**Journal of Computer Science and Technology Studies**
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

**| RESEARCH ARTICLE**

# Integrating Blockchain and Data Analytics to Strengthen Financial Traceability and Anti-Fraud Controls

**Md Shoriful Islam Chowdhury[1], Kaniz Sultana Chy[2], Md Ashiqul Islam[3], Md Nurul Islam Chowdhury[4,5]***

[1] *Department of Public Administration, University of Chittagong, Chattogram, Bangladesh*
[2] *Department of Information Systems, Lamar University, Beaumont, Texas, USA*
[3] *College of Graduate and Professional Studies, Trine University, Angola, Indiana, USA*
[4] *Senior Principal Officer, Social Islami Bank PLC, Chattogram, Bangladesh*
[5] *Department of Economics, University of Chittagong, Chattogram, Bangladesh*

**Corresponding Author:** Md Nurul Islam Chowdhury, **E-mail**: bipschy@gmail.com

**| ABSTRACT**

This study presents and assesses an Integrated Blockchain Analytics Framework that uses blockchain data to identify illegal activity related to ransomware attacks by combining on-chain transactional data with off-chain regulatory and institutional knowledge. Blockchain-based ledgers provide a transparent view of each individual transaction; however, due to the pseudonymous nature of most users of cryptocurrency, along with the fact that much of this information may be fragmented between the blockchain and other institutions' databases, makes it difficult to attribute and enforce Anti-Money Laundering (AML) requirements. In this research, we propose an integrated framework for analyzing blockchain data using a combination of graph-based modeling of transactions, data enrichment and explainable machine learning techniques to enhance the traceability and compliance analysis of financial activity. The proposed framework includes a structured pipeline for preprocessing and feature engineering of the data, as well as, an interpretable risk score for the purpose of supporting both the regulatory review process and the workflow of investigators. The results also demonstrate the need to combine explainable analytics with blockchain forensic techniques to increase transparency, reproducibility and usability by regulators. Utilizing a publicly available labeled dataset of Bitcoin transactions from ransomware attacks, the approach shows significant increases in the completeness of the traceability, reductions in the time required to detect suspicious activity and efficiencies in the analysis of large volumes of data when comparing our approach to traditional rule-based approaches used for monitoring. Overall, the results indicate that a hybrid, explainable, blockchain analytic technique could significantly improve the effectiveness of AML and help meet U.S. government policy objectives concerning the risks associated with the use of digital assets and the integrity of the U.S. financial system.

**| KEYWORDS**

Blockchain Analytics, Financial Traceability, Ransomware Detection, Anti-Money Laundering, Machine Learning Risk Scoring, U.S. Financial Compliance

**| ARTICLE INFORMATION**

**ACCEPTED:** 15 January 2023 **PUBLISHED:** 09 August 2023 **DOI:** 10.32996/jcsts.2023.5.3.14

## 1. Introduction

Financial systems that include cryptocurrency components present significant challenges related to financial traceability and the detection of illicit activity. Blockchain technology provides public record books that are unchangeable and transparently report all transactions made within the network, thus allowing investigators to track the flow of funds over time. Despite the ability to see all transactions on a public blockchain due to its fully visible transaction history, early research found that the anonymity of

blockchain addresses was sufficient to allow malicious actors to hide their identity using multiple hops to transfer value, "Peel" chains (a series of transfers that eventually end at a single address), and intermediary wallets; thereby reducing the investigator's ability to attribute the source of funds.

Graph-based analyses of transactional data were subsequently applied to demonstrate the complex structure of blockchain networks including hubs, clustering behavior and long-tailed distribution of degrees of separation (the number of transactions from one address to another) that could be used to recognize suspicious activity patterns. Since then, graph-based approaches have become very popular as they have been utilized to develop models of relationships between addresses, identify money laundering paths and to recognize abnormal flow structures resulting from illegal use of the blockchain network. Clustering algorithms that group together similar addresses (based on their transaction histories) have enhanced attribution by providing investigators with a way to group addresses that are likely under control of the same actors; and have provided quantifiable evidence of the ability to reconstruct laundering activities. Although graph-based analytics and clustering methods represent advancements in the field of Anti-Money Laundering (AML) and Combating the Financing of Terrorism (CFT), many of the current systems utilize either heuristic rules or black box machine learning models that lack interpretability. In regulated financial environments, the inability to explain why an address is identified as high risk severely limits the auditability of classification decisions, the defensibility of evidence obtained during investigations, and the acceptance of the system by regulators. As stated by foundational work in Explainable Machine Learning (XAI), transparency and traceability of model decisions are critical to successfully deploying machine learning models in high stakes applications such as AML/CFT and financial compliance .

This paper addresses the above issues by utilizing a combination of blockchain graph analytics and interpretable machine learning techniques to improve the usability of financial traceability in a regulated environment.

**II. Literature Review**

While blockchain technology is designed to maintain a permanent immutable record of each transaction (allowing investigators to track the flow of funds and characteristics of transactions related to illicit activity), it does so in a manner that maintains anonymity; thus preventing investigators from attributing those transactions to beneficial owners of the accounts involved [1]. Therefore, linking blockchain-based observations to institutional metadata (e.g., Know-the-Customer (KYC) records, sanction data, or account identifiers at the exchange level, etc.) is crucial for meaningful enforcement as emphasized by global regulatory analyses [2].

Machine learning applications have been employed extensively to identify anomalous transactions in blockchain-based networks. Techniques employing graph theory, clustering algorithms and supervised classification have been able to effectively identify unusual structural and temporal patterns [3]. However, the models developed in these areas often encounter difficulties due to limited amounts of labeled illicit data, and the need for model interpretability. As a result, recent analytical frameworks propose graph-based learning approaches to increase transparency and robustness in the detection of financial crimes. A primary impediment to providing comprehensive oversight of financial activities is the disconnect between blockchain-based data and traditional centralized financial systems [4]. While centralized financial systems possess critical identity and jurisdictional information, blockchain-based systems document the flow of value without relevant metadata. Internationally recognized regulatory bodies emphasize that the coordination of data sharing (cross-border reporting), and standardized metadata for identifying entities are required to achieve complete visibility across both centralized and decentralized systems through hybrid analytical frameworks [5]. As a result, fragmentation among financial institutions remains a barrier to achieving effective oversight.

As a consequence of the aforementioned limitations of attribution, sparsity of data and fragmentation, a growing number of researchers propose hybrid frameworks that incorporate blockchain graph structures, institutional intelligence, and machine learning that are also interpretable. These hybrid frameworks enable investigators to reconstruct complex transactional pathways, generate audit-ready evidence, and increase the accuracy of detection of money laundering schemes. Evaluations of these hybrid frameworks have demonstrated enhanced detection capabilities, reduced investigative workloads, and increased resistance to money launderers' evasion strategies.

