Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів аналізу транзакцій криптовалют для перевірки їх законності

(тема)

Виконав:
Випускник 2 курсу, групи ІПЗм-19-2
Войтенко О.О.
(прізвище,ініціали)

Спеціальність 121- Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми Освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Керівник доц. Каук В.І.
(посада,прізвище)

Допускається до захисту
Зав. кафедри _____ З.В. Дудар
(підпис) (прізвище, ініціали)

2021р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Програмної інженерії
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 121- Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми Освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Інженерія програмного забезпечення
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« 26 » березня 2021 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студента Войтенка Олега Олександровича
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів аналізу транзакцій криптовалют для перевірки їх законності
затверджена наказом університету від 26.03.2021 № 385
2. Термін подання роботи до екзаменаційної комісії 09 травня 2021р.
3. Вихідні дані до роботи *електронні ресурси теми, методи кластеризації криптовалютних гаманців, датасет Chainalysis, середовище розробки WebStorm, мови програмування: Node.js, Angular, Python, Scikit.*
4. Перелік питань, що потрібно опрацювати в роботі *аналіз предметної області і постановка задачі, аналіз методів кластеризації криптовалютних гаманців, огляд методів оцінки ризику та категоризації кластерів, створення та навчання моделей нейронних мереж для прогнозування, проведення експерименту, формування рекомендацій.*
5. Перелік графічного матеріалу із зазначенням креслеників, схем, слайдів, ілюстрацій *аналіз предметної області, аналіз конкурентів, мета роботи, основні задачі, діаграма розгортання, типи ризиків, огляд бази Chainalysis, огляд алгоритмів машинного навчання, результати експерименту, висновки.*

6. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Спецчастина	доц. Каук В.І		11.05.21

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної області	25.01.21 – 11.02.21	виконано
2	Огляд існуючих методів	10.02.21 – 12.02.21	виконано
3	Постановки задачі	13.02.21 – 17.02.21	виконано
4	Дослідження існуючих способів кластеризації криптовалютних гаманців	18.02.21 – 24.02.21	виконано
5	Пошук датасетів для експерименту	16.02.21 – 21.02.21	виконано
6	Планування експерименту, огляд інструментів	24.02.21 – 07.03.21	виконано
7	Створення та тренування моделей	07.03.21 – 28.03.21	виконано
8	Програмна реалізація веб-додатку	28.03.21 – 10.04.21	виконано
9	Проведення експерименту	05.04.21 – 14.04.21	виконано
10	Оформлення статті	16.04.21 – 26.04.21	виконано
11	Підготовка пояснювальної записки	01.04.21 – 4.05.21	виконано
12	Підготовка презентації та доповіді	4.05.21 – 07.05.21	виконано
13	Нормоконтроль	10.05.21 – 18.05.21	виконано
14	Рецензування	10.05.21 – 18.05.21	виконано
15	Занесення диплома в електронний архів	19.05.21	виконано
16	Попередній захист	19.05.21	виконано
17	Допуск до захисту у зав. кафедри	20.05.21	виконано

Дата видачі завдання 25 січня 2021р.

Студент _____
(підпис)

Керівник роботи _____ доц. Каук В. І.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 88с., 19рис., 45 дж.

БЛОКЧЕЙН, ETHEREUM, BITCOIN, AML, KYC, FRAUD, NODE.JS, MONGODB, NEO4J, WEBSITE, KYT.

Об’єктом дослідження є структура блокчейн мереж та засоби пошуку і верифікації транзакцій. Предмет дослідження – існуючі методи аналізу транзакцій криптовалют.

Методи розробки базуються на мові програмування Node.js і графовій базі даних Neo4j. У роботі проаналізовані принципи роботи блокчейн мереж, методи кластеризації криптовалютних гаманців, способи оцінки ризику та категоризації кластерів, а також області використання графових баз даних. Також розглянуті можливості використання комбінації оригінальних публічних даних блокчейну, згрупованих і кластеризованих даних наявних систем, публічної інформації та використання алгоритмів машинного навчання під наглядом, для покращення швидкості роботи системи та розв'язання задачі кластеризації ще невизначених об’єктів блокчейн мереж.

У результаті роботи побудовано та розроблено веб-додаток з можливістю ввести геш криптовалютного гаманця та отримати повноцінний звіт про потенційну загрозу та «чистоту» транзакції за шкалою обраних ризиків.

Explanatory note to the qualification work: 88p., 19fig., 45sources.

BLOCKCHAIN, ETHEREUM, BITCOIN, AML, KYC, FRAUD, NODE.JS, MONGODB, NEO4J, WEBSITE, KYT.

The object of research is the structure of blockchain networks and tools for finding and verifying transactions. The subject of research - the existing methods of analysis of cryptocurrency transactions.

Development methods are based on the Node.js programming language and the Neo4j graph database. The paper analyzes the principles of blockchain networks, methods of clustering cryptocurrency wallets, methods of risk assessment and categorization of clusters, as well as the use of graph databases. Possibilities of using a combination of original public blockchain data, grouped and clustered data of existing systems, public information and the use of supervised machine learning algorithms to improve system speed and solve the problem of clustering still undefined blockchain network objects are also considered.

As a result, a web application was developed with the ability to enter the wallet’s hash and obtain a full report on the potential threat and "purity" of the transaction on a scale of selected risks.

Я, Войтенко Олег Олександрович, студент групи ІПЗм-19-2, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів аналізу транзакцій криптовалют для перевірки їх законності», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу ElAr KhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

ВСТУП

В останні роки криптовалюти стали невід’ємною частиною світової фінансової системи. З того часу, як у 2009 році була представлена перша децентралізована криптовалюта Bitcoin, загальна вартість криптовалют та різноманітність типів криптовалют різко зросли. Згідно з оцінкою, у першому кварталі 2021 року глобальна ринкова капіталізація криптовалют перевищила $1.1 трлн доларів[1].

