DOI 10.31891/2307-5732-2024-333-2-51
УДК 004.65:519.17

ЖИКІН ЮРІЙ
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"
https://orcid.org/0009-0001-5930-1444
e-mail: yzhykin@protonmail.com
ОНАЙ МИКОЛА
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"
https://orcid.org/0000-0002-4938-8355
e-mail: onay@pzks.fpm.kpi.ua

ПАТЕРНОВИЙ АНАЛІЗ ГРАФА БІТКОЇН-ТРАНЗАКЦІЙ

Біткоїн є найбільш економічно успішною системою електронних платежів, в основі якої лежить прокотол децентралізованого консенсусу, який дозволяє підтримувати однакову копію бази даних впорядкованих транзакцій у кожного учасника системи. Для забезпечення роботи протоколу база даних з транзакціями є повністю прозорою і містить детальний опис кожної транзакції, включно з інформацією про місце знаходження коштів транзакції до та після того, як вона відбулась. Ця інформація дозволяє побудувати граф Біткоїн-транзакцій і за допомогою різних методів аналізу встановлювати зв'язки між місцями перебування коштів з метою деанонімізації користувачів. Методи аналізу графа Біткоїн-транзакцій є важливим напрямком досліджень як з точки зору отримання знань про злочинну фінансову діяльність, так і з точки зору захисту користувачів від незаконного фінансового шпигунства.

У даному дослідженні пропонується метод аналізу графа Біткоїн-транзакцій, що полягає у пошуку патернів, що відповідають типовим складним багатокроковим операціям. Для цього пропонується модель графа транзакцій, що містить як позиції, в яких грошові одиниці знаходяться в певний момент часу (адреси), так і переходи, що переводять грошові одиниці з одних позицій на інші (власне, транзакції). Включення транзакцій в модель графа є необхідним, оскільки в патернах, що розглядаються, ключовою інформацією є кількість вхідних та вихідних позицій в транзакції. Також у дослідженні описується процес початкової побудови і послідовної добудови та розширення такого графа Біткоїн-транзакцій, що включно з інформацією про асоціації між позиціями та сторонніми даними про користувачів утворює граф знань про Біткоїн-транзакції.

Описаний підхід може бути використаний на практиці для побудови глобального графа знань про Біткоїн-транзакції від початку існування мережі. Для цього можна зокрема використовувати графові бази даних, що дозволяють зберігати великі кількості вершин та зв'язків між ними, а також здійснювати ефективний пошук патернів.

Ключові слова: біткоїн, криптовалюта, блокчейн, граф транзакцій, граф знань.

ZHYKIN YURI, ONAI MYKOLA
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

PATTERN-BASED BITCOIN TRANSACTION GRAPH ANALYSIS

Bitcoin is the most economically successful electronic payment system, which is based on a decentralized consensus protocol that allows each participant of the system to maintain an identical copy of the database of ordered transactions. To ensure the protocol's operation, the database of transactions is entirely transparent and contains detailed descriptions of each transaction, including information about the location of funds before and after the transaction happened. This information can be used to construct a graph of Bitcoin transactions and use various methods of analysis to establish connections between the locations of funds for the purpose of de-anonymizing users. Methods of Bitcoin transaction graph analysis are an important direction of research both for gaining knowledge about criminal financial activity and for protecting users from illegal financial espionage.

This paper proposes a method of analyzing the Bitcoin transaction graph, which involves searching for patterns that correspond to typical complex multi-step operations. For this purpose, a model of a transaction graph is proposed, which includes both positions where monetary units are located at a certain point in time (addresses) and transitions that transfer monetary units from one set of positions to another (transactions themselves). Including transactions in the graph model is necessary because the key information in the patterns considered in this paper is the number of input and output positions in the transaction. The paper also describes the process of initially building and consequently expanding and extending such a Bitcoin transaction graph, which in combination with information about associations between positions and external user data forms a knowledge graph of Bitcoin transactions.

The described approach can be used in practice to build a global knowledge graph of Bitcoin transactions from the start of the network operation. Graph databases can be used for this purpose since they allow storing large numbers of nodes and connections between them, as well as performing efficient pattern searches.

Keywords: Bitcoin, cryptocurrency, blockchain, transaction graph, knowledge graph.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Біткоїн – це грошова система, що була описана особою під псевдонімом Сатоші Накамото [1] у 2008 році і введена в експлуатацію спонтанно сформованою групою криптографів на початку 2009 року. Ключовою особливістю системи Біткоїн, яка зацікавила криптографічну спільноту, стало те, що це була перша повністю децентралізована транзакційна система, яка могла гарантувати достовірність поточного стану володіння грошовими одиницями без будь-якого центрального контролюючого органу. Протокол створення і підтвердження Біткоїн-транзакцій побудовано так, що обчислювальна вартість перевірки правильності стану володіння грошовими одиницями є дуже низькою і може бути виконана навіть на найбільш простих обчислювальних пристроях, тоді як обчислювальна вартість створення підробленого стану володіння

грошовими одиницями є надзвичайно високою як з точки зору наявності спеціальних обчислювальних пристроїв, так і з точки зору спожитої енергії.

