbling services by checking for rapid, multi-destination transactions and visualized results with scatter plots to analyze their behavior over time.

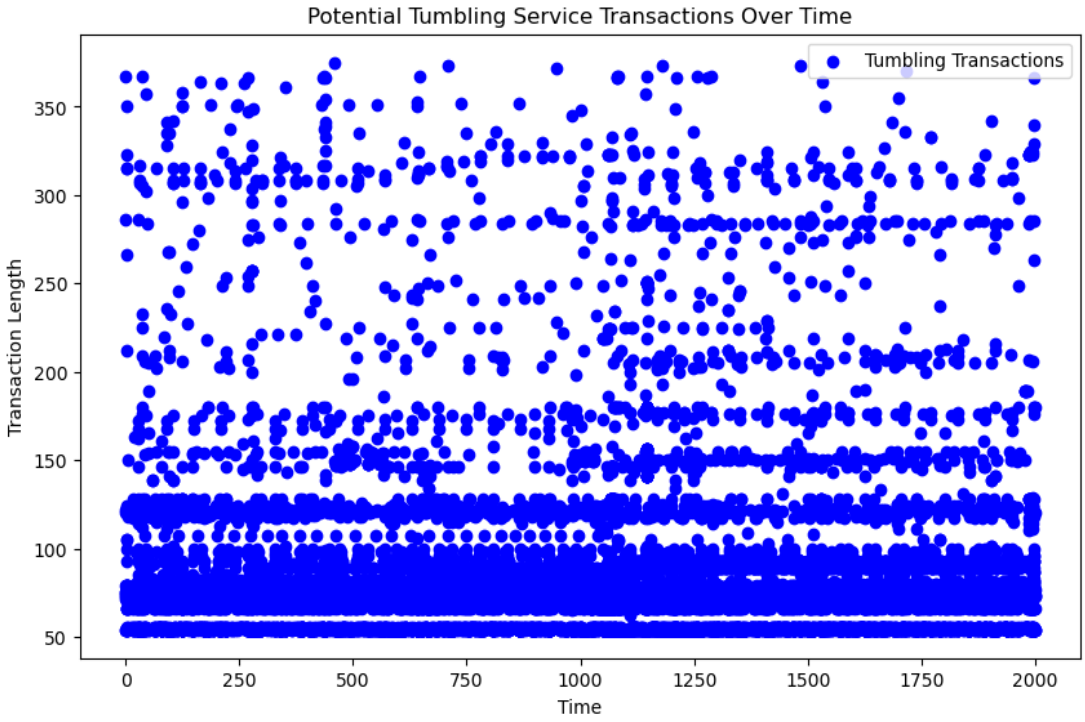

Figure 3: Tumbling Services in NetTran3

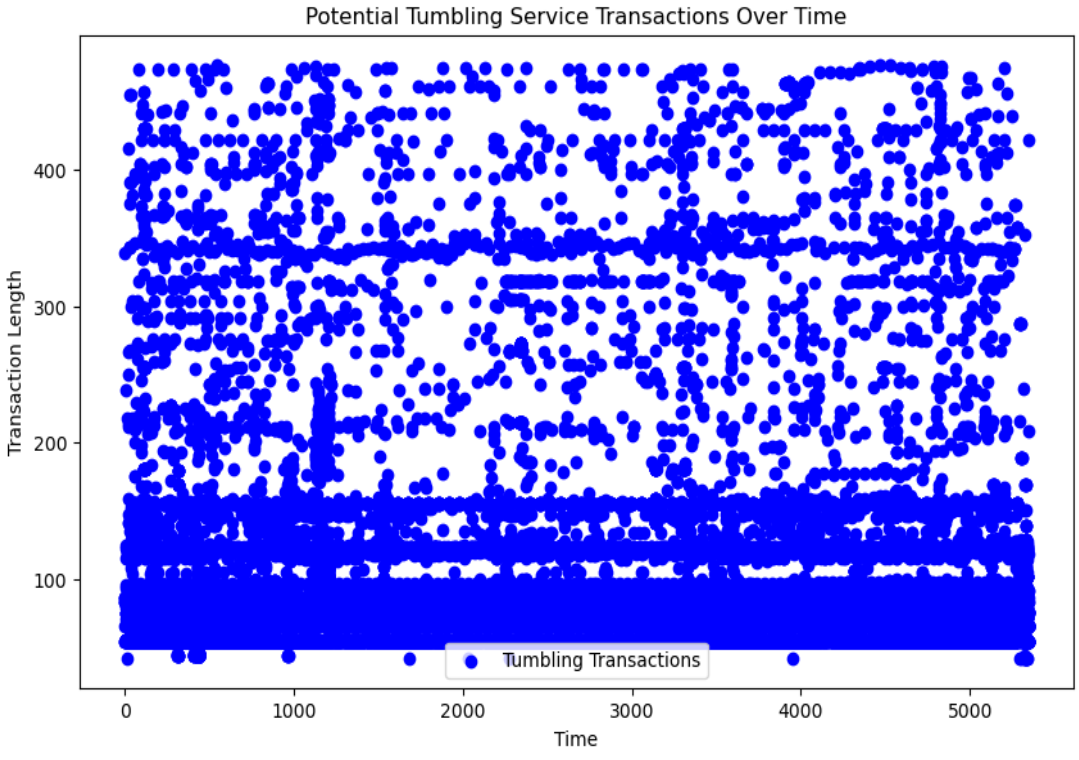

Figure 4. Tumbling Services in NetTran4

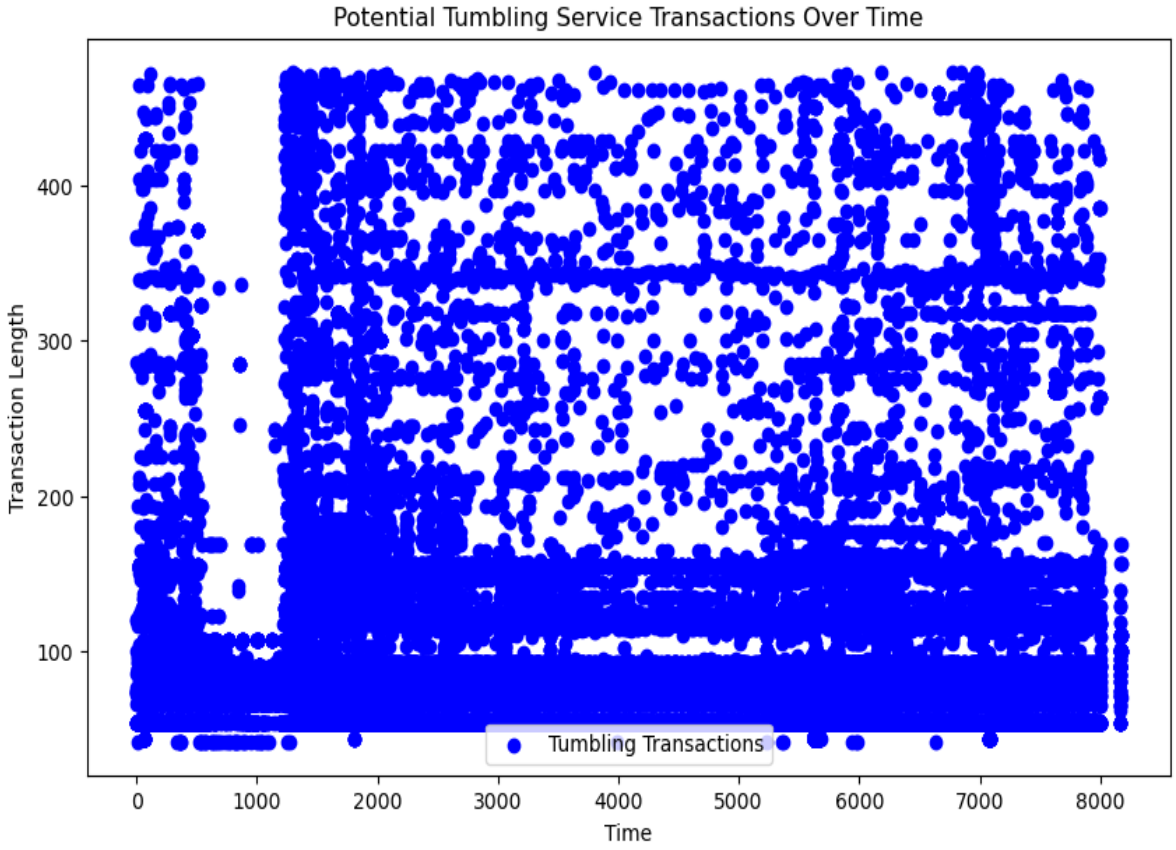

Figure 5. Tumbling Services in NetTran5

Figure 3, 4 and 5 reveals significant use of tumbling services, with 34.42%, 34.01%, and 40.22% of entries, respectively. Tumbling services obscure fund origins by fragmenting transactions and routing them through intermediate addresses. This obfuscation, combined with decentralized networks' lack of oversight, complicates fraud detection by undermining traceability and transparency. The high prevalence of such techniques highlights their widespread use and the challenges they pose to traditional fraud detection methods reliant on tracking transaction paths.

Table 4. Tumbling Services in the Datasets

Dataset	**Tumbling Services**	**Count**	**Percentage**
NetTran3	12730	36985	34.42
NetTran4	38865	114289	34.01
NetTran5	90780	225688	40.22

The significant percentages across datasets underscore the widespread use of tumbling services in financial data, effectively obscuring transaction histories by fragmenting and rerouting funds through numerous addresses. These methods, prevalent in high-volume networks, render forensic tracing nearly impossible. Traditional fraud detection tools, reliant on traceable pathways, face substantial challenges adapting to these anonymization techniques, complicating financial crime prevention efforts.

False Positives (FPs) and False Negatives (FNs) in Detection:

Fraud detection systems face high false positive and negative rates due to the overlap between legitimate and fraudulent activities, compounded by encryption and tumbling services (e.g., 34.01% in NetTran4, 40.22% in NetTran5). Encryption masks transactional details, leading to misclassification of legitimate users and missed detection of fraud. These structural properties inherently increase the likelihood of false classifications in downstream detection systems. These challenges highlight the difficulty in distinguishing obfuscated behaviors in financial networks, as machine learning models struggle with ambiguity in such datasets.