Завдяки цьому зростанню, криптовалюти стали також однією із формою особистого багатства. Для обслуговування сектору криптовалют з’явився широкий спектр підприємств, пов’язаних з ними. Це підприємства, які беруть безпосередню участь у торгівлі та розробці криптовалют, такі як біржі, а також ті, що надають допоміжні послуги, пов'язані з ринком, або іншими учасниками опосередковано, включаючи компанії у банківському та ігровому секторах. Зростання ринків також спричинено значним інтересом інвесторів на фоні глобальної кризи та падіння долара, як платіжної одиниці.

На фоні глобального зростання ринків криптовалют, зростає також кількість шахраїв та кібертерористичних угрупувань. За останні роки, кількість атак збільшилася в декілька разів[2], атаки зазнають як великі, так і малі біржі, а також звичайні користувачі криптовалют. Один із найбільших інцидентів за останній час – порушення криптобезпеки японської біржі Coincheck у 2018 році. Наразі через порушення криптобезпеки було вкрадено приблизно 3 мільярди доларів, а через шахрайство та маніпуляції – 4,8 мільярда, тому загалом із 2011 року викрадено більш ніж 7,8 мільярда доларів в еквіваленті криптовалют.

Інша зворотна сторона криптовалют – використання їх у Darknet[3]. Згідно з австралійським дослідженням, приблизно 47% усіх операцій з Bitcoin приходять на Darknet. Термін Darknet з'явився ще у 1970-ті роки і використовувався для позначення приватних мереж, що характеризуються високим ступенем конфіденційності.

Спочатку "анонімний Інтернет" був безпечний, та поступово Darknet став засобом комунікації для незаконної діяльності та злочинності. Це частина мережі з багатомільярдними оборотами на чорних ринках. Майданчики для торгівлі наркотиками, зброєю, підробленими документами – це далеко не повний список того, що є в Darknet.

Криптовалюти ідеально підходять для незаконних операцій, забезпечуючи їм необхідний рівень анонімності. Відстежити власника криптовалютного гаманця вкрай складно. І це при тому, що блокчейн транзакції достатньо прозорі та відкриті. Можна стверджувати, що цифрові гроші з їх анонімними транзакціями можуть створювати екосистему для процвітання тіньового інтернету і незаконної торгівлі.

Спецслужби і органи держбезпеки багатьох країн почали серйозну боротьбу з нелегальною діяльністю в даркнеті. Спостерігаючи зв’язок злочинної діяльності і криптовалютного обміну, вони налаштовані критично до питання легалізації цифрових грошей. Тому зараз вкрай важлива та актуальна тема розроблення механізму регулювання операцій з криптовалютами.

Отже, метою роботи є дослідження актуальних методів аналізу транзакцій криптовалют для перевірки їх законності. У ході дослідження необхідно проаналізувати сучасні методи перевірки законності транзакцій криптовалют, та вирішити задачу класифікації кластерів криптовалютних гаманців.

Об’єкт дослідження – структура блокчейн мереж та засоби пошуку і верифікації транзакцій. Предмет дослідження – актуальні методи аналізу транзакцій криптовалют.

Результати проведеного дослідження будуть корисні з теоретичної точки зору для розуміння шляхів вирішення актуальних проблем криптобезпеки. З практичної точки зору є можливим використання системи, що була створена у ході роботи, для аналізу та перевірки законності транзакцій у криптовалютах. Результати дослідження можуть бути корисні традиційним фінансовим установам, фінтех стартапам, а також державним фінансовим регуляторам.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз засобів регулювання відмивання грошей в області криптовалют

У зв’язку з низкою прикладів нелегального використання криптовалют, їх часто зображують як ідеальний засіб платіжної системи для злочинців, які прагнуть приховати свої незаконні кошти.

Банківські та фінансові регулятори країн G20 впроваджують низку законів на тему фінансового моніторингу криптовалют та зробили спільну заяву, в якій наголошено на активізації зусиль у боротьбі з фінансуванням тероризму, і закликали інші країни долучитись.

Українське законодавство розвивається повільніше, але не стоїть осторонь світових процесів. «AML/CFT. Фінансовий моніторинг в Україні»[4] – проект, який запроваджено з метою розуміння, що таке фінансовий моніторинг, як відрізнити звичайні операції від операцій з відмивання коштів, які дії повинні виконати фінансові установи, щоб не впустити в фінансову систему «брудні» кошти, не втратити свою репутацію, уникнути застосування до них штрафних санкцій.

Головне завдання фінансового моніторингу – це протидія та запобігання відмиванню коштів, розповсюдження зброї масового знищення та фінансуванню тероризму.

Ця боротьба сьогодні – одна з головних тем світової фінансової системи.

Автоматизовані рішення для моніторингу блокчейн мереж можуть стати важливим компонентом дотримання нормативних вимог у криптопросторі, дозволяючи криптовалютним компаніям, банкам та іншим регульованим фінансовим установам зменшувати ризики та задовольняти світові вимоги у сфері AML/CFT.

1.2 Аналіз рішень автоматизованого моніторингу криптовалютних транзакцій для зменшення ризику незаконного використання коштів

У світі фіатних валют широко використовують рішення для автоматичного моніторингу транзакцій, що базуються на вивченні поведінки клієнтів. Такі методи є ефективним способом задовольнити вимоги щодо відмивання коштів та протидії фінансування тероризму.