Протягом останніх років Біткоїн все більше інтегрується в світові економічну і технологічну екосистеми. Так, перша відома комерційна Біткоїн-транзакція, що відбулась в травні 2010 року [2], встановила приблизний курс обміну $0.003 за один біткоїн, тоді як вже восени 2017-го року він сягнув історичного максимуму $65000.00 за один біткоїн, при капіталізації ринку 1.1 трильйона доларів США. Станом на 2024 рік Біткоїн є легальною грошовою одиницею в Сальвадорі та Центральноафриканський Республіці. Біткоїн також є важливим інструментом для здійснення економічної діяльності в умовах тоталітарних режимів чи гіперінфляції. На даний момент вже існують сотні компаній, що розробляють сервіси та нові технології на основі Біткоїн-протоколу. Багато з цих компаній також працюють над вдосконаленнями самого Біткоїн-протоколу. Розробляються технології мікроплатежів [3] та «розумних» контрактів [4]. Усе це дає підстави вважати, що в майбутньому роль Біткоїна як фінансового інструменту і навіть глобальної грошової системи ставатиме все більш важливою.

Втім, децентралізованість Біткоїна досягається тим, що у кожного учасника Біткоїн-мережі є повна копія всієї історії транзакцій від моменту початку роботи мережі, через що ця система має деякі недоліки порівняно з традиційними централізованими транзакційними системами. Одним з найбільш важливих таких недоліків є те, що інформація про стан володіння грошовими одиницями та факти передачі володіння від одного користувача системи до іншого (транзакції) перебувають у вільному доступі як частина історії транзакцій, і можуть агрегуватись та комбінуватись з сторонніми даними з метою деанонімізації користувачів. Ця проблема порушує взаємозамінність грошових одиниць та приватність фінансових потоків індивіда чи підприємства, що може становити загрозу фінансовій чи навіть фізичній безпеці користувача Біткоїна.

Фундаментальною складовою протоколу Біткоїн є база даних з історією транзакцій [1]. Ця база даних містить упорядковані записи про всі транзакції, що відбулись в мережі Біткоїн від початку її роботи в січні 2009 року. База даних транзакцій розділена на сторінки, кожна з яких може містити не більше 1 Мб даних (2-4 тисячі транзакцій у поточній версії протоколу) [1]. Ці сторінки називаються блоками і об'єднуються у ланцюг у такий спосіб, що зміна хоча б одного біта у блоці потребує повного переобчислення всіх наступних блоків у ланцюгу. Обчислення одного блока відбувається шляхом перебору значень криптографічної хеш-функції і складність цього перебору динамічно змінюється так, щоб в кожен момент часу тривалість підбору блока всіма учасниками мережі становила приблизно 600 секунд [1]. Результат такого обчислення слугує цифровим підписом для блока і називається доказом виконаної роботи, тому що вказує на середню кількість енергії, яка була витрачена учасником для знаходження даного результату.

Таким чином Біткоїн-протокол є протоколом децентралізованого консенсусу щодо порядку всіх транзакцій: кожен учасник мережі може бути впевненим, що його власна копія ланцюга блоків повністю збігається з копіями всіх інших учасників, і будь-які спроби змінити історію транзакцій потребують колосальних витрат енергії за умови, що достатньо багато учасників беруть участь у обчисленні нових блоків.

Кожна сторінка (блок) бази даних містить упорядковану послідовність транзакцій, що були певним чином вибрані з множини непідтверджених транзакцій під час створення блока. Коли транзакція потрапляє в черговий блок, вона вважається підтвердженою.

Для вирішення задачі захисту від повторного використання Біткоїн-протокол використовує особливу структуру транзакції. На рис. 1 зображено три транзакції T1, T2 та T3. Весь біткоїн у обігу міститься у структурах, що називаються транзакційними виходами (O1, O2, O3, O4, O5 та O6 на рис. 1), кожен з яких складається з двох значень: кількості неподільних одиниць біткоїна (A1, A2, A3, A4, A5 та A6 на рис. 1) та ідентифікатора власника, або адреси (P1, P2, P3, P4 та P5 на рис. 1). Кожна транзакція знищує один або більше таких транзакційних виходів, і створює один або більше нових. На рис. 1 транзакція T3 знищує виходи O1 та O4, і створює нові виходи O5 та O6. Транзакційний вихід, який знищується, представляється в транзакції структурою-посиланням, яка називається транзакційним входом даної транзакції (I1, I2, I3, I4 та I5 на рис. 1). Транзакційний вихід є невикористаним, якщо не існує транзакції, що його знищує. Наявність у користувача певної кількості біткоїна означає, що у множині всіх невикористаних транзакційних виходів в системі є кілька виходів, які він може використати (має доступ до відповідних криптографічних ключів тощо). Оскільки кожна транзакція посилається на виходи інших транзакцій (зв'язки між I4 і O1 та I5 і O4 на рис. 1), які вона знищує, можна прослідкувати грошовий потік між транзакціями аж до моменту, коли певна кількість біткоїна вперше з'явилась у мережі як винагорода учаснику, що витратив енергію на створення чергового блока. Ці зв'язки між транзакціями та їх виходами утворюють граф грошових потоків в системі Біткоїн і оскільки кожен учасник системи змушений підтримувати повну копію всієї історії транзакцій для регулярної перевірки консенсусу, він має доступ до всього графу грошових потоків в системі від початку її існування.

