УДК 004.056

Н. Карапетьянц[1]✉

Исследования методов анализа транзакций сети Bitcoin

[1]Национальный исследовательский ядерный университет «МИФИ»,
г. Москва, Российская Федерация
e-mail: nkarapetyants@mephi.ru

Аннотация. В данной статье представлено описание этапов процесса анализа транзакций сети Bitcoin и применяемых методов. Цель данного исследования – оценить применимость существующих методов анализа транзакций сети Bitcoin. В работе приводится перечень и описание методов анализа транзакций в сети Bitcoin. Результаты исследований показали, что рассмотренных инструментов недостаточно для полной автоматизации процессов сбора, агрегации и обработки данных транзакций. Рассматривается реализация хранения данных транзакций сети Bitcoin с использованием СУБД PostgreSQL и плагина Citus. Также приводятся рекомендации по оптимизации процессов сбора данных из Bitcoin Core и агрегации данных в СУБД PostgreSQL. Результаты настоящей работы предоставляют возможность осуществить автоматизацию процессов сбора, обработки и агрегации данных сети Bitcoin для разработки новых методов анализа транзакций сети Bitcoin, используемых в незаконной деятельности, и их источников.

Ключевые слова: Bitcoin, Blockchain, KYT, мониторинг, анализ данных

N. Karapetyants[1]✉

Research of transaction analysis methods of the Bitcoin network

[1]National Research Nuclear University MEPhI, Moscow, Russian Federation
e-mail: nkarapetyants@mephi.ru

Abstract. This article delineates the stages of Bitcoin network transaction analysis and the methodologies employed. The objective of this study is to evaluate the applicability of existing methods for analyzing Bitcoin network transactions. The work presents a comprehensive list and description of transaction analysis methods utilized within the Bitcoin network. The research findings revealed that the examined tools are insufficient to fully automate the processes of data collection, aggregation, and processing of transaction records. The implementation of a Bitcoin transaction data storage system using the PostgreSQL database management system (DBMS) and the Citus extension is discussed. Additionally, recommendations are provided for optimizing data collection procedures from Bitcoin Core and enhancing data aggregation within the PostgreSQL DBMS. The results of this study establish a foundation for automating the collection, processing, and aggregation of Bitcoin network data, thereby facilitating the development of novel methods to analyze transactions and their sources associated with illicit activities.

Keywords: Bitcoin, Blockchain, KYT, Monitoring, Data analysis

Введение

Использование децентрализованной сети блокчейн Bitcoin в качестве финансово-технического средства несет в себе определенные риски. При проведении транзакций необходимо соблюдать требования международных и государственных регуляторов. Законодательная база Российской Федерации, как и во

всем мире, отстает от развития этой области. Но, несмотря на это, регуляторными органами, такими как Центральный Банк, Росфинмониторинг, МВД, ФНС разрабатываются нормативно-правовые акты и инструменты для проведения транснациональных платежей с использованием криптовалют.

Одним из подходов для идентификации средств и их источников в сети Bitcoin, применяемых биржами и интернет-магазинами, использующими криптовалюту в качестве средств платежа, является процедура «Знай свою транзакцию» (англ. KYT, «Know Your Transaction»). Данная процедура позволяет произвести анализ и оценку рисков при проведении платежей с учетом отсутствия персональных данных контрагента, как это осуществляется в процедуре «Знай своего клиента» (англ. KYC, «Know Your Client»). Сбор информации о транзакции и самом контрагенте для проведения процедуры KYT осуществляется из общедоступных источников сети Интернет, в том числе и из источников оверлейных сетей, таких как Tor. Из-за недостатка информации в инструментах, реализующих процедуру KYT, возникает необходимость в создании интеллектуальных методов анализа данных транзакций, которые позволят повысить эффективность процесса оценки рисков при проведении криптовалютных платежных операций.

В данной области проводят исследования следующие ученые: Басыня Е. А. [1–3], Худяков Д.С [1], Стрельцов А. С. [2], Французова Г. А. [2], Сафронов А. [3], Фельдман Е. В. [4], Ручай, А. Н. [4], Матвеева, В. К. [4], Самсонова, В. Д. [4], Олссон А. [5], Андерссон Д. [5].

В предшествующей работе [6] был реализован сценарий автоматизированного развертывания виртуального испытательного стенда для проведения экспериментов в области анализа транзакции с использованием средств автоматизации Ansbile и Terraform в среде виртуализации Proxmox Virtual Environment. В данной работе приводится исследование методов анализа транзакций Bitcoin, что расширяет возможности применимости ранее разработанного виртуального стенда в рамках разработки решений по анализу криптовалютных транзакций.

Методы и материалы

Процесс анализа транзакций в рамках виртуального испытательного стенда включает себя четыре этапа:

- сбор данных о транзакциях из источников;
- агрегация в систему хранения (база данных);
- обработка данных;
- анализ обработанных данных.