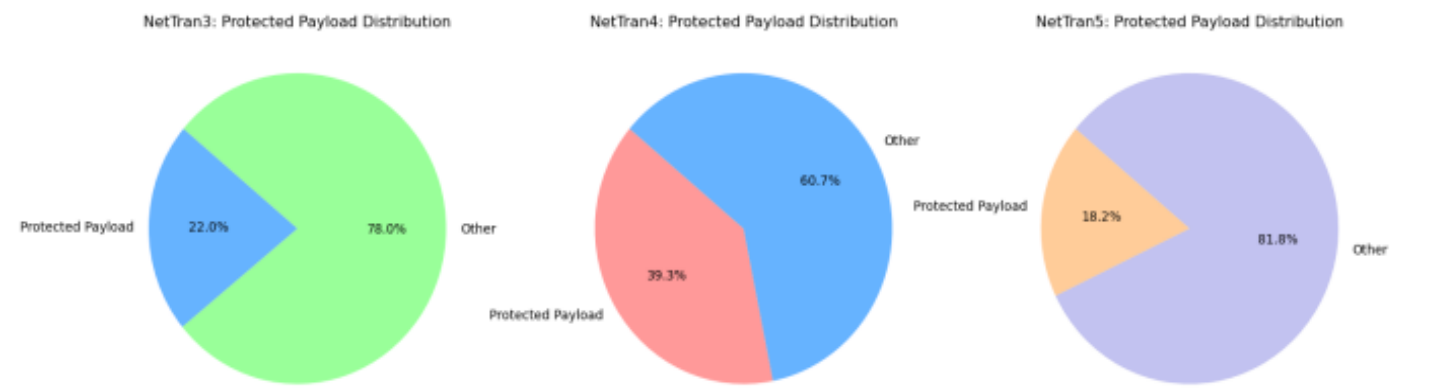

Figure 6: Protected Payload Distribution Chart

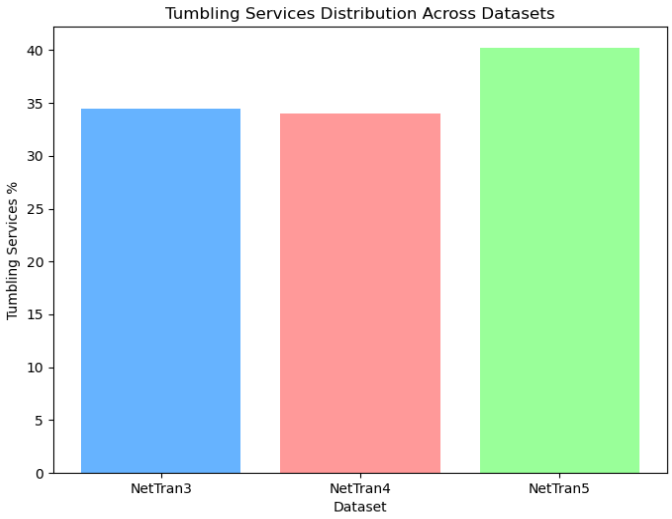

Figure 7. Tumbling Services Distribution Across Datasets

Adversarial Attacks on Detection Models:

Figures 8, 9, and 10 illustrate key patterns that highlight potential adversarial threats. The Packet Size Distribution reveals common sizes and size-based anomalies, while the Time vs. Packet Length Scatter Plot exposes timing irregularities that may indicate adversarial timing manipulations. The Protocol Usage Distribution charts protocol-specific patterns, showcasing potential exploitations of certain protocols. Together, these analyses emphasize the need for robust detection models capable of countering adversarial techniques in encrypted and complex financial networks.

