ANNOTATION

of the doctoral dissertation by Shaizat Medet Zhanbolatuly on the topic: "Development of Methods and Machine Learning Models for the Cyber-Forensic Analysis of Bitcoin Transactions", submitted in fulfillment of the requirements for the Doctor of Philosophy (PhD) degree in the educational program "8D06301 – Information Security Systems"

Relevance of the research topic. The analysis of digital asset transactions is important for legal regulation, financial monitoring, and digital forensics. For the Republic of Kazakhstan, this issue is particularly relevant due to the Law of the Republic of Kazakhstan «On Digital Assets in the Republic of Kazakhstan», which creates a practical need for tools to control and analyses the relevant operations, including in financial monitoring and the investigation of offences.

The international significance of the problem is confirmed by the fact that the implementation of the FATF standards regarding virtual assets and providers of the corresponding services remains incomplete and uneven. Difficulties persist in risk assessment, supervision, and the application of the Travel Rule, which increases the need for analytical tools for transaction monitoring.

The relevance of the research is also determined by the scale of illicit activity in the cryptocurrency environment. According to Chainalysis, in 2025, addresses associated with illicit activity received at least USD 154 billion. Despite the relatively small share of such operations in the total volume of attributed cryptocurrency transactions, their absolute volume remains significant and requires the development of methods for detecting, analyzing, and documenting the corresponding flows.

The choice of Bitcoin is determined by the public architecture of the network, which makes it possible to analyses the transaction graph, trace the movement of funds, and study behavioral patterns while preserving the pseudonymity of participants. However, the availability of blockchain data does not guarantee their unambiguous interpretation. The results of existing methods depend on the completeness of labelling and attribution procedures, which complicate their verification and comparison and increase the risk of erroneous conclusions.

Thus, the scientific problem lies in the insufficient development of methods that make it possible to detect Bitcoin transactions with indicators of illicit activity and explain the analysis results while considering the transactional, temporal, and graph context.

Connection of the research topic with research program: The dissertation was carried out within the framework of the «Zhas Galym» grant funding program for young scientists for 2025–2027 under project AP25793719 «Development of a System for Analyzing Illicit Cryptocurrency Transactions», implemented within the priority area «National Security and Defense, Biological Safety» and the specialized area «Ensuring Information Security».

Research aim. The aim of the dissertation research is to develop methods for constructing interpretable features and machine learning models for the cyber-forensic analysis of Bitcoin transactions.

Research objectives. To achieve the stated aim, the following objectives are addressed in the dissertation:

1. To analyses existing approaches to the cyber-forensic analysis of Bitcoin transactions based on machine learning and graph analysis.
2. To investigate existing Bitcoin transaction datasets and methods for constructing transaction graphs and to determine their main limitations.
3. To develop a genetic algorithm-based feature selection method for the anonymized Elliptic and Elliptic++ datasets.
4. To conduct a comparative evaluation of tabular and graph models for detecting illicit Bitcoin transactions.
5. To construct a specialized Bitcoin transaction dataset with semantically transparent features based on data from the public Bitcoin blockchain.
6. To develop a system of forensically interpretable features comprising features of local deviation from the neighborhood, temporal features of subsequent spending of funds, and features of prior transaction history.
7. To develop and evaluate machine learning models on the constructed dataset, determine the contribution of the developed features, and interpret the classification results.
8. To develop a software system for the cyber-forensic analysis of Bitcoin transactions that integrates transaction classification and address-based, behavioral, and graph analysis.

Object of the research: the process of cyber-forensic analysis of Bitcoin transactions.

Subject of the research: methods for feature construction and selection, as well as machine learning models for the cyber-forensic analysis of Bitcoin transactions.

Research methods. The study employs systems and comparative analysis, a genetic algorithm, and methods of data analysis, machine learning, and model interpretation.

Scientific novelty of the research

A system comprising 61 forensically interpretable descriptors of Bitcoin transactions that formalize patterns of fund movement identified through the analysis of practical cases is proposed. The features characterize local deviation from the graph neighborhood, the dynamics of subsequent spending of outputs, and the deviation of current characteristics from values expected based on prior transaction history. Adding them to the 26 baseline features improved the performance of all tested models: the average increase in PR-AUC was 3.97 percentage points, while that in F1-score was 4.26 percentage points. Ablation analysis confirmed the contribution of all three feature groups.

A CatBoost-based model for detecting illicit Bitcoin transactions was developed using the constructed dataset with the disclosed semantics of 87 features instead of the anonymized Elliptic characteristics. The graph structure, temporal

slices, and Elliptic labels were retained, public transaction identifiers were restored, and the feature computation procedure was formally defined. The model achieved average values of PR-AUC 90.27 %, F1-score 83.90 %, and Recall 81.11 %.

A wrapper-based method for selecting anonymized features is proposed, based on genetic search and the evaluation of subsets by a Random Forest model using the F1-score of the illicit class. Selection is performed only on the training temporal slices. On the Elliptic dataset, the method reduced the feature space from 166 to 53 features and achieved Precision 99.37 %, Recall 73.13 %, and F1-score 84.26 %, exceeding the published RF+GCN results by 2.3, 5.6, and 4.7 percentage points, respectively. On Elliptic++, 85 of 183 features were selected, and the F1-score exceeded the result obtained with the full feature set by 1.57 percentage points.

