УДК 004.056

Н. Карапетьянц[1]*

Инструменты анализа транзакций сети Bitcoin с открытым исходным кодом

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

Аннотация. В данной статье рассматривается текущее состояние инструментов для анализа транзакций сети Bitcoin с открытым исходным кодом трех категорий: сбор, анализ, визуализация. Цель данного исследования – провести анализ функциональных возможностей существующих инструментов, исходный код которых доступен в сети Интернет. Результаты исследования показали, что каждый из рассмотренных инструментов обладает уникальными возможностями и решает определенную задачу. Полученные результаты работы позволяют определиться с набором инструментов в зависимости от поставленной задачи. К примеру, некоторые инструменты представляют собой библиотеки Python, которые можно использовать для изучения и проведения соответствующих экспериментов. Проведенный анализ представляет текущее состояние существующих инструментов анализа транзакций сети Bitcoin. Результаты проведенной работы могут быть использованы для разработки передовых инструментов и методов идентификации средств, связанных с незаконной деятельностью, и их источников.

Ключевые слова: Bitcoin, Blockchain, KYC, KYT, ETL

N. Karapetyants[1]*

Open source Bitcoin network transaction analysis tools

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

Abstract. This article discusses the current state of open source tools for analyzing Bitcoin transactions in three categories: collection, analysis, visualization. The purpose of this study is to analyze the functionality of existing tools, the source code of which is available from the Internet. The results of the study showed that each of the considered tools has unique capabilities and solves a specific problem. The obtained results of the work allow us to determine the set of tools depending on the task. For example, some tools are Python libraries that you can use to learn and experiment with. The analysis carried out represents the current state of existing Bitcoin network transaction analysis tools. The results of this work can be used to develop advanced tools and methods for identifying funds associated with illegal activities and their sources.

Keywords: Bitcoin, Blockchain, KYC, KYT, ETL

Введение

Пиринговая платежная система Bitcoin, работающая на основе распределенной технологии блокчейн, обеспечивает прозрачность проводимых операций сети и, соответственно, децентрализованный доступ. Обеспечить полную анонимность пользователей сети невозможно ввиду двух причин: особенности про-

граммной реализации протокола обмена данными между участниками и несоблюдение пользователями цифровой гигиены. Таким образом, сопоставление информации о пользователе из общедоступной сети Интернет и приложения Bitcoin Core позволяет с высокой степенью достоверности идентифицировать пользователей, но при условии использования соответствующих средств и методов интеллектуальной обработки больших данных.

Основными клиентами инструментов анализа транзакций сети Bitcoin являются правоохранительные органы, регуляторы и биржи. Правоохранительные органы используют инструменты анализа транзакций Bitcoin для расследования преступлений, связанных с отмыванием денег, финансированием терроризма, киберпреступлений и других видов незаконной деятельности. С точки зрения регулятора, например, центрального банка, контроль за криптовалютой осуществляется с точки зрения обеспечения экономической безопасности в сфере кредитно-финансовой системы. Во многих развитых странах для проведения той или иной операции, связанной с денежными средствами, должна осуществляться процедура «Знай своего клиента» (англ. KYC, Know Your Client), а в некоторых случаях осуществляется процесс «Знай свою транзакцию» (англ. KYT, Know Your Transaction). Данные требования направлены на борьбу с отмыванием денег.

В этой статье рассматриваются инструменты анализа транзакций Bitcoin с точки зрения их применения в исследовательском сообществе и сообществе разработчиков. На основе проведенного анализа научных статей, индексируемых в базах РИНЦ, Web of Science, Scopus за последние пять лет, был составлен перечень инструментов, их описание и область применения.

В данной области проводят исследования Басыня Е. А. [1, 2], Сердечный А.Л. [3], Родивилина В. А., Родивилин И. П. [4], Сонг В. [5], Мун, Х. [6], Бакуменко Л.П. [7], Фельдман Е. В., Ручай А.Н. [8] и др.