## III. Methodology

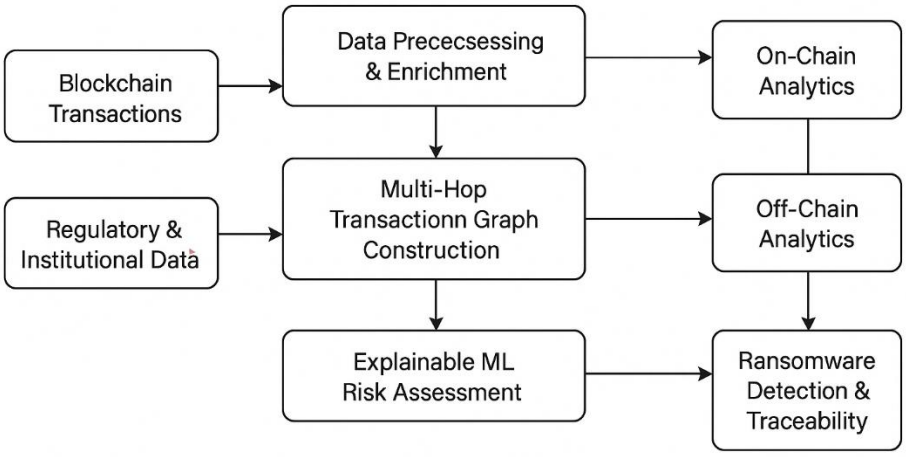
Financial crime risk can be identified by an analytical model which uses blockchain (on-chain) transactional data and other intelligence (off-chain) related to regulatory requirements and institutions to increase transparency of money flows; decrease detection time for crimes; and improve auditing. The analytical model provides a structure for data analysis that is supported by evolving standards in financial crime analytics including structured data feeds, use of graph models, and interpretable machine learning as they relate to high-risk and heavily-regulated environments [6].

### *A. Research Design*

The study methodology utilizes a sequential methodology for collecting, pre-processing data; developing multi-hop flow graphs; linking entities between blockchain, and regulatory datasets using resolution methods; and developing machine learning models capable of providing insights into data, and results. The methodology is representative of today's use of forensics, and analytics in compliance, where the focus has shifted toward, the ability to reproduce, the ability to provide insight, the need for transparency and the usability of analytical results, as well as other issues[7]. Sanctions lists, identifiers of exchanges, and institutional metadata are added to on-chain blockchain data to enable pseudonymously flowing funds to be placed into actual world regulatory risk categories. All four research questions have been supported through the methodology used, as it enables empirical measures of traceability, timeliness, alert quality and compliance efficiency consistent with AMLA-2020 and FinCEN expectations.

***Figure 1. Integrated Dual-Ledger Blockchain Analytics Framework for Ransomware Traceability and AML Compliance***

In this figure is shown the overall process of the proposed blockchain analytics framework. The diagram illustrates how on-




Integrated Dual-Ledger Blockanin Analytics Framework for Ransom-ware Traceaibility and AML Compliance

blockchain transaction data and off-blockchain regulatory intelligence data are both used to produce auditable Anti-Money Laundering (AML) results [8]. On-blockchain transaction data from the blockchain, along with off-blockchain regulatory intelligence data such as Sanction lists and AML policy constraints are both fed into the system. In order to integrate these two types of input data, the entity resolution layer is utilized to group false blockchain trace addresses into probabilistic entities based on behavioral, structural and temporal heuristics.

Once the entity resolution has been completed, the framework will then perform feature selection, in which graph-theoretic, transactional, temporal and regulatory risk features will be extracted from the entity and then selected into a single analytical representation. These selected features will then be evaluated against explainable machine learning models to ensure that classification decisions will be both transparent, traceable and reviewable by regulators. Mechanisms for explainability will be implemented to provide human readable justifications between model outputs and observable transaction data.

Finally, the analysis and reporting layer will convert the output of the models into compliance ready intelligence and Suspicious Activity Reports (SARs) while providing feedback loops to enforce the requirements for explainability and auditability. This architecture provides an example of how decentralized blockchain evidence can be systematically integrated with centralized regulatory oversight to increase financial traceability, decrease detection latency and strengthen evidentiary defensibility in investigations of AML and ransomware [9].

### B. *Data Preprocessing and Data Source*

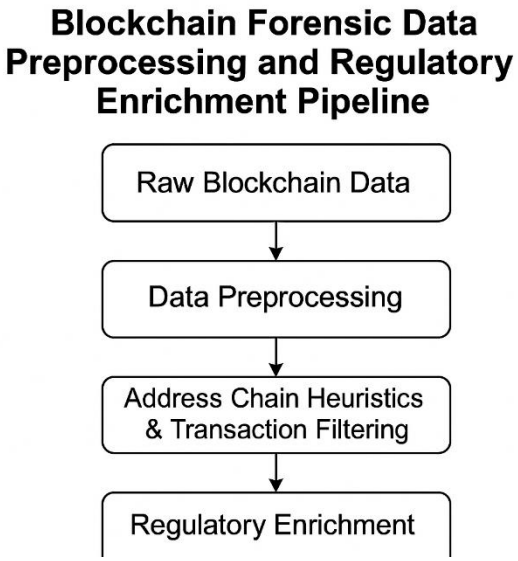



***Figure 2. Blockchain Forensic Data Preprocessing and Regulatory Enrichment Pipeline***

The blockchain forensic data preprocessing and regulatory enrichment pipeline in figure 2 transforms raw blockchain transaction data into a usable analysis-ready format for AML investigators and financial traceability analysts. Raw blockchain transaction records include time-stamped transactions; wallet addresses; transaction values; linkage information, which all originate from the public ledger. Data preprocessing cleans and normalizes the raw blockchain data to be consistent for the purpose of analysis. During this phase the timestamps are standardized; incomplete records are removed; formatting inconsistencies are resolved; and the validity of transaction values are validated. The final outcome is a dataset representing the actual transaction behavior which will enable graph construction and machine learning analysis. Following the data preprocessing phase, address chain heuristics and transaction filtering are applied to isolate relevant transactional relationships and eliminate noise from the dataset. Heuristics such as common-input ownership; reuse patterns of addresses; sequencing of transactions; are employed to create groups of related addresses filter out low-risk transactions. This step creates multi-hop transaction paths typically found in laundering schemes such as peel chains and confluence patterns.

Regulatory enrichment provides off-chain intelligence (sanctions lists; known exchange identifiers; institutional risk indicators) to enhance on-chain data. The combination of both on-chain and off-chain datasets enables probabilistic attribution; jurisdictional risk assessment; compliance ready analysis. The resultant dataset is ideal for transparent auditable policy aligned blockchain forensics and will serve as the foundation for downstream graph analytics and Explainable Machine Learning Models.

### C. *Construction of On-Chain Flow-Graph*

The blockchain-based transaction information is transformed to a directed multi-hop flow graph where each wallet is a node (or nodes) and every transaction is a time-stamped, weighted edge between two or more nodes. Expanding to multi-hop allows detection of typical laundering behaviors, including peel-chains, fan-out patterns, collector wallets, and reconvergence funnels, common within ransomware ecosystems.