Найпростішим прикладом використання інформації з графа Біткоїн-транзакцій, що може мати негативні наслідки, є встановлення кількості біткоїна, що належить тій чи іншій людині. На ранніх етапах розвитку Біткоїн-екосистеми поширеною практикою було використання єдиного криптографічного ключа як ідентифікатора власника біткоїна: транзакційні виходи прив'язувались до публічного криптографічного ключа, і для їх використання потрібного було включити в транзакцію криптографічні підписи, здійснені відповідним приватним криптографічним ключем. Оскільки бінарне представлення ключа безпосередньо включене в транзакційний вихід (P1, P3, тощо, на рис. 1), для обчислення повної кількості біткоїна у власності у певної людини достатньо лише знайти всі невикористані транзакційні виходи, що містять її

криптографічний ключ. На рис. 1 виділені два транзакційні виходи O3 та O6, що містять один і той же криптографічний ключ P3. Це однозначно вказує, що обидва виходи належать одному й тому ж користувачу. Пізніше використання публічних ключів поступово було витіснене використанням хешів криптографічних ключів, які почали називати Біткоїн-адресами, і оскільки ця зміна сама по собі не вирішувала цієї проблеми, з'явилось поняття транзакційної гігієни, одним з ключових принципів якої є уникання повторного використання однієї і тієї ж Біткоїн-адреси [5].