Методы и материалы

На первых этапах процесса анализа транзакций сети Bitcoin осуществляется сбор, обработка и агрегация данных о транзакциях и участниках сети из разных источников. Источники данных о транзакциях можно разбить на три группы: данные из сети Bitcoin, общедоступная сеть Интернет, провайдер данных. Данные о транзакциях сети Bitcoin можно получить с помощью толстого клиента Bitcoin Core, которых хранят всю актуальную информацию сети. На данный момент общий объем информации главной ветки сети достигает более 450 гигабайт. Информацию о владельцах можно найти в общедоступной сети Интернет. В таких случаях используются поисковые роботы, которые осуществляют автоматизированный поиск определенной информации, например, адрес кошелька, среди ресурсов сети Интернет. Структурированная информация может быть предоставлена сторонним поставщиком данных. Доступ к такой информации, как правило, осуществляется с помощью API, предоставляемой поставщиком данных. С учетом всего объема данных о транзакциях и участниках сети блокчейн возникает необходимость в использовании специализированных инструментов для работы с большими данными.

Получение данных о транзакциях с помощью толстого клиента Bitcoin Core осуществляется двумя способами: использование встроенного протокола JSON RPC и прямое извлечение структур данных из файлов вида “blkXXXXX.dat”. Первый способ значительно проще, поскольку требует осуществления запросов RPC, в ответ на которые сервер Bitcoin Core RPC [9] выдает информацию о блоках и транзакциях в текстовом формате JSON. Преимуществом данного способа является последовательное получение актуальной информации о транзакциях в сети Bitcoin, не требующее высоких навыков программирования. Недостатком является низкая скорость, когда речь идет о переводе в реляционную базу данных. В таком случае заполнение базы данных может занять до нескольких дней. Чтобы ускорить процесс необходимо воспользоваться алгоритмами, написанными на быстрых компилируемых языках программирования – C++ или Rust. Примером таких утилит является fast-dat-parser [10], который позволяет извлекать данные транзакций со скоростью до 450 Мб в секунду при использовании твердотельного накопителя. Последний коммит проекта датируется мартом 2020 года и может быть использован в качестве примера для реализации быстрых алгоритмов извлечения данных о транзакциях с учетом новых актуальных версий протокола Bitcoin.

Извлечение информации о владельцах кошельков Bitcoin из общедоступных ресурсов сети Интернет предполагает использование поисковых роботов. Среди распространенных поисковых роботов с открытым исходным кодом можно выделить следующие: Apache Solr [11], Elasticsearch [12].

Apache Solr является платформой поиска на основе Apache Lucene с открытым исходным кодом и обладает следующими преимуществами: высокая скорость работы, масштабируемость, высокая гибкость настройки. К недостаткам можно отнести отсутствие визуального поиска.

Elasticsearch основан на базе Lucene, распространяется с открытым исходным кодом, обладает большими возможностями конфигурации и настройки, имеет возможность масштабирования и значительно быстрее, чем многие другие системы, но требует больше ресурсов и не всегда корректно работает с кириллицей.

Выбор между Apache Solr и Elasticsearch зависит от объема и типа анализируемых данных транзакций сети Bitcoin. В случае, если необходимо провести анализ структурированных и неструктурированных данных, лучше всего подойдет Elasticsearch. В остальных нетребовательных задачах целесообразнее использовать Apache Solr.

Задачей определения принадлежности адресов их владельцам является кластеризация. К инструментам для осуществления кластеризации адресов сети Bitcoin с открытым кодом можно отнести следующие проекты: Blockhair [13], OXT [14], pycoin [15].

Программа Blockchair используется для проведения кластеризации адресов Bitcoin. Может использоваться для анализа крупных объемов данных. Главным недостатком можно считать низкую скорость работы при обработке очень большого объема данных.

Инструмент OXT помимо кластеризации адресов Bitcoin предоставляет инструменты визуализации и построения наглядных графиков. К недостаткам относится низкая скорость работы при обработке очень большого объема данных, а в некоторых случаях результаты чуть менее точны по сравнению с другими инструментами.

Pycoin является библиотекой Python и позволяет осуществлять кластеризацию сети с использованием различных алгоритмов кластеризации. По сравнению с другими инструментами обладает высокой гибкостью, но требует больше вычислительных ресурсов при использовании некоторых алгоритмов кластеризации и требует владения навыками программирования на языке Python.

Отображение результатов кластеризации или других результатов анализа транзакций сети осуществляется специализированными программами визуализации с открытым исходным кодом: Blockstream [16], Bitbonkers [17], 3D Chain [18].