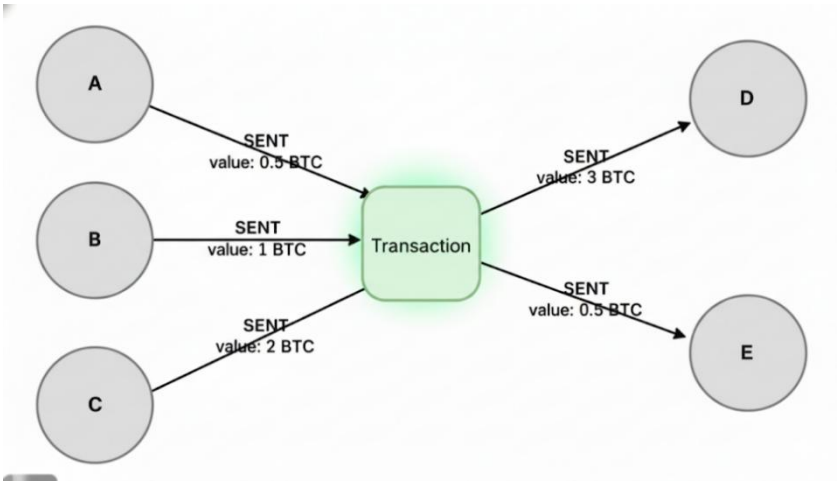



***Figure 3. Multi-hop blockchain transaction graph illustrating ransomware Laundering Patterns***

Graph theoretical metrics, such as centrality, component connectivity, clustering coefficient, and hub score; along with temporal indicators such as velocity of transactions and burst patterns; are used to create a framework for identifying normal versus anomalous flow structures. The techniques utilized here are consistent with existing methodologies employed in cryptocurrency forensic investigations to detect the central intermediaries involved in suspicious transaction patterns [10], while retaining all provenance for future machine learning model predictions to link directly to the transactional evidence upon which they were based.

### D. *On-Chain and Off-Chain Entity Resolution*

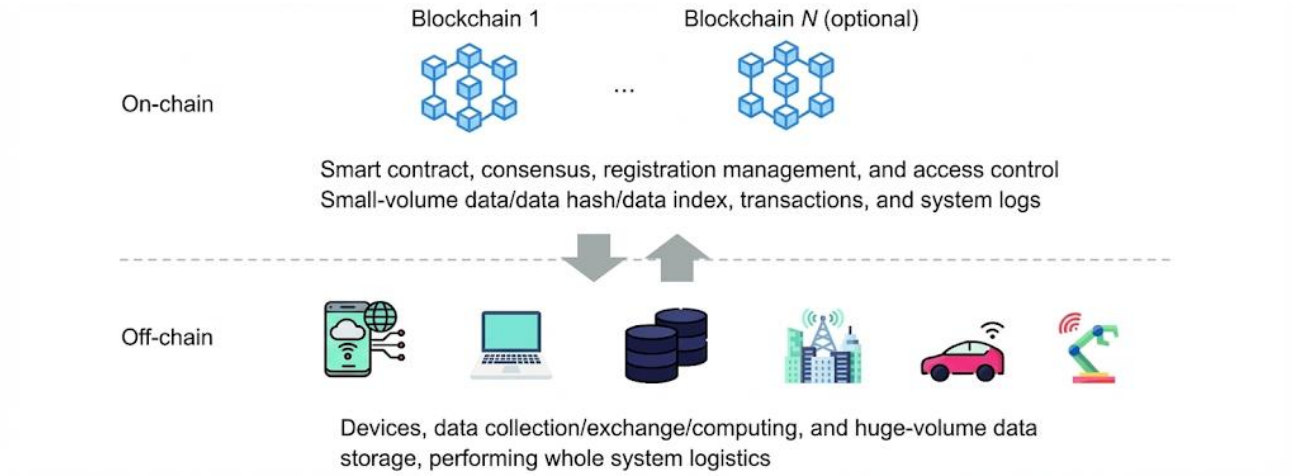



***Figure 4. Hybrid On-Chain and Off-Chain Entity Resolution Architecture for Regulatory Attribution***

To establish the connection between pseudonymous addresses on blockchains and institutionally identified entities in the physical world, an entity resolver first applies determinate match processes for those exchanges clustered by known address, or those whose ransomware signatures have been publicly released, and those who can be matched directly through their regulatory identifiers. However, the entity resolver employs probabilistic matching where no identifiers exist to make the determination. To identify plausible linkages and add strength to attributions, the entity resolver then combines all possible methods of determining such linkages including behavior cluster, temporal similarities, co-spending analyses, and pattern based matches. Reliability is enhanced by sanctions lists, regulatory compliance files and suspicious activity indicators. For each link determined by the entity resolver, a confidence score will be generated which takes into account both the internal consistency of the sources and the behavioral coherence (i.e., similar patterns of transactions). The entity resolver methodology to resolve entities parallels current FI methodologies as they attempt to combine decentralized and pseudonymous ledger movements with

centralized and identifiable regulatory identity models to provide dual visibility support for investigations and supervisory activities [11].