Незважаючи на поширену теорію того, що транзакції криптовалют неможливо ідентифікувати та відслідкувати, існує можливість створити автоматичне рішення для моніторингу блокчейну, що буде мати змогу ідентифікувати потенційно небезпечні транзакції в мережі, а також перевіряти їх законність за визначеними критеріями. Слід зазначити що моніторинг транзакцій звичайних фіатних валют, та моніторинг криптовалютних транзакцій мають важливі відмінності.

На малюнку нижче (див. рис. 1.1) видно, що традиційні методи моніторингу фіатних транзакцій зосереджуються на виявленні аномалій поведінки в момент, коли клієнтські кошти вносяться або вилучаються з фінансової установи:

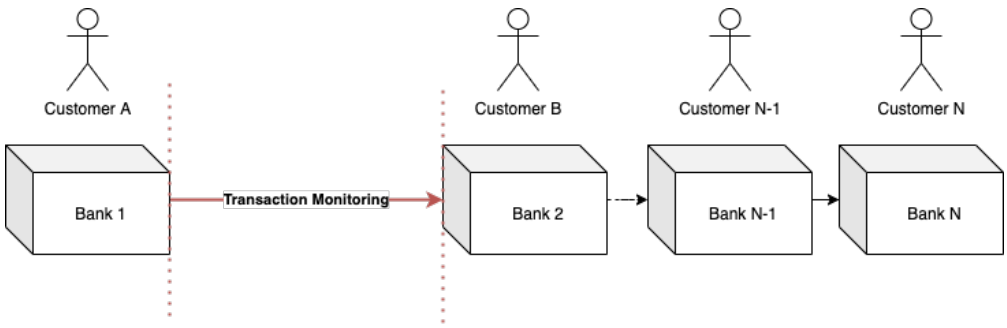

Рисунок 1.1 – Моніторинг транзакцій фіатних коштів

«Bank 1» володіє інформацією про те, що його клієнт переводить кошти іншому клієнту «Bank 2», при цьому банк не володіє інформацією про історію коштів до заходження їх до банку та після виведення коштів з цього банку.

Розглянемо більш детально ситуацію у світі криптовалют. Як видно з малюнку нижче (див. рис. 1.2), біржа криптовалют не завжди володіє інформацією про фізичну або юридичну особу, яка здійснює транзакцію.

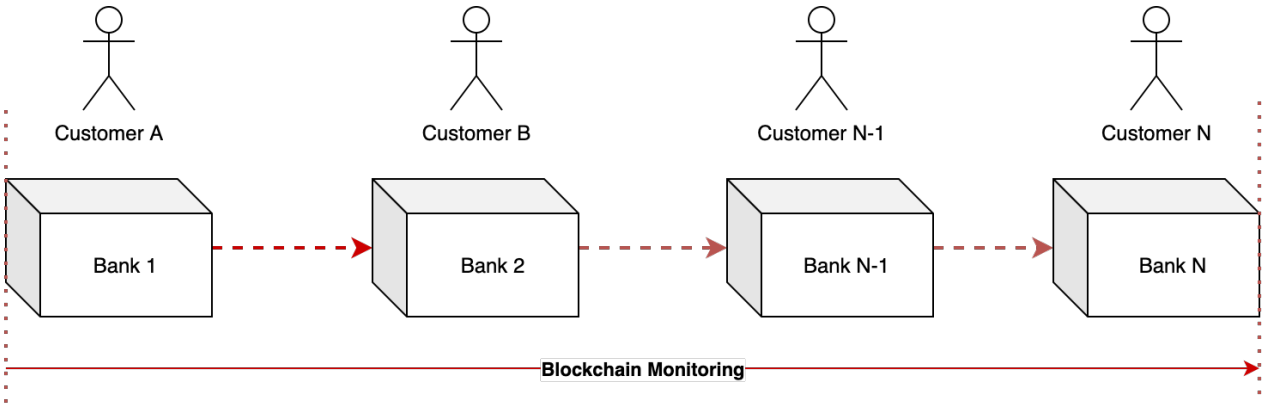

Рисунок 1.2 – Моніторинг криптовалютних транзакцій у блокчейн мережі

Проте прозорість публічних блокчейн мереж надає можливість повністю переглядати історію коштів від початкового до кінцевого джерела виникнення, розкриваючи додаткову інформацію, що може свідчити про ризик транзакції та ймовірність заходжень «брудних» коштів, які були пов'язані з відмиванням коштів, фінансуванням тероризму або іншим показникам AML/CFT регуляторів.

1.3 Аналіз законодавчих актів у ринку віртуальних активів та сфері блокчейн моніторингу

Наразі не існує єдиного способу правового регулювання криптовалют у всьому світі, кожна країна має власний підхід та досвід. У Європі криптовалюти – це децентралізовані конвертовані валюти. European Central Bank у 2014 році радив банківським організаціям не проводити транзакції з криптовалютами, до тих пір поки

для них не буде створено режим регулювання. Суд Євросоюза постановив у 2015 році, що під час використання Bitcoin як засобу платежу, операції щодо його обміну на фіатні кошти не повинні обкладатися ПДВ.

У 2017 році були внесені поправки до директиви ЄС щодо боротьби з відмиванням грошей, які спрямовані на зниження ризику використання віртуальної валюти для відмивання коштів, здобутих злочинним шляхом. Згідно з цими змінами, «платформи віртуальних валют» і провайдери криптовалютного сервісу зобов'язані дотримуватись таких самих вимог ідентифікації своїх клієнтів і відстежувати підозрілі операції, як інші фінансові організації, у тому числі банки.