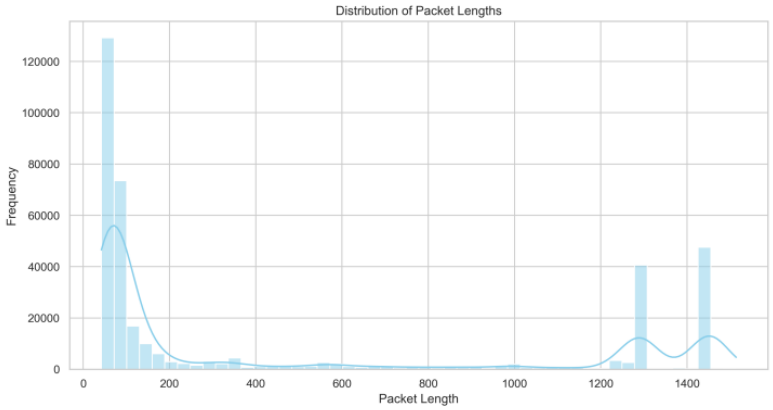

Figure 8. Protocol Usage in NetTran3

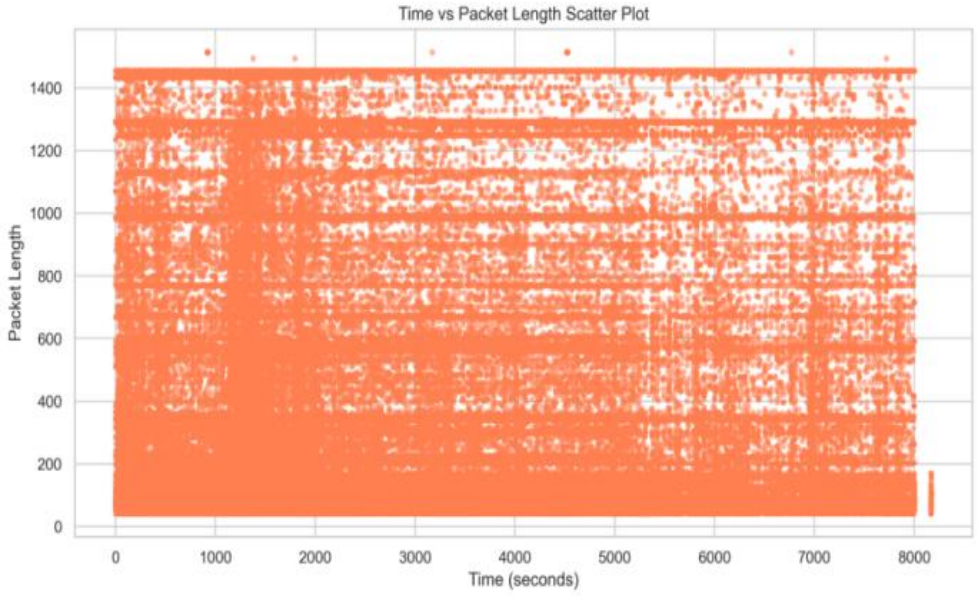

Figure 9. Protocol Usage in NetTran4

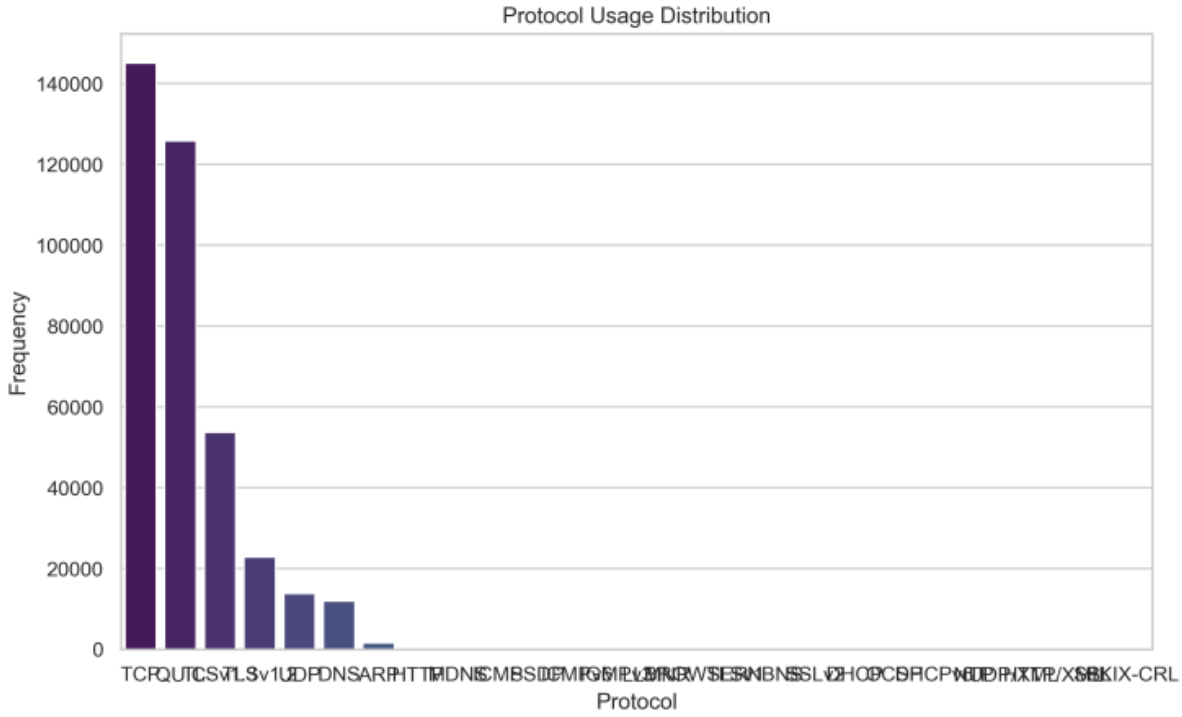

Figure 10. Protocol Usage in NetTran5

Figure 7, 8 and 9 shows that attackers could exploit encrypted protocols like TLSv1.2 and QUIC, commonly seen in network traffic, to conceal malicious activity. Encrypted protocols limit visibility into packet contents, allowing adversaries to embed modifications that evade detection while appearing statistically normal. This obfuscation enables them to manipulate the data transmitted, often avoiding detection by models that depend on recognizing clear anomalies or irregularities.

Moreover, adversaries can use slight variations in packet size and timing to mimic legitimate traffic patterns, bypassing anomaly-based detection systems. By carefully crafting data to fit into expected patterns, attackers can evade detection while maintaining a cover of legitimate-looking traffic. These findings underscore the vulnerability of machine learning models to adversarial attacks, where encrypted traffic and manipulated input data create substantial barriers to effective fraud detection.