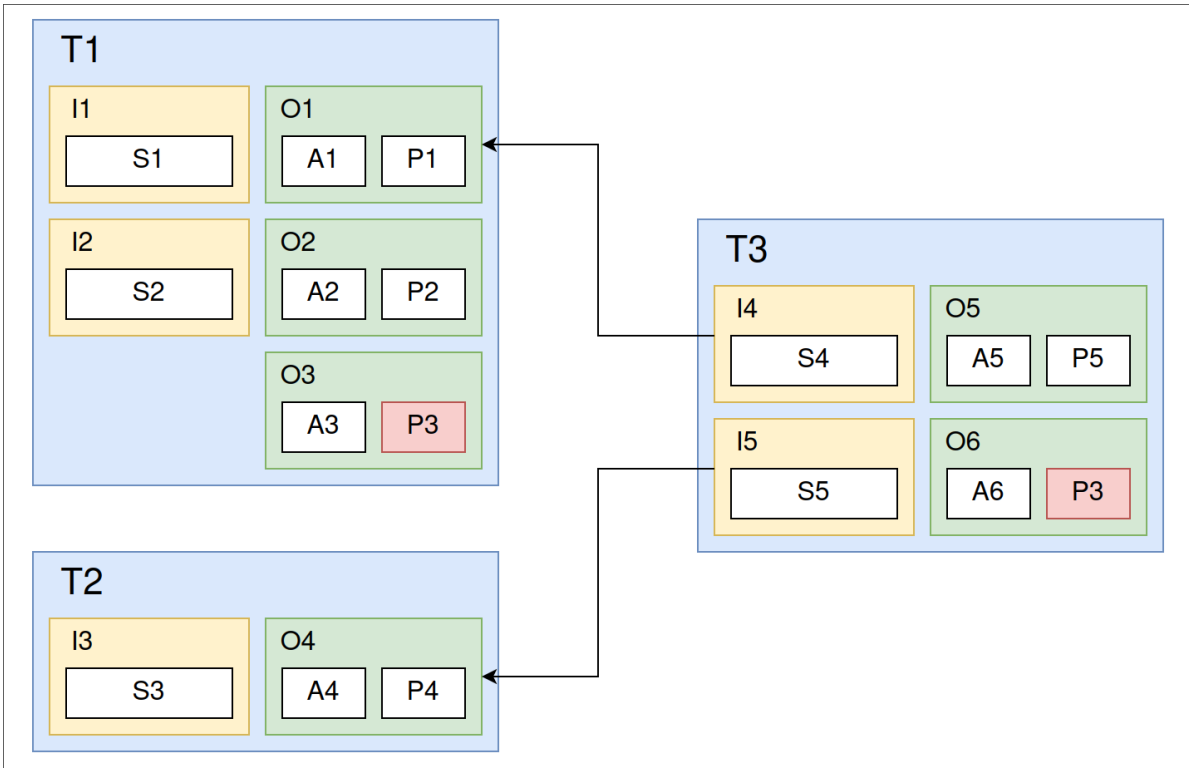

Рис. 1. Структура та зв'язки між Біткоїн-транзакціями

За умови, що більшість користувачів використовують нові Біткоїн-адреси для кожного нового транзакційного виходу, задача обчислення кількості біткоїна у конкретного власника ускладнюється, але час від часу користувач може здійснювати консолідацію транзакційних виходів (об'єднання менших виходів у більші з метою зменшення комісій за транзакції в майбутньому, які залежать від розміру транзакції а не від кількості біткоїна у ній). У графі Біткоїн-транзакцій такі операції виглядають як транзакції, що знищують багато виходів, але при цьому створюють значно меншу кількість нових (зазвичай 1). При цьому достатньо, щоб хоча б одна Біткоїн-адреса у виході, що знищується, була прив'язаною до будь-яких даних, що ідентифікують користувача (того ж публічного ключа чи випадкової публікації на веб-ресурсі, щоб можна було зробити припущення, що всі інші виходи, що були знищені цією транзакцією, належать тому ж користувачу.

Проблему публічності транзакційних даних системи Біткоїн складно вирішити на рівні протоколу, оскільки відкритий доступ до транзакційних даних є фундаментальним принципом роботи цього протоколу. Тому дослідження, присвячені вивченню методів аналізу графу Біткоїн-транзакцій та зовнішніх методів захисту транзакційних даних від такого аналізу є актуальними.

Аналіз досліджень та публікацій

Простий метод аналізу графу транзакцій полягає у побудові графу зв'язків між транзакційними виходами для певного проміжку часу та подальшого анотування його інформацією з сторонніх публічних джерел (форумів, соціальних мереж, інших веб-ресурсів) [6, 7]. Цей підхід використовує в якості основної евристики припущення, зазначене вище, а саме: транзакції, що мають декілька входів, з великою ймовірністю знищують виходи, що належать одному і тому ж користувачу. Однак ця евристика є недостатньою в умовах поточного стану Біткоїн-екосистеми, зокрема через те, що протягом останніх років значної популярності набув протокол розподіленого змішування виходів CoinJoin [8], в основі якого лежить агрегація виходів, що належать різним користувачам, в одну транзакцію, що створює певну кількість нових виходів однакового розміру. Тому для вдосконалення методів аналізу графа транзакцій необхідні більш складні евристики, що описують типові операції, а також інструменти, що можуть ефективно розпізнавати відповідні патерни графа Біткоїн-транзакцій. При цьому граф зв'язків між транзакційними виходами в загальному випадку є недостатнім для розпізнавання типових операцій, оскільки не відображає самих транзакцій, кількості входів та виходів у них, тощо. Відсутність інформації про кількість входів та виходів у транзакції обмежує типи евристик, що можуть бути застосовані до графа для ідентифікації типових операцій. Цю проблему можна

вирішити шляхом включення в граф транзакцій інформації і про транзакційні виходи, і про самі транзакції.

Ще однією проблемою методу аналізу анотованого графа транзакцій, запропонованого у [7] є те, що деякі складні операції в Біткоїн-мережі можуть відбуватись протягом довгих проміжків часу, і тому застосування аналізу може бути безрезультатним, якщо ключові компоненти операції не попали в проміжок часу, для якого було побудовано граф транзакцій. Цю проблему можна вирішити побудовою і послідовною добудовою графа транзакцій, що відображає всю історію транзакцій у Біткоїн-системі, але розмір такого графа значно ускладнить застосування аналізів.

Існують комерційні сервіси, що надають послуги з аналізу графа Біткоїн-транзакцій, але методи аналізу, що використовуються цими сервісами, є корпоративною таємницею, і тому ми не можемо включити їх у наше дослідження.

Формулювання цілей статті

Метою роботи є: розробка методу аналізу графа знань про Біткоїн-транзакції шляхом розпізнавання у зв'язках між транзакціями та транзакційними входами і виходами патернів, що відповідають типовим потенційно багатокроковим грошовим операціям в системі Біткоїн, з метою визначення асоціацій між транзакційними виходами в системі Біткоїн та даними про користувачів, що отримані з зовнішніх джерел.