У Японії, Німеччині, Франції криптовалюти визнали легальним засобом платежів.

У Великій Британії віртуальні кошти ще не легалізовані, але обмінники та криптовалютні біржі працюють та підлягають державної реєстрації, а усі операції з криптовалютами – оподаткуванню.

Продаж або купівля у Польщі криптовалют, а також майнінг визнані одним із видів комерційної діяльності. Постачальники віртуальних валют повинні здійснити процедуру держреєстрації.

Білорусь регулює операції із криптовалютами декретом «Про розвиток цифрової економіки». У рамках економічної зони «Парк високих технологій» дозволяється реєстрація криптовалютних бірж та обмінників, майнінг, ведення криптовалютного бізнесу, при цьому операції не оподатковуються.

У Російській Федерації криптовалюти визнані засобом платежів, засобом інвестицій та накопичень. Наразі заборонено оплачувати криптовалютами товари та послуги. Закон про цифрові фінансові активи[5, 6] набув чинності від 1 січня 2021 року.

За останні декілька років в Україні сформувався також ринок віртуальних активів, але він перебуває поза правовим полем країни. Експерти вважають, що в Україні щоденно проходять транзакції на суму 170-220 млн доларів на день, а

капіталізація – становить більше 2 млрд доларів. Вся галузь криптовалют України перебуває у тіні. Однак декілька місяців тому ВРУ у першому читанні ухвалила законопроєкт №3637, який повинен врегулювати поняття та правовий статус віртуального активу, а також питання прав власності та здійснення угод із криптовалютами в Україні. Якщо підтримають законопроєкт вдруге, в Україні з'явиться можливість відкривати та використовувати для проведення операцій з віртуальними активами рахунки в банках.

1.4 Загальний аналіз криптовалютних транзакцій у період COVID пандемії

У зв'язку з економічною кризою, що спричинена пандемією, розвиток низки галузей був значно повільніший ніж за останні роки. Однак ринок криптовалют продовжує залишатися лідером у розвитку, а законодавство в галузі блокчейн мереж стає пріоритетним для урегулювання державними інститутами.

Як видно з малюнку нижче (див. рис. 1.3), згідно зі статистикою аналітичної групи Crystal[7] у 2020 році, кошти Bitcoin, отримані постачальниками різних послуг збільшилися на 118,6 млрд. доларів.

Сума Bitcoin, переведена до Darknet, зросла до 1,6 млрд. доларів. Отримана сума, що пройшла через систему міксерів збільшилася загалом до 1,4 млрд. Bitcoin цього року.

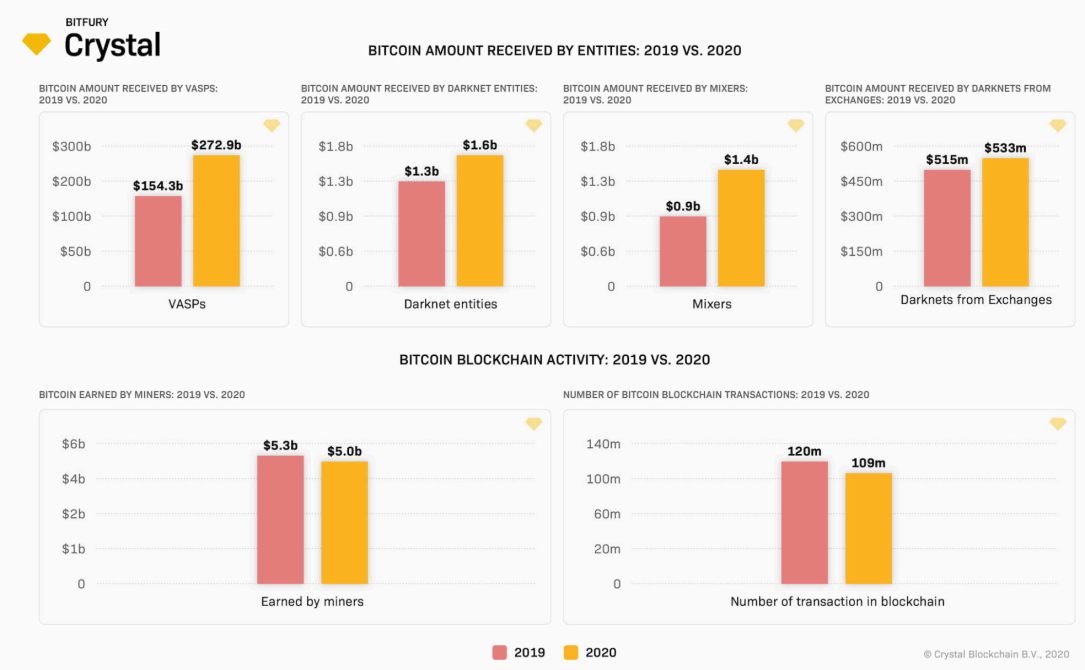

Рисунок 1.3 – Розподілення Bitcoin транзакцій за напрямками

Також, як видно з малюнку нижче (див. рис. 1.4), необхідно зазначити, що кримінальна злочинність зростає паралельно з криптовалютним ринком.