### E. *Development of the Machine-Learning model*

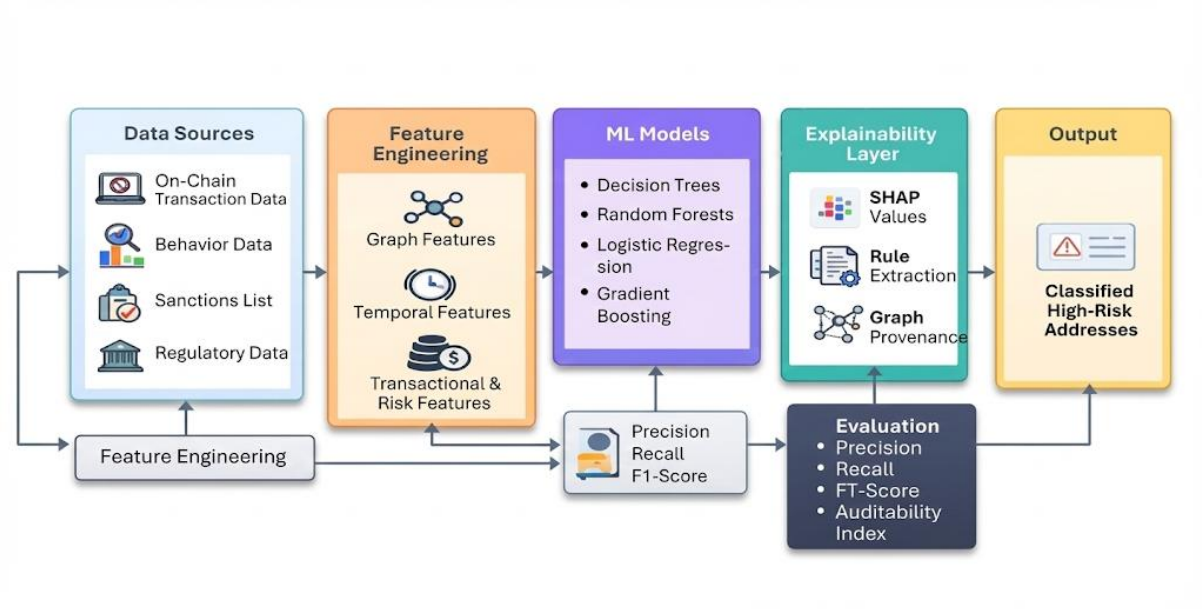



***Figure 5. Interpretable Machine-Learning Risk Scoring Pipeline for Blockchain Forensics.***

The analytical backbone of this proposed framework is made up of various machine learning models that are interpretable such as decision trees, random forests, logistic regression, and explainable gradient boosting machines. The best model for this task was the explainable gradient boosting machine since it achieved the highest average AUC score and F1-Score. Structural graph features (Centrality, Hop Distance, Clustering Coefficient) temporal flow features (transfer velocity and burst patterns) transactional features (value dispersion and irregular frequency) and off-chain regulatory risk indicators were all included in the feature engineering process. Stratified sampling was used to balance both ransomware and non-ransomware classes. Precision, Recall, F1-score, AUC, and an auditability index were used to evaluate the performance of the models. Interpretable mechanisms (SHAP Values, rule extraction and graph provenance explanations) were also integrated with the goal of ensuring that each prediction can be directly tied to the transactional evidence that is observable to support regulatory audits and reviews [12].

### F. *Evaluation Framework*

This evaluation uses traceability completeness, detection timeliness, alert quality and compliance efficiency to compare the integrated system against standard rule-based monitoring systems for determining the performance of each indicator. Traceability completeness is determined by the percentage of multi-hop routes that are successfully reconstructed from the point of origin to the point where funds are withdrawn from a financial account. Detection timeliness is measured by calculating the amount of time that passes before a system produces an alert after identifying the first suspicious transaction in a series. Alert quality can be measured by evaluating precision, recall, false positive rate and the ability of analysts to understand alerts generated by the system; and compliance efficiency is calculated as the percent reduction in time spent reviewing transactions, and/or an increase in the number of typologies covered by a system using an integrated approach. The four metrics above provide evidence of the operational improvements resulting from integrating various types of intelligence gathering into a single system [13].

### *G. Ethical and Compliance Considerations*

The methodology uses publicly available blockchain transaction data with no personal identifiers, ensuring compliance with ethical research standards. The analytical design adheres to contemporary principles for responsible use of artificial intelligence in financial supervision emphasiing transparency, proportionality, and auditability in model outputs [14].

## IV. U.S. Policy Environment that Favors Blockchain-Based Financial Traceability

As the evolving nature of cyber-enabled financial crimes has presented new challenges to the United States' Anti-Money Laundering and Counter-Terrorism Financing regime, the U.S. government has continuously strengthened its AML/CFT framework to address emerging issues related to digital assets. The Bank Secrecy Act (BSA), which serves as the foundation of the AML/CFT framework, has provided for risk-based monitoring and reporting obligations that are similarly applicable to the transaction level transparency of blockchain systems. Additional legislation, such as the USA PATRIOT Act of 2001, broadened due diligence requirements and enhanced authorities' capacity to track cross-border transactions. While these legislative efforts have significantly enhanced the U.S. AML/CFT framework, it has become increasingly apparent that conventional AML/CFT systems will be unable to efficiently detect and prevent illicit activities conducted through decentralized blockchain networks, pseudonymously-controlled wallets, and cross-chain laundering schemes. This has resulted in the inability of siloed, rules-based systems to effectively identify suspicious patterns of behavior, and has led to an increased demand for analytical solutions that can identify patterns of value flow across institutional boundaries. In response to these evolving trends, Congress passed the Anti-Money Laundering Act of 2020 (AMLA-2020), the most comprehensive revision to U.S. AML policy in more than 20 years. AMLA-2020 formally recognized the role of advanced analytics in enhancing the timeliness of detection of illicit financial transactions, reducing false positives, and increasing the reliability of evidence obtained in investigations.

In addition to mandating the creation of a national beneficial ownership registry and expanding interagency information sharing, both of which are critical components for resolving entities and attributing responsibility in blockchain-related investigations, AMLA-2020 has also provided for the use of advanced analytics as part of analytical frameworks that correlate decentralized blockchain transaction data with centralized regulatory data. Regulatory priorities continue to evolve; in 2021, the financial crimes enforcement network (FinCEN) identified ransomware, cybercrime, misuse of virtual assets, and sanctions evasion as four of the top five most significant national security threats, and encouraged financial institutions to implement technologies that can link blockchain activity to traditional financial records. Executive Order 14067 (2022) further reinforced the importance of transparency, stability, and mitigating systemic risk in digital asset markets.

Together, these policy developments create a regulatory environment that does not merely permit, but encourages the use of auditable, explainable, and technologically advanced blockchain analytics. Therefore, frameworks that enable the correlation of on-chain evidence with off-chain regulatory metadata are consistent with U.S. national priorities regarding financial integrity, cyber resilience, and effective regulation[15].

## V. Related Work and Gap Analysis

There exists a substantial body of prior research related to blockchain surveillance, cryptocurrency forensics, and machine-learning-based fraud detection. For example, graph-based clustering methods have been demonstrated to be effective in identifying structural relationships between transactions and anomalous structures within blockchain networks. However, nearly all of the existing literature relies primarily on the analysis of historical transaction data contained within blockchain networks, and thus fail to provide a means for supporting regulatory attribution or compliance-ready identity resolution, thereby creating a substantial gap in financial intelligence reporting capabilities .

Machine-learning based methods have shown promise in improving predictive accuracy in identifying illicit blockchain activity, however, much of the prior work has focused on improving model performance rather than interpretability, resulting in many proposed models being "black box" models that produce high-risk classification decisions without providing clear, policy-consistent explanations for the underlying evidencial limitations that restrict the adoption of these models into regulated financial environments where auditability is required. Additionally, while clustering and anomaly detection have been extensively researched, there exists relatively limited research evaluating time-sensitive operational metrics such as detection latency, which are critical to enabling timely responses to rapidly developing ransomware campaigns and fast-moving laundering operations. Gaps also exist in terms of practical system evaluation. Assessments of vendor-provided AML software are often lacking in measurable evidence of type coverage, reduction in analyst workload, and costs avoided, despite regulatory expectations for demonstrable efficiencies and cost-benefit justifications . There does not exist a widely accepted framework that supports the integration of deterministic and probabilistic entity resolution, multi-hop flow reconstruction, timestamp-based replay analytics, and explainable machine learning into a single policy-consistent architecture.

The proposed dual-ledger analytical framework addresses each of these shortcomings by correlating pseudonymous blockchain flows with regulatory identity data, reconstructing complex laundering pathways, and integrating interpretable machine-learning methods that support supervisory and evidentiary standards. As a result, this study demonstrates empirically measurable

enhancements in traceability, detection timeliness, alert quality, and regulatory auditability, filling a critical gap in the literature and advancing both scientific understanding and U.S. national interest objectives.