Этап сбора данных заключается в извлечении информации из клиента Bitcoin Core с использованием протокола JSON RPC, а также из других источников сети интернет путем прямого парсинга веб-страниц или интерфейса API, если веб-сайт его предоставляет. Скачивание всей истории транзакций сети Bitcoin с помощью Bitcoin Core, размер которого на данный момент составляет приблизительно 650 гигабайт, занимает приблизительно 48 часов при скорости доступа в интернет 1 Гбит/с. Получение самих данных о транзакциях может быть осуществлено с использованием библиотеки «Bitcoin ETL» [7], написанной на

языке Python. Получение информации из других источников требует реализации программы парсинга, который будет в фоновом режиме запрашивать необходимые страницы с ресурса, а после извлекать нужные данные в соответствии с заданным шаблоном. Подобная программа парсинга может быть реализована с использованием библиотек, предоставляющих методы для реализации собственного поискового робота с необходимым функционалом. Примером ресурса, для которого можно реализовать программу парсинга, является walletexplorer.com. Данный веб-ресурс содержит информацию о транзакциях, адресах, а также перечень кошельков и их владельцев сети Bitcoin.

На *этапе агрегации* осуществляется сохранение данных транзакций сети Bitcoin в систему хранения. В качестве системы хранения может выступать любая система управления базами данных (СУБД). В качестве основной СУБД в разработанном ранее испытательном стенде была выбрана Apache Cassandra по двум причинам: наличие функции горизонтального масштабирования и возможность интеграции с Apache Spark. В результате проведенного эксперимента выяснилось, что выгрузка данных из Bitcoin Core осуществляется со скоростью в среднем 1 блок в секунду после 200 000 блока. Это связано с медленной работой самого интерпретатора Python, а также с большим количеством транзакций в блоке. Уменьшение времени загрузки блока в базу данных может быть достигнуто путем реализации многопоточности, которая не требует существенных изменений в коде за счет использования встроенной библиотеки multiprocessing.

Для повышения производительности при работе с данными транзакций Bitcoin предлагается заменить Apache Cassandra на модифицированную версию PostgreSQL. Стандартная реализация PostgreSQL имеет ограничения в этом сценарии: ее архитектура и подход к хранению данных плохо адаптируются к структуре блокчейн-транзакций, которые подразумевают большое количество динамических полей и высокую частоту записей. Эти ограничения связаны с жесткой схемой таблиц, отсутствием встроенной горизонтальной масштабируемости и особенностями оптимизации под OLTP-нагрузки. Модификации PostgreSQL могут включать внедрение расширений для работы с JSONB-полями, шардинга данных или интеграции с колоночными хранилищами, что позволит эффективнее обрабатывать специфические требования Bitcoin-транзакций. Для сохранения колоночной структуры и горизонтальной масштабируемости, как это реализовано в Apache Cassandra, можно использовать плагин с открытым исходным кодом Citus [8]. Он позволяет организовать таблицы в виде колоночной структуры данных и распределять данные между несколькими серверами PostgreSQL.

В рамках *этапа обработки* производится модификация структуры данных в соответствии с требованиями методов, используемых в рамках анализа. Полученные данные о транзакциях в рамках этапа сбора из источника Bitcoin Core не позволяют напрямую оперировать адресами входов и выходов транзакций. Для этого необходимо осуществить преобразования таблиц в соответствии с предлагаемым решением Graphsense [9]. Для анализа транзакций с использованием графовых СУБД, таких как Neo4j, необходимо также произвести обработку исходных данных [10]. Для осуществления потоковой обработки данных в реальном времени в рамках исследо-

ваний больших данных и машинного обучения используется Apache Spark, который в качестве источника данных поддерживает PostgreSQL и Neo4j [11].

Этап анализа включает в себя построение моделей с использованием различных методов: визуальный анализ, кластеризация адресов, классификация. Методы визуального анализа осуществляются с использованием графов транзакций и адресов, что позволяет отслеживать цепочку транзакций для выявления кошельков и их владельцев, связанных с незаконной деятельностью. Поскольку в Bitcoin нельзя осуществить идентификацию пользователей сети, применяются методы классификации и кластеризации адресов. Методы кластеризации позволяют определить принадлежность адреса к конкретному кошельку. Используемые в данном методе признаки основаны на определенных эвристических правилах, где каждое отдельное правило не дает точного результата, в отличии от их совокупного использования [12–14]. В свою очередь, методы классификации позволяют определить принадлежность адресов и их владельцев к конкретному виду деятельности, например: биржи, интернет-магазин, использование сервисов для запутывания цепочек транзакций и т. д. В таком случае производится анализ моделей поведения различных типов адресов Bitcoin с использованием конкретных признаков [15–17].

Результаты и обсуждения

Исследование методов анализа транзакции сети Bitcoin позволило определить четыре этапа процесса анализа: сбор, агрегация, обработка и анализ.

В рамках первых двух этапов было произведено развертывание инфраструктуры, содержащей в себе четыре сервиса: Bitcoin Core, кластер Postgresql с плагином Citus, сервис получения данных транзакций сети Bitcoin на основе инструмента bitcoint-etl с поддержкой многопоточности, а также парсер интернет-сервиса walletexplorer.com. Было проведено сравнительное тестирование предлагаемой ранее СУБД Apache Cassandra и PostgreSQL с плагином, результаты которого представлены в табл. 1.