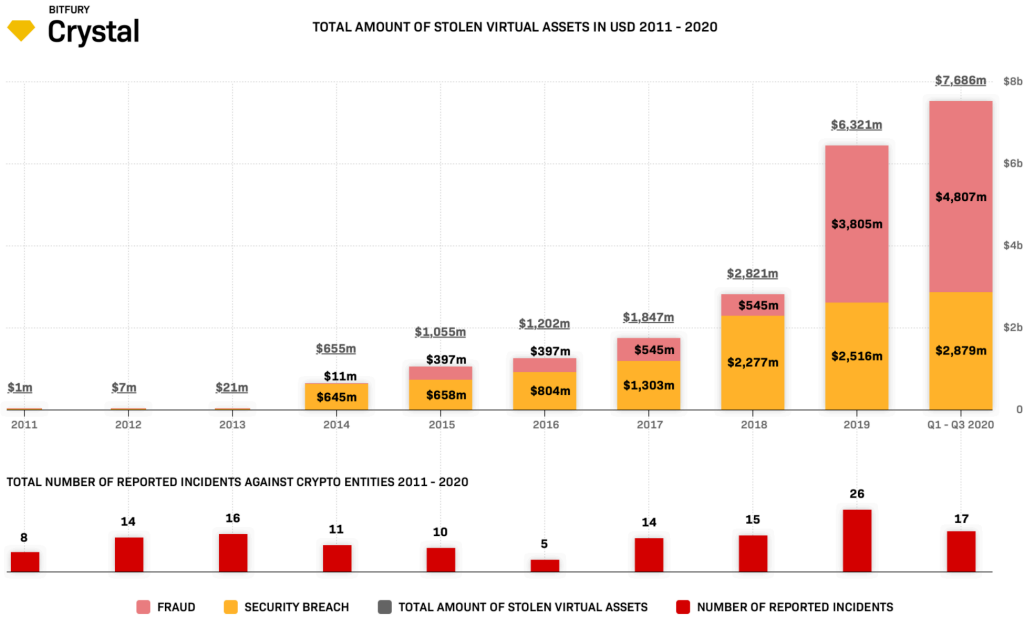

Рисунок 1.4 – Загальна кількість викрадених коштів у блокчейн мережі

На малюнку нижче (див. рис. 1.5), показана географічне розподілення транзакцій між світовими біржами:

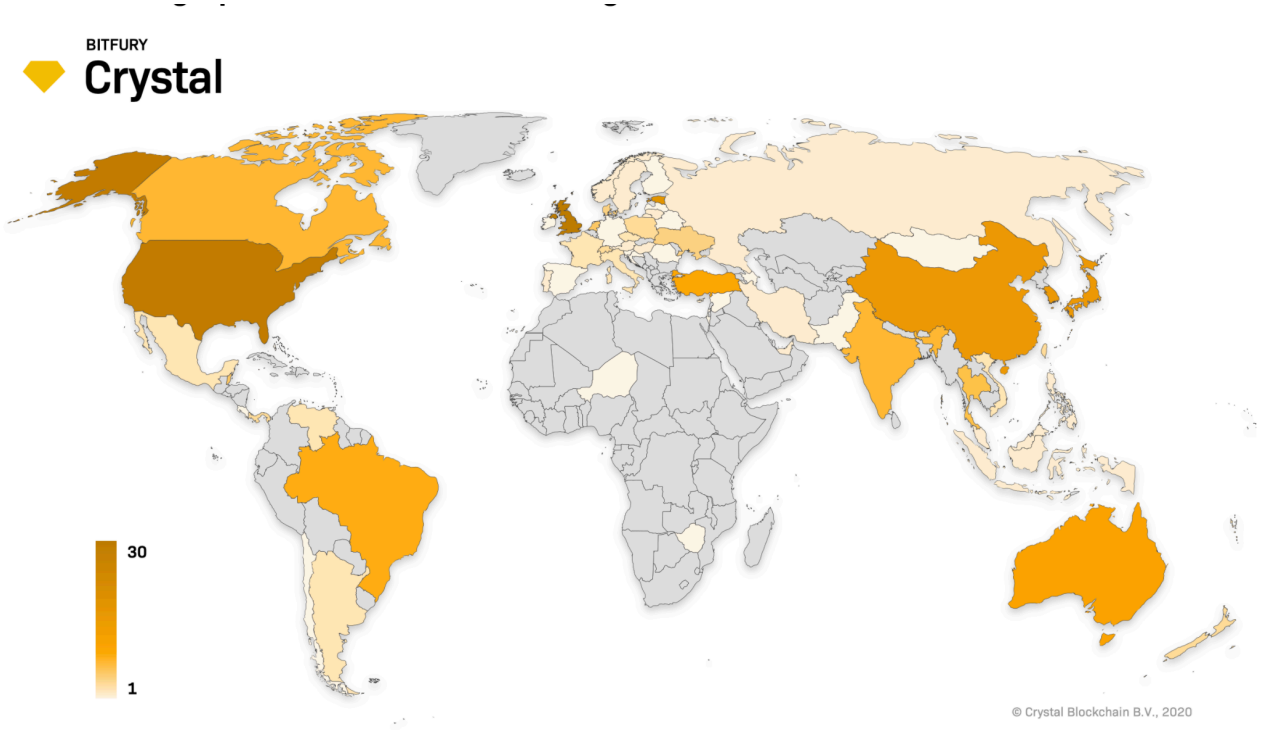

Рисунок 1.5 – Географічне розподілення криптовалютних транзакцій

Загалом на період 2020 року нараховується близько 455 ліцензованих міжнародних бірж більш ніж у 70 країнах.

1.5 Аналіз використання криптовалют у сфері Darknet

Darknet — це нелегальний інтернет, який існує на базі основного, але використовує виключно захищені проксі-сервери та інтернет-з'єднання, щоб було неможливо відслідковувати користувачів та адреси сайтів. Darknet — простір для існування нелегальної торгівлі.

Незмінність та прозорість блокчейн мереж дозволяє виявляти підозрілу активність на глобальному рівні, що було б неможливо за допомогою фіатних платежів.