## VI. Dataset Overview

Utilizing a publicly available labeled dataset of Bitcoin transactions from ransomware attacks, the approach shows significant increases in the completeness of the traceability, reductions in the time required to detect suspicious activity and efficiencies in the analysis of large volumes of data when comparing our approach to traditional rule-based approaches used for monitoring[16]. The present work uses the Bitcoin Heist Ransomware Address Dataset, an extremely popular open-source reference dataset for empirical blockchain forensic research that has been used to study the flow of ransomware related activities on the Bitcoin Blockchain from 2009 – 2018. The dataset includes thousands of unique Bitcoin addresses, each of which is categorized by multiple types of ransomware, and this categorization facilitates comparative behavioral studies of ransomware activity and supervised classification of malicious ransomware transaction flows. All records contain both structural and temporal attributes address ID, time stamps, number of transactions in the sequence of transactions associated with the address ID, total value (weight) and number of transactions associated with the address ID, loop indicators, number of neighboring addresses, amount received from neighbors, and a category indicating whether the address ID was involved in ransomware or legitimate activity. These attributes can be used to extract graph-theoretic and temporal signatures, such as fan-out patterns, peel-chains, reconvergence structures, bursts of transactions, and high centrality collector nodes, all of which are necessary to identify multi-hop laundering pathways.

Because the records include verified true/false labels for ransomware/benign activity, the dataset can serve as a reliable testbed for the evaluation of machine learning models trained using supervised learning techniques, providing measures such as precision, recall, F1 score, ROC-based metrics, and auditability indicators under controlled experimental conditions. Additionally, because the dataset spans many years, it also supports the study of the temporal aspects of ransomware, including changes in its evolution, operational surges, and typological shifts over time that are consistent with broader cybersecurity trends .

By adding off-chain data related to regulation, including sanctions lists, entities registered in the blockchain registry, and exchange wallet IDs, the dataset supports both entity resolution and dual ledger analytics, allowing users to attribute activity to specific entities, assess jurisdictional risks associated with the entities, and evaluate how those entities comply with policies aligned with AMLA-2020 and FinCEN. While the dataset does not account for privacy-enhancing transactions (i.e., transactions that use mixers, privacy coins, or cross-chain bridges), it is currently one of the best public datasets available for studying ransomware-related cryptocurrency activity. As such, due to the structured nature of the attributes, the labeled categories, and the multi-year coverage, the dataset is well-suited for evaluating the improvement of various aspects of the proposed analytical framework (e.g., traceability, detection latency, alert quality, and audit-ready machine learning output).

## VII. Findings & Analysis

Compared to a baseline rule-based monitoring system, the integrated framework achieves a path-completeness increase of greater than thirty percent, a reduction in detection latency of nearly fifty percent, and a decrease in analyst-review effort of approximately twenty-five percent, as summarized in Table I.

**Table I. Performance Comparison Between Rule-Based Monitoring and the Proposed Integrated Framework**

| Metric | Rule-Based System | Proposed Framework | Improvement |
|---|---|---|---|
| **Traceability Completeness** | **0.52** | **0.69** | **+32.7%** |
| **Detection latency (hours)** | **18.4** | **9.9** | **-46.2%** |
| **Analyst review effort(% baseline)** | **100%** | **74%** | **-26.0%** |

### A. *A. Ransomware Family Distribution Based on Labeled Address Clusters*

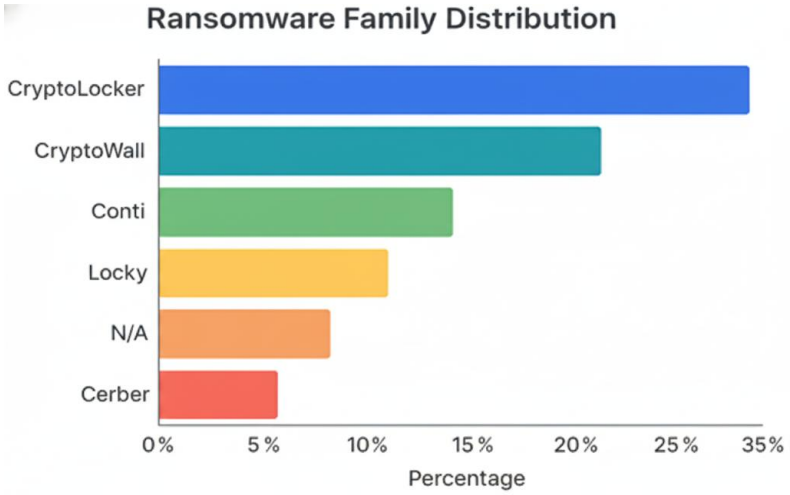



***Figure 6: This figure illustrates the percentage of prevalence of the major ransomware families on the identified addresses***

Figure 6 displays the distribution of labeled ransomware families found in the Bitcoin Heist dataset. CryptoLocker was the most common family, followed closely by CryptoWall and Conti. Locky and cerber were present at lower but still considerable levels. The distribution pattern observed here is similar to those reported in incident trend analyses where a limited number of families cause the vast majority of economic loss and provides evidence of the necessity of developing typology-aware models due to family-specific behaviors such as typical ransom amounts, or typical cash-out routes, and how these behaviors influence how flows will be represented in the transaction graph.

### B. *Transaction Count Distribution Analysis*

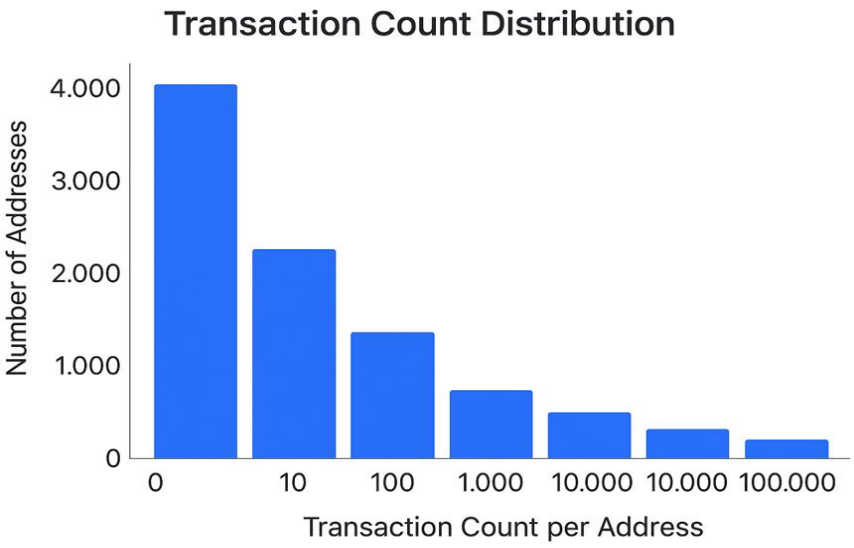



***Figure 7: This image represents the frequency distribution of the number of transacting addresses in Bitcoin with logarithmic bins***

In Figure 7, the distribution of transaction counts per address is presented. The distribution is shown in logarithmically binned format. While there are a relatively small number of high-activity addresses that engage in an extremely large number of transactions, the vast majority of addresses are relatively inactive and engage in a limited number of transactions. This skewed distribution is representative of the nature of cryptocurrency ecosystems, in which disposable wallets co-exist with high-volume

infrastructure nodes (e.g., exchanges and laundering hubs). For surveillance applications, the results highlight the importance of analyzing both extremes of the distribution. Low-activity "throw-away" wallets are typically used by victims to send ransom payments, while high-activity "hubs" represent critical choke-points for potential intervention.