Blockstream помимо визуализации обладает широким набором функционала для анализа блокчейна Bitcoin, предоставляет информацию о транзакциях в реальном времени и имеет удобный интерфейс. Но данный инструмент не позволяет проводить анализ построенного графа транзакций.

Инструмент Bitbonkers обладает уникальной по сравнению с другими инструментами данной категории функцией интерактивной визуализации и предоставляет информацию о новых блоках в режиме реального времени.

Уникальным инструментом визуализации является 3D Chain, который позволяет визуализировать транзакции в трехмерном пространстве, но в тоже время обладает небольшим функционалом по сравнению с другими инструментами визуализации.

Результаты

В результате проведенного анализа инструментов сбора, анализа и визуализации были выявлены их основные преимущества и ограничения, которые представлены в табл. 1.

Результаты анализа показали, что существует множество инструментов, предназначенных для осуществления сборки, анализа и визуализации транзакций блокчейн Bticoin. У каждого из них имеются свои преимущества и ограничения. Из всех рассмотренных инструментов стоит выделить fast-dat-parser, Elasticsearch, pycoin, 3D Chain, которые обладают уникальными функциональными возможностями.

Таблица 1

Сравнительный анализ

Инструмент	Преимущества	Ограничения
Инструменты сбора данных о транзакциях из сети Bitcoin		
Bitcoin Core RPC	Возможность получения проверенных данных о транзакциях Bitcoin в реальном времени. Можно легко реализовать на языке программирования Python	Низкая скорость работы. Занимает много времени в случае, если требуется получить данные сразу всех транзакций сети Bitcoin
fast-dat-parser	Высокая скорость работы благодаря использованию языка программирования C++ и прямого извлечения данных из файлов	Требует наличия навыков программирования на языке C++
Инструменты сбора данных об участниках сети Bitcoin в общедоступной сети Интернет		
Apache Solr	Расширенные возможности конфигурации и настроек, масштабируемость, высокая гибкость настройки	Отсутствие визуального поиска
Elasticsearch	Высокие скорость работы, масштабируемость, гибкость настройки	Требует больших ресурсов. Не всегда корректно работает с кириллицей
Инструменты кластеризации адресов сети Bitcoin		
Blockhair	Наличие возможности обработки большого объема данных. Поддерживает кластеризацию других криптовалют. Легко интегрируется с другими продуктами	Низкая скорость работы при обработке крупных объемов данных
OXT	Обработка крупных объемов данных. Встроенные инструменты визуализации	В некоторых случаях может быть менее точен по сравнению с другими инструментами. Показывает низкую скорость работы при обработке сверхвысоких объемов данных
pycoin	Включает несколько алгоритмов кластеризации транзакций Bitcoin. Легко интегрируется с другими библиотеками Python	Требует знания программирования на Python. В некоторых случаях требует больше ресурсов для работы, чем другие инструменты
Инструменты визуализации транзакций сети Bitcoin		
Blockstream	Позволяет отслеживать данные транзакций и смотреть все узлы блок-	Не позволяет проводить анализ графов

Инструмент	Преимущества	Ограничения
	чейна в реальном времени. Удобный интерфейс	
Bitbonkers	Имеет уникальную интерактивную функциональность. Показывает новые блоки и транзакции в режиме реального времени	Обладает ограниченными функциональными возможностями
3D Chain	Интерактивная визуализация данных в трехмерном пространстве	Ограниченный функционал в сравнении с другими инструментами

Заключение

Рассмотренные в данной работе инструменты обладают рядом уникальных возможностей и могут быть использованы для эффективного решения конкретных задач, связанных с анализом транзакций Bitcoin. Некоторые инструменты позволяют производить анализ с крупным объемом данных, что может быть использовано для проведения исследований в области идентификации участников сети Bitcoin.

В сравнении с другими исследованиями по данной тематике результаты данного исследования подтверждают, что анализ транзакций блокчейн представляет собой комплексный процесс, который включает в себя работу с большими данными и интеллектуальными методами обработки данных.

Существующие исследования по данной тематике уделяют мало внимания инструментам анализа транзакций Bitcoin с открытым исходным кодом. В некоторых исследованиях осуществляется сравнение только лишь конкретной категории инструментов анализа с точки зрения удобства использования без детального рассмотрения их технических возможностей и показателей работы.