Графік, який зображений малюнку нижче (див. рис. 1.6), показує взаємозв'язок між кількістю надісланих і отриманих Bitcoin до Darknet.

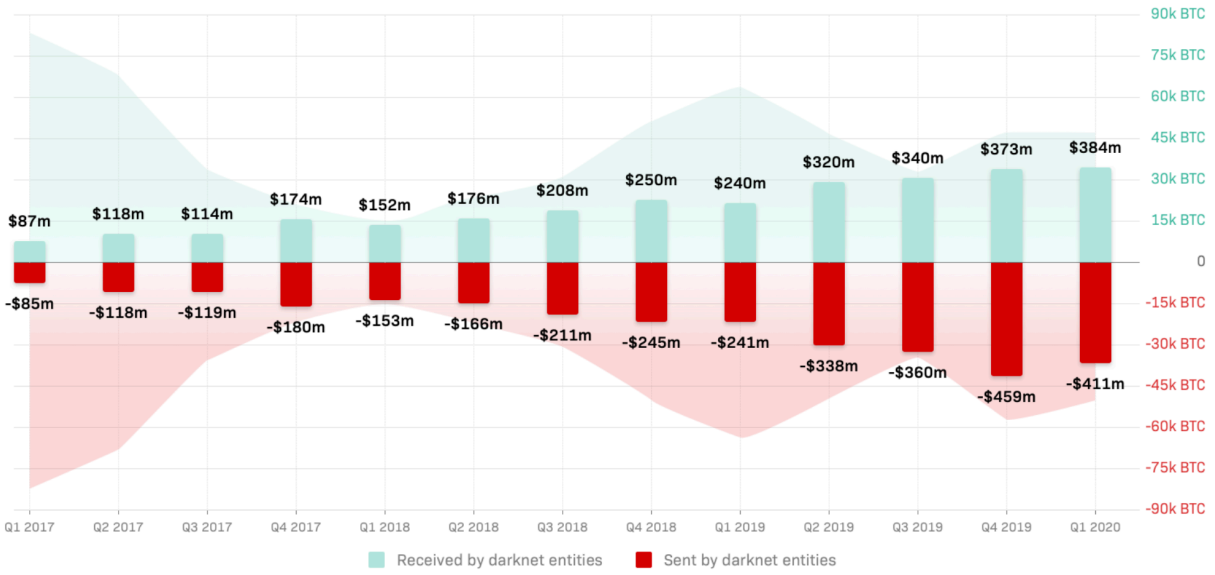

Рисунок 1.6 – Географічне розподілення криптовалютних транзакцій

Необхідно зазначити що загальна сума, отримана Darknet, зменшилася до 47 тисяч bitcoin у I кварталі 2020 року. Загальна сума, надіслана організаціями Darknet, також зменшилася до 50 тисяч bitcoin у I кварталі 2020 року.

Однак, якщо врахувати вартість Bitcoin в доларах США, ми бачимо, що Darknet отримав та надіслав збільшену суму грошей - з 384 млн. доларів США в 1 кварталі 2019 року, до 411 млн. доларів США в 1 кварталі 2020 р. Це частково пояснюється зростанням капіталізації bitcoin, а також з тим, що використовувати криптовалюту стає все простіше, популярність стрімко зростає.

Суб’єкти Darknet часто використовують біржі без суворих вимог до перевірки (імовірно, щоб уникнути обмежень KYC / AML). Частка всіх Bitcoin, отриманих

суб'єктами Darknet від таких бірж, зменшилася з 62% у I кварталі 2019 року до 45% у I кварталі 2020 року.

На малюнку нижче (див. рис. 1.7) видно, що частка Bitcoin, отриманих від бірж з вимогами до верифікації, зросла до 29% у I кварталі 2020 року (порівняно з 22% у попередньому році).

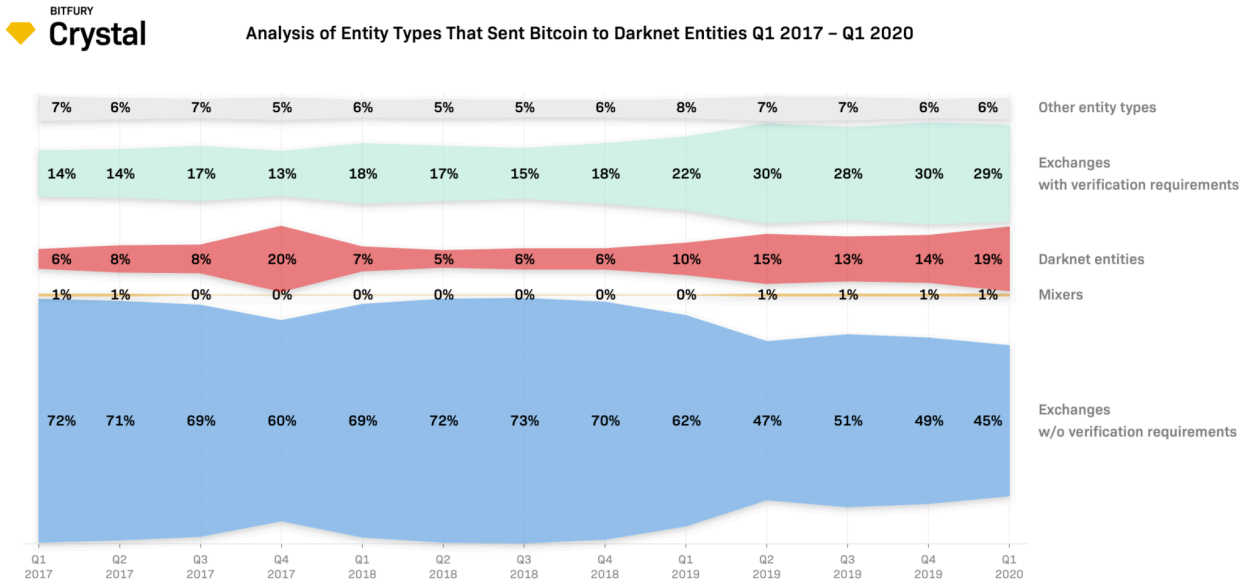

Рисунок 1.7 – Розподілення коштів від суб’єктів Darknet

Частка Bitcoin, що надсилаються від одного суб'єкта даркнету до іншого (тобто всіх транзакцій між різними суб'єктами даркнету), також зросла з 10% у I кварталі 2019 року до 19% у I кварталі 2020 року, що свідчить про загальний ріст доходів від Darknet.