Для досягнення даної мети у даному дослідженні ставляться наступні задачі:

- розробити модель графа Біткоїн-транзакцій, що відображає транзакційні виходи, що належать певному користувачу, транзакції — події, що змінюють множину невикористаних транзакційних виходів та стан володіння ними, та асоціації між транзакційними виходами і інформацією про користувачів, що отримана з зовнішніх ресурсів;
- сформулювати процес побудови і розширення графа транзакцій з метою отримання графа знань про Біткоїн-транзакції;
- розглянути типові операції у Біткоїн-системі як патерни у графі знань про Біткоїн-транзакції та їх використання для пропагації асоціацій між транзакційними виходами та ідентифікуючими даними про користувачів системи.

У даній роботі пропонується метод аналізу графа Біткоїн-транзакцій, що полягає у використанні інструментів пошуку патернів у графі для ідентифікації типових операцій в системі Біткоїн і встановлення зв'язків між транзакційними виходами та користувачами. Для цього формулюється модель графа Біткоїн-транзакцій, що містить усю необхідну інформацію для розпізнавання таких патернів незалежно від того, як довго триває подібна операція.

В роботі було використано теоретичні методи моделювання графів, що складаються з різних типів вершин та зв'язків, зокрема графів знань.

Також було розроблено програмне забезпечення [9] для побудови об'єктної моделі Біткоїн-блоків з подальшою трансформацією їх у графове представлення RDF (Resouce Description Framework) [10].

Візуальний аналіз типових графових патернів здійснювався з використанням популярного програмного забезпечення для візуального дослідження графів AllegroGraph та Gruff [11].

Виклад основного матеріалу

Для представлення анотованого графа Біткоїн-транзакцій запропонована наступна модель:

$$G_T = (T \cup P \cup E, I \cup O \cup L)$$

$$I \subseteq T \times P$$

$$O \subseteq P \times T$$

$$L \subseteq P \times E$$

де T — множина Біткоїн-транзакцій, P — множина транзакційних виходів, E — множина сутностей, що представляють абстрактних користувачів системи Біткоїн, I — множина зв'язків між використаними транзакційними виходами транзакцій, які їх використали, O — множина зв'язків між транзакціями та їхніми виходами, L — множина зв'язків між виходами та сутностями, яким ці виходи належать.

На рис. 2 зображено підграф графа Біткоїн-транзакцій, ще моделює типову Біткоїн-транзакцію Т3 з двома входами та двома виходами, один з яких цільовий, а інший — повертає решту з транзакції користувачу. При цьому якщо в графі присутня інформація про користувача E1, якому належать виходи O1 та O2, то можна зробити припущення, що один з виходів O4 та O5 є виходом-рештою, і тому теж належить користувачу E1.

Дана модель відрізняється від моделей, розглянутих у [6-7] тим, що граф транзакцій окрім множини вершин-виходів містить ще й множину вершин-транзакцій. Це необхідно для того, щоб відобразити в моделі транзакції як події, що мають конкретну кількість входів та виходів, а також час та джерело. В моделі, що має лише вершини-виходи, транзакція — це сукупність зв'язків між виходами, що значно ускладнює опис патернів, які шукають певні типи транзакцій за кількістю входів та виходів, тощо.

В початковому стані множина зв'язків L між вершинами-виходами P та вершинами-сутностями E є порожньою, і встановлення цих зв'язків є результатом процесу анотації графа транзакцій, описаного вище. В деяких системах представлення графів, зокрема у RDF [10], дозволяється додавання довільних атрибутів (пар з ключа та значення) до будь-яких елементів графа, що може бути дуже корисним у ситуаціях, коли зв'язок між вершиною-виходом та вершиною-сутністю існує з ймовірністю, відмінною від 0 та 1 (ключем атрибута у такому випадку буде символ «ймовірність», а значенням — відповідна величина). Наприклад, для транзакції з рис. 2 можна позначити ймовірність зв'язку (E1, O5) 0.5.