Imbalanced Datasets:

Fraud detection in financial transactions is hindered by imbalanced datasets, where rare fraudulent transactions are overshadowed by legitimate ones. This bias leads to models that struggle to detect anomalies, mistaking subtle fraud patterns for normal variations. Encrypted protocols like TLSv1.2 and QUIC further obscure fraudulent activity, compounding the issue by masking packet contents. As a result, fraudulent transactions remain hidden both by their rarity and encryption, highlighting the need for advanced methods to address this dual challenge.

Cross-Chain Transactions:

Cross-chain transactions, where funds are moved between different financial networks, present unique challenges for fraud detection in financial systems. Encrypted protocols like TLSv1.2 and QUIC in these datasets could mask data exchanged across platforms, adding further complexity in distinguishing legitimate from fraudulent cross-chain activities. Monitoring cross-chain transactions requires detection models that can correlate behaviors across networks and protocols. Frequent protocol or IP switching, as seen in the dataset's diversity, allows attackers to mimic cross-chain activities and evade models confined to single-chain analyses.

This demonstrates the need for enhanced cross- network monitoring capabilities in fraud detection systems to address the growing sophistication of cross-chain techniques used by fraudsters in cryptocurrency transactions.

These limitations suggest the need for more sophisticated fraud detection methods that can handle the anonymity, scalability, and evolving tactics of cryptocurrency fraudsters.

CONCLUSIONS AND FUTURE DIRECTIONS

This study presented a dataset-driven validation of structural constraints in network traffic–based financial crime detection. By synthesizing prior literature and empirically examining three progressively scaled real-time datasets (NetTran3, NetTran4, and NetTran5), the research demonstrated how encryption prevalence, tumbling services, scalability pressures, cross-chain complexity, adversarial manipulation, and data imbalance manifest as measurable infrastructural limitations. The findings show that these structural characteristics intensify detection difficulty as dataset size and transactional complexity increase.

Rather than benchmarking detection algorithms, this work provides foundational empirical grounding that clarifies environmental constraints confronting network-based detection systems. Such constraint-oriented validation is essential before designing or deploying anomaly detection models, as algorithmic performance is inherently shaped by underlying dataset properties.

Future research should extend this validation framework to additional network environments, incorporate emerging decentralized financial infrastructures, and examine how structural constraints interact with adaptive detection architectures. By grounding algorithm development in empirically validated infrastructural realities, more robust, scalable, and privacy-aware financial crime detection systems can be systematically advanced.

Recommendations

Enhance Scalability of Detection Systems: Future fraud detection systems must be capable of scaling to meet the demands of increasingly large and complex blockchain networks. This includes optimizing computational efficiency to analyze massive transaction volumes in real-time without sacrificing accuracy or speed.

Develop Cross-Chain Monitoring Solutions: With the rise of cross-chain transactions, it is essential to develop fraud detection models that can track illicit activities across multiple blockchain ecosystems. This requires creating frameworks that can integrate data from various blockchain protocols and link fraudulent activities across different networks.

Improve Data Imbalance Handling:

Machine learning models should be better equipped to handle the imbalance between legitimate and fraudulent transactions. Techniques such as data augmentation, synthetic data generation, and more sophisticated anomaly detection algorithms can help improve detection rates and reduce false positives.

Integrate Privacy-Conscious Techniques:

As privacy concerns grow in decentralized networks, fraud detection systems need to balance the need for transparency with user privacy. Federated learning models and other privacy-preserving techniques can be explored to maintain the integrity of personal data while effectively identifying fraud.

Leverage Advanced Cryptographic Analysis:

Further research into cryptographic techniques, such as de-anonymization and cryptographic key analysis, is necessary to mitigate the challenges posed by obfuscation methods like mixing and tumbling services. Enhancing the ability to trace encrypted and pseudonymous transactions will be vital in combating cryptocurrency fraud.

Adapt to Adversarial Threats:

Fraud detection models must evolve to counter adversarial attacks that exploit weaknesses in machine learning systems. This includes developing robust defense mechanisms that can detect and neutralize manipulation attempts by attackers aiming to bypass detection.