Цього року обсяг Bitcoin, відправлених мікшерам від суб'єктів Darknet суттєво збільшився, на малюнку нижче (див. рис. 1.8) видно зріст – із 790 Bitcoins у I кварталі 2019 року до 7946 Bitcoins у I кварталі 2020 року. Такий же ріст спостерігався у еквіваленті доларах США. Сума Bitcoin, отриманих суб'єктами даркнету від мікшинг-послуг, також зросла втричі – зі 106 Bitcoin у I кварталі 2019 року до 288 Bitcoin у I кварталі 2020 року.

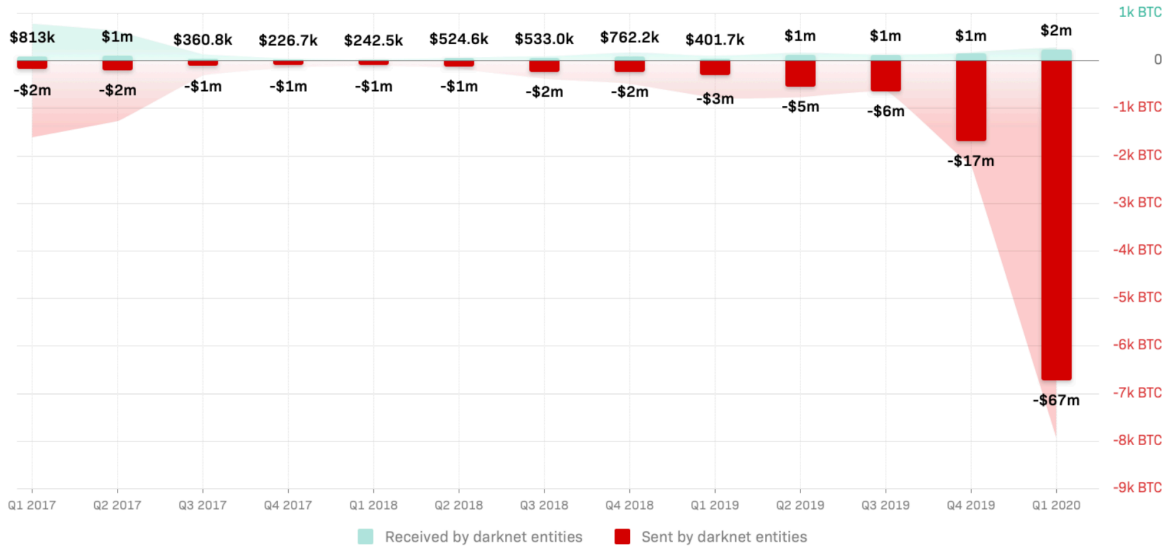

Рисунок 1.8 – Взаємодія Darknet з Crypto Mixers

З цього аналізу видно, що біржі з вимогами до верифікації стають менш популярними як спосіб виведення Bitcoin, тоді як міксери стають все більш популярними для виведення з Darknet.

1.6 Аналіз патернів поведінки виведення та передачі викрадених коштів у криптовалютах

Згідно статистики аналітичної групи Crystal, на малюнку нижче (див. рис. 1.9) видно, у 2020 році криптозлочинці намагалися вивести викрадені та отримані шахрайством активи зі швидкістю в 13 разів швидше, ніж п’ять років тому. У 2015 році найпопулярнішим способом виведення викрадених коштів – біржі з вимогами до верифікації клієнта. Ця кількість різко впала в 2020 році і склала лише 8% з загальної кількості викрадених коштів. Міксери та біржі без вимог верифікації були основними напрямками для виведення криптовалютного фонду у 2020 році. Зазвичай криптозлочинці використовують декілька додаткових операцій з криптовалютами на

невідомими проміжними адресами, перед тим як надіслати та вивести вкрадені кошти через відомі біржі.