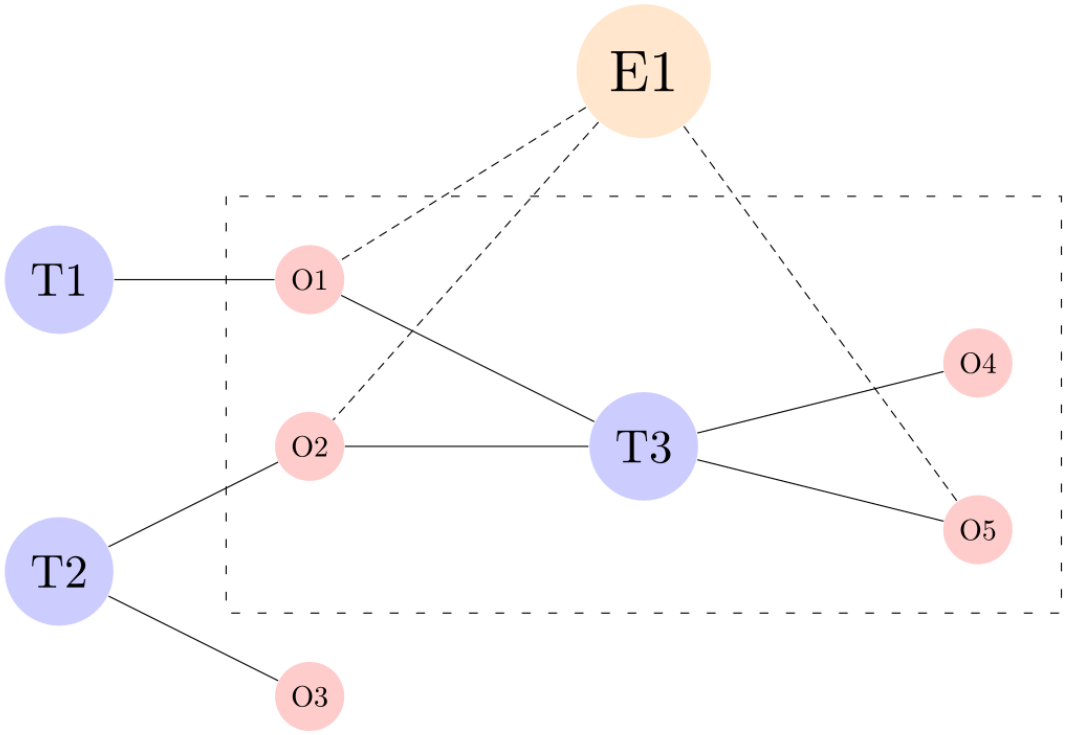

Рис. 2. Модель типової Біткоїн-транзакції

Процес побудови початкового графа Біткоїн-транзакцій є відносно простим. Оскільки кожна вершина однорангової Біткоїн-мережі — це сервер, що підтримує власну копію всієї бази даних з транзакціями, достатньо розгорнути такий сервер та реалізувати програмне забезпечення, що може отримувати від сервера блоки з транзакціями у hex-кодуванні, використовуючи протокол JSON RPC, декодувати їх у об'єкти об'єктної моделі і формувати з цих об'єктів вершини та зв'язки обраного представлення графа.

Потрібно також звернути увагу на те, що окрім підтверджених транзакцій, що входять в блоки, в анотований граф Біткоїн-транзакцій слід включати і непідтверджені транзакції, оскільки не всі непідтверджені транзакції стають підтвердженими. Деякі з них можуть бути замінені іншими транзакціями з іншими входами під час підбору комісії за транзакцію, і включення таких транзакцій у граф може додати до нього інформацію про зв'язки між транзакційними виходами, яка інакше була б назавжди втрачена. У деяких випадках кількість непідтверджених транзакцій може значно перевищувати кількість підтверджених, тому необхідно передбачити механізм «видалення сміття».

Наступний етап побудови графа Біткоїн-транзакцій аналогічний процесам, описаному у [6-7]: за допомогою обходу веб-ресурсів на зразок форумів (bitcointalk.org), соціальних мереж (x.com), платформ колективної розробки програмного забезпечення (github.com) необхідно зібрати інформацію про асоціації між користувачами цих сервісів та Біткоїн-адресами, і додати до графа початкову множину вершин E та відповідні зв'язки з множини L. Цей початковий підграф слугуватиме основою для подальшого розширення множини L за допомогою пошуку патернів у графі, деякі з яких описані нижче.

Після цього починається етап добудови графа Біткоїн-транзакцій: в середньому кожних 600 секунд з'являється новий блок, транзакції необхідно включати в граф транзакцій, також у мережі постійно з'являються нові непідтверджені транзакції, які теж необхідно додавати до графу транзакцій, і окрім цього необхідно періодично повторно обходити зовнішні ресурси для пошуку нових асоціацій між їх користувачами та Біткоїн-адресами.

Результатом цього процесу є граф Біткоїн-транзакцій, анотований інформацією про асоціації (та можливі асоціації) між користувачами зовнішніх сервісів та Біткоїн-адресами. Цей граф разом з усіма атрибутами, присвоєними його елементам, утворює граф знань про всі Біткоїн-транзакції в історії існування мережі, який постійно доповнюється новою інформацією.

І на завершення, на етапі добудови графа знань, що описаний вище, множина зв'язків між транзакційними виходами та користувачами E може розширюватись автоматично з результатів аналізів на основі патернів у графі на зразок тих, що описані нижче, оскільки такі патерни можуть бути подані у вигляді пошукових запитів у графовій базі даних, тощо.

Розглянемо деякі патерни графа знань про Біткоїн-транзакції, що відображають типові операції.