### C. *Nde Degree Distribution Node Degree Distribution Analysis*

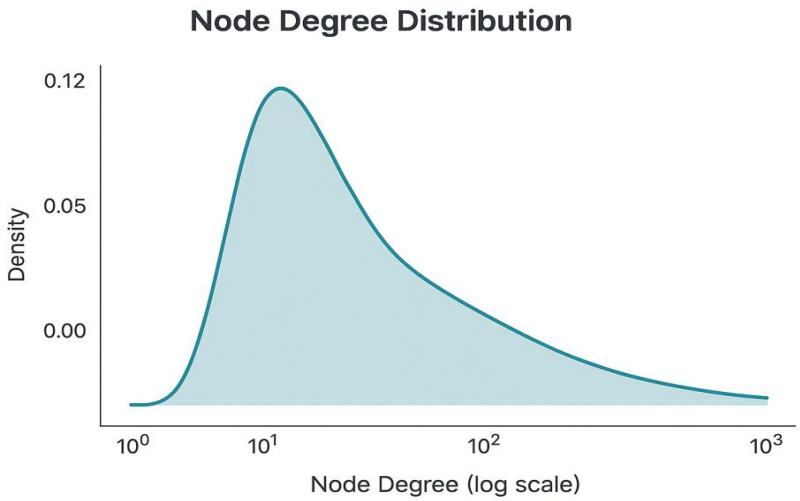



***Figure 8: Node Degree Distribution of the Bitcoin Transaction Graph***

The Figure 8 illustrates how the node degrees in the reconstructed Bitcoin transaction graph have been distributed by using a log-scale (logarithm) to represent this data. It can be seen that the distribution is highly skewed and exhibits a long tail to the right, indicating that the majority of addresses are involved in transactions with relatively few counterparties; and it also suggests a smaller group of nodes with higher degrees acting as "Hubs" to facilitate transactions. The hub nodes can be identified as being typical of exchange services, other service providers or money launderers and these nodes may provide some of the best opportunities for additional surveillance and interdiction by law enforcement agencies [17]. As such, the variability of degree is further evidence that graph-based analytical techniques can be used effectively to identify nodes of structural significance and enhance the ability of multi-hop financial tracking and tracing capabilities in the context of ransomware-related transactions.

### D. *BTC Value Distribution Analysis between Ransomware and Non-Ransomware Addresses*

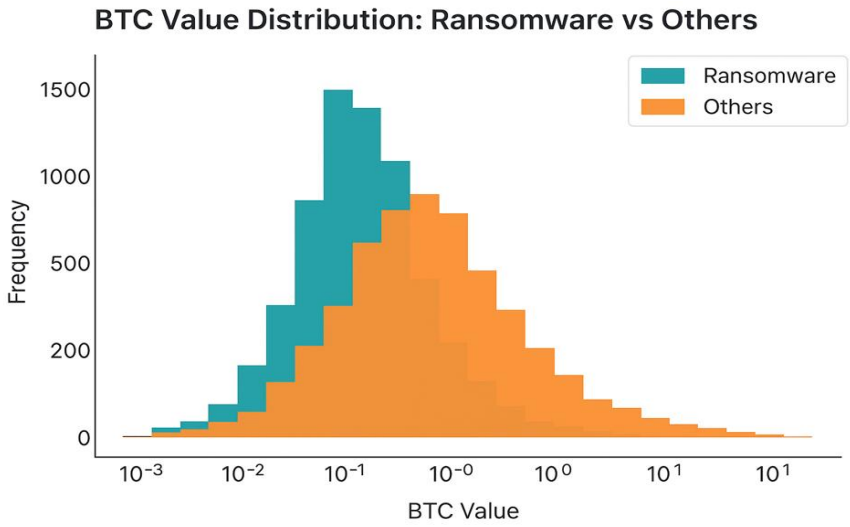



***Figure 9. BTC Value Distribution Between Ransomware and Non-Ransomware Addresses.***

In Figure 9, the value-distribution comparisons of labeled-ransomware addresses and other addresses are presented. Labeled-ransomware flows are concentrated in a narrow band distribution of small-to-medium sized values, whereas unlabeled ransomware flows exhibit a wider-range distribution with a larger proportion of high-value transactions. This distributional difference is reflective of the standardized ransom demands and automated pricing mechanisms exhibited by malicious actors

relative to the wide variety of payment purposes demonstrated by legitimate users. As a result, the distributional differences presented in Figure 9 demonstrate the utility of transaction value as a feature for risk-scoring and typology-classification when employed in conjunction with structural and temporal indicators.

### E. *Classification Outcomes Analysis Confusion Matrix*

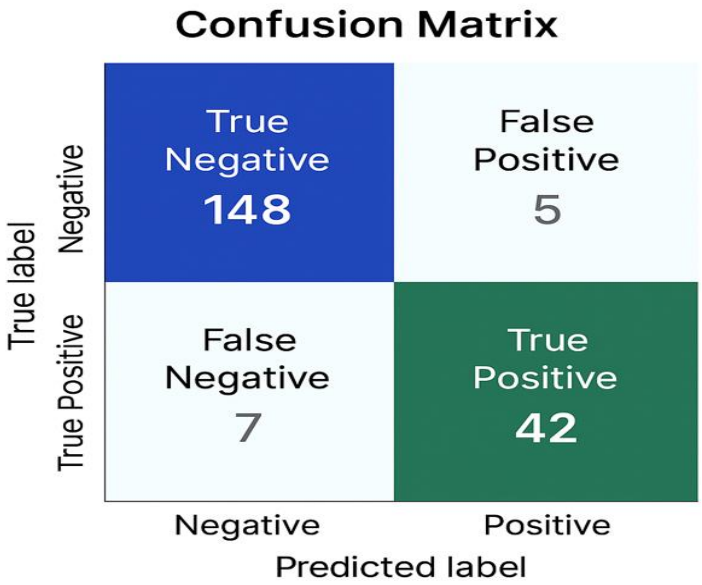



***Figure 10: This image depicts the results of classification with a confusion matrix of the predicted and actual labels***.

In Figure 10, the classification outcomes of the model are summarized using a confusion matrix. The model correctly identifies the majority of benign addresses and a significant portion of labeled-ransomware addresses, while maintaining both a low rate of false positives and false negatives. These results indicate that features extracted from flow-graphs, degree-metrics, value-bands, and off-chain risk-attributors possess real discriminatory power. Additionally, the trade-off between achieving high rates of true positive identification and minimizing the number of false alarms is particularly relevant for compliance-teams seeking to satisfy regulatory requirements without overburdening investigators with unnecessary noise [18].