Таблица 1

Результаты сравнительного тестирования баз данных при сборе данных с ресурса walletexplorer.com

Критерий	Apache Cassandra	PostgreSQL+Citus+ Multiprocessing
Общее количество блоков, шт.	800 000	
Среднее количество блоков в секунду	1	6
Объем базы данных, ГБ	4134	1033
Количество узлов	2	2

Сравнительное тестирование методов выполнения SQL-запросов через библиотеку psycopg2 показало, что использование функции «copy_from» для пакетной вставки данных работает быстрее, чем последовательное выполнение операторов «execute». Это связано с оптимизацией транзакций при групповой загрузке. В рамках эксперимента удалось собрать 37623351 адрес кошельков с сервиса WalletExplorer.com, однако процесс занял больше времени из-за ограничений лимита на количество одновременных запросов со стороны веб-ресурса, что существенно замедлило сбор данных. Для обхода этих ограничений потребовалась дополнительная реализация динамических пауз между запросами.

Заключение

В ходе проведенного исследования были рассмотрены методы анализа транзакций сети Bitcoin и их применимость на каждом из этапов процесса анализа. Практическим результатом работы являются результаты сравнительного анализа СУБД Apache Cassandra и PostgreSQL с точки зрения хранения данных о транзакциях в сети Bitcoin, а также перечень методов и инструментов, которые могут быть использованы для улучшения существующих или разработки новых инструментов KYT. Также были получены данные из Bitcoin Core и сервиса walletexplorer.com.

В рамках дальнейших исследований планируется воспроизвести результаты исследований методов кластеризации и классификации адресов сети Bitcoin.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Басыня Е. А., Худяков Д. С. Комплексный мониторинг информационной инфраструктуры предприятия //Защита информации. Инсайд. – 2020. – №. 6. – С. 28-33.

2. Стрельцов А. С., Французова Г. А., Басыня Е. А. Разработка системы сбора, обработки, анализа, идентификации корреляции событий информационной инфраструктуры предприятия //Системы анализа и обработки данных. – 2023. – №. 1 (89). – С. 101-113.

3. Басыня Е. А., Сафронов А. Метод формирования децентрализованного реестра событий информационной инфраструктуры предприятия //Bulletin of LN Gumilyov Eurasian National University Technical Science and Technology Series. – 2022. – Т. 139. – №. 2. – С. 40-50.

4. Фельдман Е. В., Ручай, А. Н., Матвеева В. К., Самсонова В. Д. Модель выявления аномальных транзакций биткоинов на основе машинного обучения //Челябинский физико-математический журнал. – 2021. – Т. 6. – №. 1. – С. 119-132.

5. Olsson A., Andersson D. The dark flows of cryptocurrency: an overview of money flow behaviors in bitcoin transactions related to online criminal activities and bitcoin mixers. – 2024.

6. Карапетьянц Н. Виртуальный испытательный стенд для проведения экспериментов в области анализа транзакций сети Bitcoin //ИНТЕРЭКСПО ГЕО-СИБИРЬ Учредители: Сибирский государственный университет геосистем и технологий. – 2024. – Т. 7. – №. 1. – С. 140-146.

7. Bitcoin ETL // Github. URL: https://github.com/blockchain-etl/bitcoin-etl (дата обращения: 01.05.2025).

8. Citus // Github. URL: https://github.com/citusdata/citus (дата обращения: 01.05.2025).

9. Haslhofer B. et al. Graphsense: A general-purpose cryptoasset analytics platform //arXiv preprint arXiv:2102.13613. – 2021.

10. Наха Р.Т., Чжан К. Криминалистика криптовалют с использованием аналитики в реальном времени и графовой базы данных: всесторонний обзор // Международная конференция IEEE по большим данным (BigData) 2024 г. – IEEE, 2024. – С. 1-12.

11. Саранчук Д. и др. Интеграция графовых технологий в SQL-базы данных //Conferinţa tehnico-ştiinţifică a studenţilor, masteranzilor şi doctoranzilor. – 2024. – Т. 1. – С. 612-617.

12. Лю Ф. и др. Кластеризация адресов Bitcoin на основе улучшения адреса смены // IEEE Transactions on Computational Social Systems. – 2023.

13. Чжао Ц. и др. Улучшение кластеризации адресов в биткойне с помощью предложения эвристик // IEEE Transactions on Network and Service Management. – 2022. – Т. 19. – №. 4. – С. 3737-3749.

14. Хе Х. и др. Метод кластеризации адресов Bitcoin на основе множественных эвристических условий //IET Blockchain. – 2022. – Т. 2. – №. 2. – С. 44-56.

15. Liu F. et al. Bitcoin address clustering based on change address improvement //IEEE Transactions on Computational Social Systems. – 2023.

16. Febrero-Bande M. et al. Functional classification of bitcoin addresses //Computational Statistics & Data Analysis. – 2023. – Т. 181. – С. 107687.

17. Xiang Y. et al. Babd: A bitcoin address behavior dataset for pattern analysis //IEEE Transactions on Information Forensics and Security. – 2023. – Т. 19. – С. 2171-2185.