Першим прикладом, зображеним на рис. 3 є операція консолідації виходів: користувач протягом деякого часу накопичує Біткоїн у вигляді окремих транзакційних виходів, і в певний момент вирішує об'єднати ці виходи у один більший вихід з метою економії на комісіях за транзакції у майбутньому. Така транзакція матиме декілька входів і лише один вихід, і якщо у графі існує асоціація між хоча б одним входом та вершиною множини E, ми можемо з великою ймовірністю стверджувати, що всі виходи у даному підграфі належать тому ж користувачу.

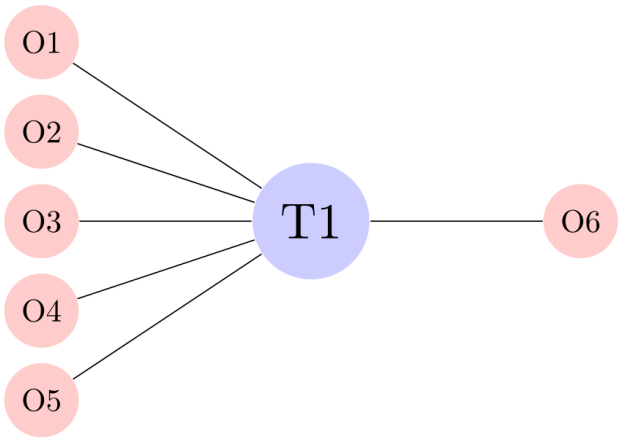

Рис. 3. Транзакція консолідації виходів

Натомість транзакція на рис. 4 виглядає як типова CoinJoin-транзакція [8], тобто колективна спонтанно створена транзакція що змішує виходи різних користувачів з метою створення перешкод аналізу грошових потоків. Така транзакція матиме декілька входів, і таку ж кількість виходів однакового розміру (в грошових одиницях), і наявна асоціація між входом даної транзакції і вершиною з множини E не може бути поширена на інші входи та виходи. З високою ймовірність якийсь з виходів транзакції належить тому ж користувачу, але ми не знаємо, який саме.

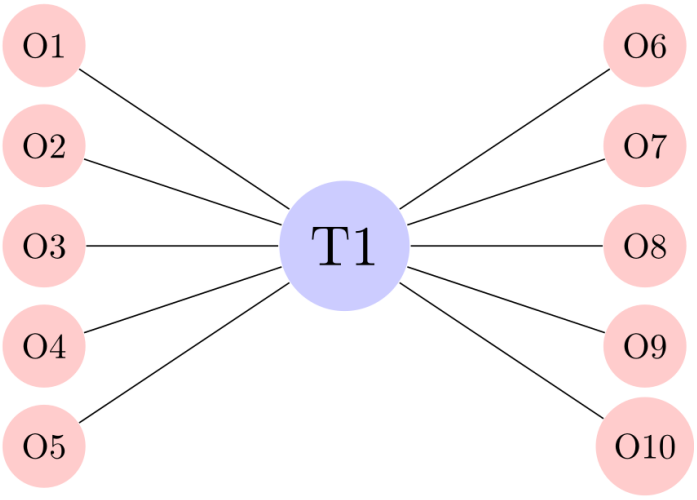

Рис. 4. CoinJoin-транзакція

Транзакція, зображена на рис. 5, є типовою примітивною спробою приховати зв'язок між користувачем E1 та користувачем E2 шляхом проведення коштів через ланцюжок тимчасових адрес. Об'єднання виходів O4 та O5 у транзакцію T4 дозволяє з великою ймовірністю ствреджувати, що вихід O2 теж належить сутності E2, і тому даний патерн відображає безпосередню транзакцію між E1 та E2, хоч це й намагались приховати.

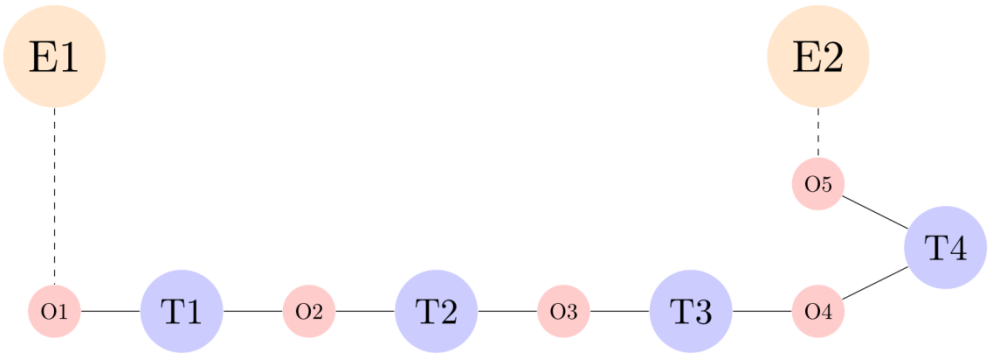

Рис. 5. Ланцюжок тимчасових адрес

Подальше вивчення графа Біткоїн-транзакцій дозволить сформулювати інші подібні евристики для виявлення поширених багатокрокових операцій у системі Біткоїн, а використання інструментів для пошуку

патернів у графа (наприклад, мови запитів SPARQL [12], що може працювати з графовим представленням RDF [10]) дозволить ефективно знаходити подібні операції та розширювати граф знань про Біткоїн-транзакції новими асоціаціями.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Розроблена модель графа знань про Біткоїн-транзакції відображає як транзакційні виходи, що складають грошову масу в обігу в Біткоїн-мережі, так і самі транзакції, з якими вони пов'язані, що дозволяє описувати складні багатокрокові операції як патерни у графі, що складаються з однієї чи більше транзакцій, що мають певні кількості входів та виходів чи деякі конкретні значення атрибутів.