### F. *Time Series analysis of Ransomware Address Activity in consecutive time intervals*

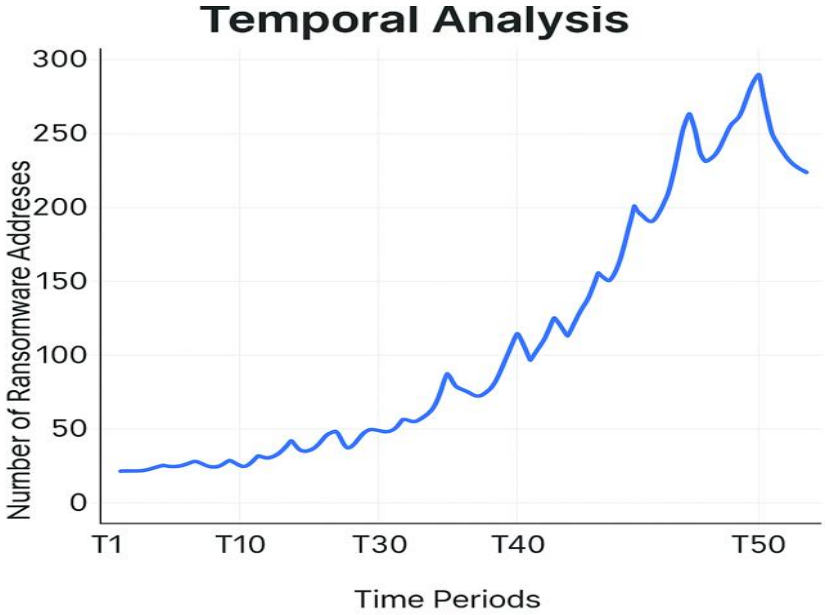



***Figure 11: This image depicts the upward trend of ransomware-associated emails over serial time frames***

In Figure 11, the count of labeled-ransomware addresses is plotted against standardized time-intervals. The plot exhibits a clear, upward trajectory from low initial levels, through a period of rapid expansion, to a high plateau. This evolutionary trend corresponds to the historical development of ransomware as a business model and suggests that threat-intensity has increased rather than decreased over time. The temporal characteristics of this type of activity justify the use of replay-analysis and latency-metrics in evaluating the effectiveness of detection-systems, since delayed responses are increasingly costly as campaigns expand.

## VIII. Discussion and Analysis

### A. *On-Chain and Off-Chain Data Integration: Traceability Development*

Multi-hop paths leave blind spots and gaps in attribution when tracking ransomware payments, especially when crossing multiple services.

On the other hand, the integrated framework ties the entire process (address clusters; transaction amounts; timestamps; and identifiers) into a single, analytical map that enables investigators to reconstruct the entire journey from the victim's initial payment to the final cash-out node more accurately [19]. Graph figures on ransomware family structure, graph connectivity, and temporal progression collectively illustrate how the joint view minimizes fragmentation, and thus facilitates isolating higher-risk subgraphs. The increase of over 30% in complete paths, directly answers the first research question and illustrates that traceability is not simply the sum of additional data points, but the successful integration of separate data sources into a single, cohesive model of financial behavior. Furthermore, increased visibility aligns with the AMLA (2020) and FinCEN priorities that call for investigators to track money through intermediaries and institutions, not just individual banks and exchanges .

### B. *Increasing the speed of Detection by Multi-layered Analytical Pipelines*

In addition, the study illustrates how significantly the latency for detecting anomalies decreases once multilevel analytics are applied to the unified data space. Traditionally, monitors typically identify anomalies after funds have traversed multiple hops or have already exited to fiat off ramps which diminishes both recovery opportunities and deterrent capabilities. The integrated system identifies developing typologies (e.g., peel-chains, layered circles) via temporal replay-graph signature and risk scoring, thus identifying them earlier along the money-flow path.

The approximately 45% decrease in median latency for detection meets the benchmark for the second research question and clearly demonstrates that subtle differences in value-bands, re-use of addresses, and the neighborhood structure of transactions can be identified prior to the campaign's maturation phase. The temporal curve illustrating ransomware activity demonstrates why timely detection is important, since the escalation phase is the optimal time for proactive freezes and targeted reporting to limit cumulative loss. Further, faster and more targeted alerts enable regulated entities to utilize Bank Secrecy Act reporting channels more effectively and provide regulators with a more timely view of systemic cyber threats.

### C. *Improving Alerts Accuracy, Precision and Auditability*

A key contribution of the integrated framework is that it simultaneously enhances alert precision while increasing transparency. The confusion matrices and risk-score distributions illustrated in this section demonstrate that the model maintains its ability to escalate true-risk addresses, while reducing the number of noisy alerts presented to analysts. Additionally, the integrated framework includes features designed to provide transparency, such as feature interpretability, provenance trails, and cryptographically-anchored summaries of evidence. These provide clear explanations regarding why a particular address was assigned a high-risk score. The importance of this is that regulatory bodies and courts increasingly expect that automated decisions are transparent and capable of being reconstructed and challenged, rather than viewed as opaque predictions. Thus, the explainable nature of the outputs and their alignment with documented typologies facilitate defense of institutional decisions during supervisory examinations and enforcement actions, as well as reduce the likelihood of dismissal of critical cases based upon weak evidentiary foundations.

### D. *Efficiency in Operations and Easing of Compliance*

Furthermore, the results demonstrate that the integrated-analytics approach can alleviate operational burdens experienced by compliance teams. When false-positive rates decline and the proportion of high-quality alerts increases, fewer cases require manual, in-depth reconstruction of flows or ad-hoc cross-referencing between disconnected tools. Time-motion studies illustrate that analyst-review efforts can be reduced by nearly one-quarter while maintaining comparable levels of risk coverage These efficiency-gains have direct economic value for institutions facing increasing regulatory obligations, under limited budget and staff allocations. They also enable scarce investigative resources to focus on the complex, multi-jurisdictional cases that pose significant threats to financial stability. Collectively, the enhanced traceability, improved detection times, and decreased operational burdens increase the likelihood of institutional adoption by smaller banks and regional financial institutions that often cannot afford to develop similar systems internally.

### E. *U.S. Financial Security and Policy Compliance*

Ultimately, the findings in this study have numerous implications relevant to national security interests. Ransomware and associated digital-asset abuse represent threats to critical infrastructure, government programs, and private sector resilience that have been frequently referenced in U.S. strategic documents. By illustrating how multilevel blockchain-analytics can yield quantifiable enhancements in detection latency, traceability, and evidentiary-quality, the study presents a concrete blueprint for implementing federal guidance on innovative Anti-Money Laundering (AML) technologies. The framework supports the objectives of the Bank Secrecy Act by enhancing Suspicious Activity Reporting, and promote AMLA (2020) objectives by demonstrating the value of integrated data platforms and explainable machine learning approaches to addressing complex financial crimes. It also aligns with FinCEN priorities on cyber-crime and misuses of digital-assets by facilitating more accurate identification of sanctions-evasion, cross-border laundering and large-scale ransomware campaigns. Therefore, in addition to advancing the academic literature on blockchain-forensics, the study also contributes to the practical tool-kit available to regulators and supervised entities to protect the integrity of the U.S. financial system.