Collaborative Regulatory Frameworks:

Governments, law enforcement, and industry stakeholders should collaborate to create more unified regulatory frameworks for cryptocurrency transactions. Cross-border cooperation is critical for addressing the global nature of cryptocurrency fraud and ensuring compliance across different jurisdictions.

By addressing these recommendations, future research and development in network-based fraud detection can better adapt to the dynamic and rapidly evolving landscape of digital financial crimes. These efforts will help enhance the accuracy, scalability, and effectiveness of fraud detection systems, ultimately improving financial security in the cryptocurrency domain.

REFERENCE

1. B. Liu, Z. Zhang, X. Xu, and M. Luo, “A comprehensive study on anomaly detection in financial systems,” *J. Netw. Comput. Appl.*, vol. 174, p. 102695, 2021. DOI: 10.1016/j.jnca.2020.102695.
2. A. Venčkauskas, R. Maskeliūnas, and R. Damaševičius, “Cryptographic methods in anomaly detection for blockchain systems,” *Future Gener. Comput. Syst.*, vol. 134, p. 203214, 2024. DOI:10.1016/j.future.2023.103214.
3. M. Turner, A. Rigby, and T. Green, “Addressing scalability issues in fraud detection: A graph-based approach,” *Pattern Recognit.*, vol. 115, p. 107968, 2020. DOI: 10.1016/j.patcog.2020.107968.
4. A. Jeyakumar, D. Nathan, and T. Sam, “Machine learning methods for detecting cryptocurrency-related financial crimes,” *Comput. Secur.*, vol. 109, p. 102212, 2021. DOI: 10.1016/j.cose.2020.102212.
5. J. Kang and Q. Buu, “Privacy-preserving techniques in anomaly detection models,” *J. Comput. Sci.*, vol. 147, p. 207345, 2024. DOI: 10.1016/j.jocs.2023.103456.
6. R. Pocher, F. Lemos, and K. Gupta, “Challenges in regulatory frameworks for cryptocurrency transactions,” *Int. J. Finance*, vol. 136, p. 409823, 2023. DOI: 10.1016/j.ijfin.2022.104823.
7. S. Nurmara, C. Lee, and D. Wong, “Advances in anomaly detection using ensemble learning techniques,” *Inf. Syst.*, vol. 153, p. 103456, 2023. DOI: 10.1016/j.is.2022.102345.
8. F. N. Mauliddiah and W. Sun, “Implementing graph neural networks for detecting financial crimes,” *Graph Theory Appl.*, vol. 12, p. 213245, 2023. DOI: 10.1016/j.gta.2023.123456.
9. C. Dumitrescu, R. Martin, and S. Andrei, “Novel approaches to address scalability in transaction

analysis," *Distrib. Ledger Res.*, vol. 94, p. 211345, 2022. DOI: 10.1016/j.dlr.2022.102345.

10. D. Gabrielli, P. Ochoa, and H. Nguyen, "Federated learning applications in anomaly detection," Artif. Intell. Rev., vol. 184, p.20385, 2024. DOI: 10.1016/j.airev.2023.103456.
11. A. Khan, F. Raza, and M. Qureshi, Privacy and scalability in blockchain-based financial systems," *Digit. Ledger J.*, vol. 120, p. 103456, 2023. DOI: 10.1016/j.dlj.2022.102345.
12. P. Ghosh, Z. Anwar, and T. Hussain, "Addressing imbalanced datasets in financial fraud detection," *Mach. Learn. Secur. Appl.*, vol. 45, p. 102345, 2023. DOI: 10.1016/j.mlsa.2022.102345.
13. https://drive.google.com/file/d/1qLVeSInr uY6a3-W5bm3ygTRFSnRJ3MY/view?usp=drive_link.
14. https://drive.google.com/file/d/1GdS3EcG pIUwKvkdOxJcGinIUwsby1LAO/view?u sp=drive_link.
15. https://drive.google.com/file/d/18nsR1g66 hQG3KxCNORtdfXFeiVHdguSC/view?u sp=drive_link.
16. Muhammad Nuraddeen Ado. (2025). Network Transaction Datasets [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/1 1300360