Описаний процес побудови і розширення графа Біткоїн-транзакцій дозволяє отримати актуальний граф знань про Біткоїн-транзакції, включно з інформацією про асоціації між транзакційними виходами та ідентифікаторами користувачів з зовнішніх ресурсів, яка може постійно доповнюватись як в ручному режимі, так і з результатів автоматизованих аналізів.

Деякі типові операції в системі Біткоїн можуть бути описані як патерни у графі зв'язків транзакцій та транзакційних виходів, тому пошук таких патернів у графі з урахуванням початкової інформації про асоціації між транзакційними виходами та користувачами дозволяє встановлювати нові такі зв'язки навіть у випадках, коли операція складається з кількох транзакцій. Подальше вивчення типових операцій дозволить сформулювати нові евристики для автоматизації цього процесу.

Граф знань про Біткоїн-транзакції, розглянутий вище, відображає зв'яки між транзакційними виходами та транзакціями, зокрема інформацію про кількість входів та виходів у кожній транзакції. Така структура графа дозволяє застосовувати евристики, що описують типові складні операції в системі Біткоїн як патерни у графі транзакцій, приклади яких було розглянуто вище. Зокрема, наявність у графі інформації про кількість виходів транзакції дозволяє легко відрізнити консолідаційну транзакцію, що вказує на єдиного власника всіх її входів, та CoinJoin-транзакцію [8], яка майже гарантовано вказує на різних власників. Процес побудови та підтримки графа знань про всі транзакції, що відбулись з моменту початку роботи мережі, що описаний вище, дозволяє гарантувати наявність всіх зв'язків між транзакціями та транзакційними виходами не залежно від того, протягом якого проміжку часу відбувалась певна багатокрокова операція, що розглядається.

Найбільш важливим напрямком подальших досліджень є вивчення графу Біткоїн-транзакцій з метою формулювання нових евристик, що описують типові складні операції в системі Біткоїн.

Також подальші дослідження можуть полягати у порівнянні способів представлення графа транзакцій, зокрема у вигляді об'єктної моделі графа, реляційної бази даних, чи графової бази даних з використанням технології RDF [10], а також дослідження способів пошуку патернів у такому графі з використанням безпосередніх обходів об'єктної моделі графа чи використання існуючих мов для пошуку патернів у графах, таких як SPARQL [12]. Коли мова йде про повний граф знань про Біткоїн-транзакції, який станом на 2024 рік містить майже мільярд транзакцій, важливим аспектом подальших досліджень є порівняння швидкодії алгоритмів пошуку патернів у графі такого розміру на основі безпосередніх обходів об'єктної моделі графа та графових мов запитів, зокрема SPARQL [12].

References

1. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf.
2. Pizza for bitcoins? https://bitcointalk.org/index.php?topic=137.0.
3. Poon, J., Dryja, T. (2016). The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments. https://lightning.network/lightning-network-paper.pdf.
4. RGB: Turing-complete, Scalable & Confidential Smart Contract Layer for Bitcoin & LN. https://blackpaper.rgb.tech/.
5. Privacy. https://en.bitcoin.it/wiki/Privacy.
6. Reid, F., Harrigan, M. (2012). An Analysis of Anonymity in the Bitcoin System. https://arxiv.org/pdf/1107.4524.pdf.
7. Fleder, M., Kester, M. S., Pillai, S. (2015). Bitcoin Transaction Graph Analysis. https://arxiv.org/abs/1502.01657.
8. Rainer Stütz, Johann Stockinger, Bernhard Haslhofer, Pedro Moreno-Sanchez, Matteo Maffei. (2022). Adoption and Actual Privacy of Decentralized CoinJoin Implementations in Bitcoin. https://arxiv.org/pdf/2109.10229.pdf.
9. Resouce Description Framework (RDF). https://www.w3.org/RDF/.
10. BRDF: Bitcoin chain data represented as RDF. https://github.com/rodentrabies/brdf.
11. AllegroGraph: Knowledge Graph + LLM Solutions. https://allegrograph.com/.
12. SPARQL 1.1 Query Language. https://www.w3.org/TR/sparql11-query/.