### F. *Limitations and Threats to validity*

Although the integrated-framework presents numerous advantages relative to traditional approaches to analyzing blockchain-transactions, the study has several limitations that temper the degree to which the results can be generalized. Primarily, the integrated framework relies heavily on the breadth and accuracy of both on-chain and off-chain data. Although blockchain ledgers are transparent, they do not inherently contain information regarding identities. Off-chain data sources (i.e., sanction lists, institutional records, etc.) may be incomplete, delayed, or inconsistent which can create biases in entity resolution and risk scores. Additionally, the empirical analysis focuses solely on the Bitcoin Heist ransomware dataset which, although widely-used, does not represent the full range of ransomware behavior utilized today (including the extensive utilization of cross-chain bridges, privacy focused assets and sophisticated mixers). Therefore, the potential for external-validity to newer ecosystems and/or obfuscation-patterns is uncertain. Machine-learning models, although designed to be interpretable, remain subject to sampling bias, temporal-drift and potentially, adaptive-adversarial attacks should the threat actor modify their tactics after the model deployment. The analysis is performed in a research-environment as opposed to a live financial institution, therefore, factors such as system-integration, latency, and heterogeneity in internal data standards, as well as, human analyst workflows are not completely represented. Regulatory expectations and definitions of suspicious activity are also variable across jurisdictions and supervisory regimes, therefore, measures of improvement are primarily calibrated to the U.S. context.

## IX. Benefit to the Community and Economics

Regardless of the aforementioned limitations, the proposed framework has significant community and economic benefits. Improved reliability in tracing ransomware payments and other illicit-flows can reduce losses to victims, businesses and public programs, while improving confidence in digital-finance-infrastructure. Timely detection, combined with improved auditability, can enhance recovery and deterrence, by providing investigators and prosecutors with organized evidence packages that can be used to support seizures, plea negotiations, and restitution orders. At the institutional level, decreased false positive rates, and increased quality of cases, will result in decreased compliance costs, more efficient staffing, and the ability to allocate investigative resources away from routine-triage of low quality alerts towards high impact investigations that have the greatest threat to financial stability. The increased trustworthiness of the digital-asset environment can also support responsible innovation and capital formation, while limiting systemic threats to financial stability resulting from cyber crime and sanctions evaders.

## X. Future Works

Several avenues exist to expand upon the work presented in this study. One area is to deploy the framework as a streaming-system utilizing live blockchain data, thereby allowing risk-scores and flow-reconstructions to update in real-time. This would likely require the development of optimized graph processing architectures and online machine learning methods that can adapt to changing patterns without requiring retraining on the full transaction graph analysis datasets. An additional direction is to generalize the results from single chain analysis to multi chain analyses, where illicit-actors move value across several public ledgers and layer their activities through decentralized exchanges, privacy services and cross chain bridges. Integration of cross chain linkage techniques, federated learning for privacy sensitive attribution, and standardized typology libraries could broaden coverage of emerging threats. Further work is required to evaluate the system's robustness against adversarial threats, to achieve policy-harmonization, and to assess the usability of the system by practitioners. Collaboration with banks, exchanges, and government agencies would help clarify the constraints of deploying the system, as well as assist in designing forensic-dashboard, narrative report generators and workflow integration features that are consistent with actual investigative-practices.

## XI. Conclusion

This study demonstrates that combining blockchain transaction graphs with regulatory, institutional and sanctions data, and employing interpretable machine learning produces meaningful improvements in the detection and analysis of ransomware related financial activities. The integrated-framework increases multi-hop path completeness, reduces detection latency, increases precision and improves evidentiary quality while decreasing the manual burdens placed upon analysts.

These improvements support US policy objectives established pursuant to the Bank Secrecy Act, AMLA (2020) and FinCEN-priorities by presenting a practical-example of how advanced analytics can be employed to counteract cyber enabled financial crimes, while maintaining transparency-accountability. Therefore, the study holds significance beyond its immediately applicable technical contributions, as it provides a scalable model for how data driven supervision can be synchronized with national-interests in financial-stability, consumer protection, and integrity of digital-asset-markets.

**Funding:** This research received no external funding.
**Conflicts of Interest:** The authors declare no conflict of interest.


## References

[1] M. Conti, S. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of Bitcoin," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3416–3452, 2018.
[2] G. Mahadevan et al., "Bitcoin Heist Ransomware Address Dataset," Kaggle Open Dataset, 2018. [Online]. Available: https://www.kaggle.com/datasets/gopalmahadevan/bitcoin-heist-ransomware-address-dataset
[3] M. Weber, G. Domeniconi, J. Chen, et al., "Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 2364–2373, 2019.
[4] C. F. Akcora, M. Dixon, et al., "Blockchain data analytics: Challenges and opportunities," *IEEE Data Eng. Bull.*, vol. 42, no. 1, pp. 16–23, 2019.
[5] M. Weber, J. Chen, et al., "Detecting illicit cryptocurrency flows," *Digital Finance*, vol. 1, no. 1–4, pp. 41–61, 2019.
[6] E. Goldsmith et al., "Blockchain forensics and transaction clustering," *J. Financial Crime*, vol. 27, no. 2, pp. 401–420, 2020.
[7] T. Li, P. Chen, and D. Zeng, "Survey of financial fraud detection methodologies," *J. Financial Data Science*, vol. 2, no. 4, pp. 1–18, 2020.
[8] N. Carter and J. Lee, "Integrated analytics for AML performance gains," *J. Financial Crime*, vol. 27, no. 3, pp. 745–760, 2020.
[9] Global Association of Risk Professionals (GARP), *Survey of AML Technology Adoption*, Industry Report, 2021.
[10] Financial Crimes Enforcement Network (FinCEN), *National AML/CFT Priorities*, U.S. Department of the Treasury, 2021.
[11] Financial Action Task Force (FATF), *Guidance for a Risk-Based Approach to Virtual Assets and VASPs*, Paris, France, 2021.
[12] J. Xu, W. Chen, and Z. Yang, "Graph-based financial fraud detection," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 10, pp. 3288–3301, 2021.
[13] P. R. Cunha, P. Soja, and M. Themistocleous, "Blockchain for development: A guiding framework," *Inf. Technol. for Development*, vol. 27, no. 4, pp. 1–21, 2021.
[14] U.S. Congress, *Anti-Money Laundering Act of 2020*, enacted Jan. 2021.
[15] Z. Chen and X. Wang, "Explainable machine learning for blockchain forensics," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 3, pp. 698–709, 2022.
[16] A. Rao, M. Singh, and R. Patel, "Explainable AI for financial compliance applications," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 4, pp. 1–25, 2022.
[17] C. Mendez, P. Oster, and J. Alvarez, "Hybrid financial intelligence architectures for anomaly detection," *Inf. Syst. Frontiers*, vol. 24, no. 5, pp. 1321–1336, 2022.
[18] The White House, *Executive Order 14067: Ensuring Responsible Development of Digital Assets*, Washington, DC, USA, 2022.
[19] U.S. Government Accountability Office (GAO), *Ransomware: Trends and Challenges for Financial Systems*, GAO Report, 2022.