Рассмотренные инструменты анализа транзакций и полученные результаты позволят исследователям в данной области выбрать оптимальный стек технологий для разработки передовых инструментов и методов идентификации средств, связанных с незаконной деятельностью, и их источников.

Дальнейшие исследования в этой области могут быть связаны с проведением экспериментов для получения оценки и показателей эффективности работы рассмотренных в данной работе инструментов. В свою очередь, существует необходимость разработки единого инструмента для проведения исследований, связанных с анализом транзакций сети Bitcoin, начиная с этапа сбора информации и заканчивая этапом визуализации результатов, полученных в ходе анализа.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Decentralized approach for collecting and processing data of the enterprise information infrastructure / E. A. Basinya, G. T. Merzadinova, A. B. Zakirova, Z. B. Akhayeva [et. al.]. - Text : direct // Journal of Theoretical and Applied Information Technology. - 2022. Vol. 100, iss. 3. - P. 788-798.

2. Басыня Е. А., Сафронов А. В. Децентрализованный подход к сбору и обработке данных информационной инфраструктуры предприятия / // Вестник УрФО. Безопасность в информационной сфере. - 2019. - № 3 (33). - С. 43–54. - DOI: 10.14529/secur190305.

3. Сердечный А. Л., Скогорева Д. А., Длинный Е. П., Ле Т. Ч., Чьеу Д. В. Картографическое исследование blockchain-транзакций и смарт-контрактов киберпреступников, атакующих автоматизированные информационные системы, и оценка ущербов от реализации их атак //Информация и безопасность Учредители: Воронежский государственный технический университет. – 2021. – Т. 24. – №. 4. – С. 471-500.

4. Родивилина В. А., Родивилин И. П., Коломинов В. В. Проблемы противодействия использованию анонимности в сети интернет в преступных целях //Криминалистика: вчера, сегодня, завтра. – 2021. – №. 4. – С. 68-79.

5. Сонг В. и др. Анализ данных блокчейна с точки зрения сложных сетей: обзор // Наука и технологии Цинхуа. – 2022. – Т. 28. – №. 1. – С. 176-206.

6. Мун, Х., Ким, С., Ли, Ю. Метод анализа биткойнов на основе RDBMS // Информационная безопасность и криптология - ICISC. 2020. Том 12593. С 235–253.

7. Бакуменко Л.П., Васильева Н.С. Классификация методом опорных векторов мошеннических программ кражи биткоина // Учет и статистика. – Учредители: Ростовский государственный экономический университет "РИНХ". – 2022 – Т. 68 – № 4. – С. 112-122

8. 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.

9. RPC API Reference [Электронный ресурс] // Bitcoin Developer. URL: https://developer.bitcoin.org/reference/rpc/ (дата обращения: 20.04.2023).

10. fast-dat-parser [Электронный ресурс] // Github. URL: https://github.com/bitcoinjs/fast-dat-parser(дата обращения: 20.04.2023).

11. Люкс З.А. и соавт. Полнотекстовый поиск проверяемых метаданных учетных данных в распределенных реестрах // 2019 Шестая международная конференция по Интернету вещей: системы, управление и безопасность (IOTSMS). – IEEE, 2019. – С. 519-528.

12. Мюле А., Грюнер А., Мейнель К. Характеристика использования прокси в одноранговой сети Биткойн //Материалы 22-й Международной конференции по распределенным вычислениям и сетям. – 2021. – С. 176-185.

13. Blockhair API [Электронный ресурс] // Github. URL: https://github.com/Blockchair/Blockchair.Support/blob/master/API.md (дата обращения: 20.04.2023).

14. Тованич Н. и соавт. Систематический обзор онлайн-визуализации биткойнов // Плакаты Европейской конференции по визуализации (EuroVis). – 2019.

15. pycoin [Электронный ресурс] // PyPi. URL: https://pypi.org/project/pycoin/(дата обращения: 20.04.2023).

16. blockstream [Электронный ресурс] // PyPi. URL: https://pypi.org/project/blockstream/(дата обращения: 20.04.2023).

17. BitBonkers [Электронный ресурс] // Privacy Pros. URL: https://privacypros.io/tools/bitbonkers/ (дата обращения: 20.04.2023).

18. 3D Chain [Электронный ресурс] // Blockchain 3d Explorer. URL: https://blockchain3d.info/(дата обращения: 20.04.2023).