The theoretical significance of the research lies in advancing the theoretical and methodological foundations of interpretable cyber-forensic analysis of Bitcoin transactions through the formalization of features of local deviation from the neighborhood, temporal features of subsequent spending, and features of prior transaction history. The proposed system combines these features with transactional and structural-graph characteristics, making it possible to consider transaction parameters, its position in the graph, deviation from the local neighborhood, the dynamics of subsequent spending, and deviation from expected values calculated from prior transaction history. The developed features are computed directly from blockchain data and have a meaningful interpretation.

The practical significance of the research lies in the possibility of applying the developed software and analytical tools to prepare data on Bitcoin transactions, analyses and classify them, interprets model results, and presents the resulting conclusions for expert review. The developed tools can be used by financial monitoring units, law enforcement agencies, cryptocurrency exchanges, and organizations engaged in combating money laundering and investigating crimes involving cryptocurrencies.

Provisions submitted for defense:

1. The system of forensically interpretable features of Bitcoin transactions, which formalizes forensic hypotheses, improves the quality of detecting illicit transactions: its inclusion in the baseline feature representation produced increases in PR-AUC and F1-score for all tested models, with the maximum increases amounting to 5.19 percentage points in PR-AUC and 7.46 percentage points in F1-score.
2. The CatBoost-based model for detecting illicit Bitcoin transactions is distinguished by being trained on the constructed dataset with semantically transparent transactional, structural-graph, and forensically interpretable features and by not using the anonymized Elliptic features, which enables its application to the initial analysis of real Bitcoin transactions; the achieved PR-AUC, F1-score, and Recall values were 90.27 %, 83.90 %, and 81.11 %, respectively.
3. The feature selection method for blockchain datasets with anonymized features, based on the genetic search for feature subsets and their wrapper evaluation by a Random Forest model using the F1-score of the illicit class,

when applied to the Elliptic dataset, made it possible to construct a subset of 53 features instead of the original 166, while the Precision, Recall, and F1-score values of the Random Forest model on the selected subset exceeded the published results of the RF+GCN model of the original Elliptic benchmark by 2.3, 5.6, and 4.7 percentage points, respectively.

Approbation of the dissertation results. The research results were discussed at scientific seminars of the Faculty of Information Technology and Artificial Intelligence and the Department of Cybersecurity and Cryptology of Al-Farabi Kazakh National University and were also presented at the «Farabi Alemi» conference in the report « Биткоин желісіндегі ауытқуларды анықтау әдісі» (Almaty, 2022).

Publications related to the dissertation: The results of the dissertation research were published in 5 scientific works, including 2 articles in journals indexed in the Scopus database, 2 articles in publications recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan, and 1 paper in the proceedings of an international scientific conference:

Scientific articles published in journals indexed in the Scopus database:

1. Shaizat M., Mussiraliyeva S., Tereikovskyi I. Forensically Interpretable Graph Descriptors for Improved Illicit Bitcoin Transaction Detection // International Journal of Computer Network and Information Security. – 2026. – Vol. 18, No. 3. – P. 63–84. The journal percentile according to Scopus data is 72.
2. Shaizat M., Mussiraliyeva S. Enhanced Identification of Illicit Bitcoin Transactions Through Genetic Algorithm-Based Feature Selection // Eastern-European Journal of Enterprise Technologies. – 2025. – Vol. 4, No. 9 (136). – P. 34–42. The journal percentile according to Scopus data is 57.

Scientific articles published in journals recommended by the Committee for Quality Assurance in Science and Higher Education:

3. Мусиралиева Ш. Ж., Шайзат М. Ж., Бекетова А. К., Абайұлы Е., Манасова А. Б. О методе идентификации подозрительных транзакций в Биткоин сети / Identification of Suspicious Transactions in the Bitcoin Network // Известия Национальной академии наук Республики Казахстан. Серия физико-математическая. – 2023. – № 1 (345). – С. 154–169.
4. Мусиралиева Ш. Ж., Шайзат М. Ж., Бекетова А. К., Манасова А. Б. Биткоин желісінде күдікті транзакцияларды анықтау: талдау, белгілер және машиналық оқыту алгоритмдері // Вестник Национальной инженерной академии Республики Казахстан. – 2023. – № 3 (89).

Scientific papers published in the proceedings of international conferences:

5. Елтай Ж., Шайзат М., Бекетова А. Биткоин желісіндегі ауытқуларды анықтау әдісі // Материалы Международной научной конференции студентов и молодых ученых «Фараби әлемі». – Алматы, Казахстан, 2022. – С. 42.

Scope and structure of the dissertation

The dissertation consists of an introduction, four sections, a conclusion, and a list of references. The first section considers the foundations of the cyber-forensic analysis of Bitcoin transactions, existing methods, and datasets. The second section develops a genetic algorithm-based method for selecting anonymized features and compares tabular and graph models. The third section constructs a dataset with semantically transparent features, develops interpretable descriptors, and evaluates machine learning models. The fourth section develops hybrid architecture for the cyber-forensic analysis of Bitcoin transactions and describes its implementation as a web application. The total length of the work is 114 pages, including 41 figures and 20 tables. The list of references contains 104 sources.

Author's personal contribution

The author's personal contribution consists in analyzing scientific literature, developing methods for feature construction and selection, preparing the dataset, building machine learning models, implementing the proposed approach in software, and interpreting the obtained results. The author also prepared scientific publications on the dissertation topic.