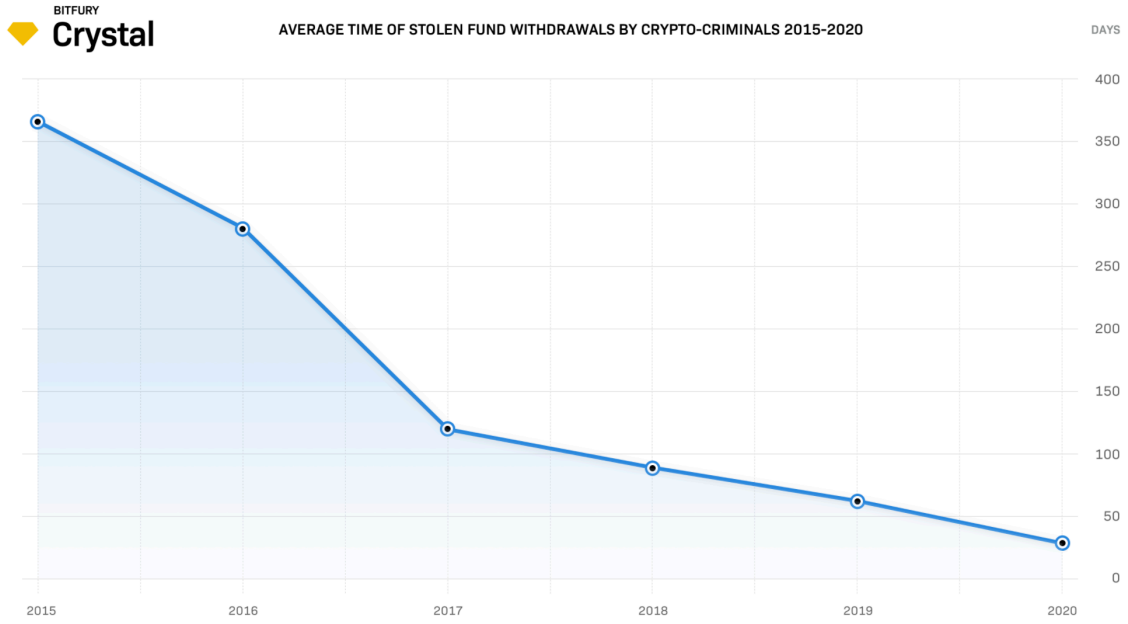

Рисунок 1.9 – Середній час виведення викрадених коштів

Між 2015–2020 роками близько 81% усіх переказів від криптозлочинців на відомі біржі було здійснено з 9 стрибками між ними.

Згідно з даними наведеними вище, можна зробити висновок, що наразі для прикриття слідів злочинці використовують скорочений час на виведення викрадених коштів зі своїх адрес, збільшують кількість проміжних операцій з невідомими адресами, використовують міксери та біржи без вимог перевірки та верифікації особи клієнта, а тема аналізу транзакцій криптовалют для перевірки їх законності сучасна та актуальна.

1.7 Аналіз сучасних сервісів для аналізу блокчейну

Сьогодні, розслідування кіберзлочинності в області криптовалюти неможливо без використання аналітичних інструментів для блокчейн мереж. Компанія CryptoLocker розробила власний спосіб аналізу блокчейн інформації[8, 9], дослідникам вдалося розробити прототип, що дозволяє ідентифікувати цифрові сліди, які можуть більш детально розкрити інформацію про особистість, що їх залишила[10]. Рід і Гарріган[11] розповіли про труднощі щодо загального питання анонімності у блокчейн мережах та виявлення поведінки реального користувача. Вони розробили Blockchain Inspector[12] – система, що використовує штучний інтелект для ідентифікації та створення профілю користувача блокчейн мережі, та дозволяє відстежувати їх поведінку.

Дослідження економічних аспектів поведінки користувачів блокчейн мереж є також одним із видів блокчейн-аналізу. Мозер і Боме[13] були сконцентровані на розгляді інформації про розподілення комісії від транзакцій Bitcoin, таким чином вони намагалися визначити поведінку користувачів[14]. Лішке і Фабіан[15], а також Рон і Шамір[16] провели аналіз ринку за допомогою блокчейну, поєднавши мережеві дані з координатами геолокацій, таким чином вони отримали уявлення про розподіл криптовалютного бізнесу. Обидва дослідження використовують blockexplorer.com[17] – веб-інструмент з відкритим кодом, який дозволяє візуалізувати інформацію щодо блоків та транзакцій у блокчейні.

BitConeView[18] – це веб-інструмент, який полегшує дослідження транзакцій у Bitcoin мережі. Інструмент також дозволяє відстежувати витрати, дозволяючи ідентифікувати закономірності та потік грошей. BitIodine – ще один інструмент для аналізу блокчейну. Надає базову інформацію, наприклад баланс гаманця та загальну кількість транзакцій. Обидва інструменти були протестовані користувачами та

демонструють ефективний спосіб аналізу та виявлення патернів поведінки всередині Bitcoin блокчейн мережі.

Blockchain.info[19] є одним з найпопулярніших сервісів, що вперше з'явився на ринку ще в 2011. Цей інструмент надає швидкі та прості у використанні можливості для відстеження окремих транзакцій, а також надає велику кількість інформації, включаючи базові діаграми та статистику, про всю Bitcoin мережу. Ортега використовував публічну інформацію з blockchain.info деякий проміжок часу, щоб деанонімізувати адреси мережі Tor та їх проксі сервера. Blockchain.info надає інформацію в зручному вигляді та дозволяє аналітикам переглянути кожну транзакцію.

Кінкельдей, Фекете та Ізенберг[20] розробили систему, яка дозволяє розпізнавати об'єкти Bitcoin мережі на основі їх публічної адреси (адреси гаманця). Інструмент називається BitConduite, він використовує топологію мережі (з її мільярдами транзакцій), надаючи оцінку до якої сутності відповідає певна адреса. Bitcoin можна використовувати у різних цілях -– починаючи від інвестицій до здійснення нелегальних платежів. BitConduite може стати корисними для вивчення та виявлення засобів використанням Bitcoin. Аналітики, які працюють з BitConduite можуть групувати та фільтрувати дані на основі різних атрибутів.

Дані про торгівлю з криптовалютних бірж можуть надати цікаве розуміння потоку грошей. Веб-сайт bitcoincharts.com[21] надає фінансову та технічну інформацію, що пов'язана з Bitcoin та може бути використана для аналізу щоденних торгових курсів, тенденцій та аномалій ринку.

Також на ринку існують комерційні програми. Chainalysis[22] був запропонований як інструмент, що дозволяє оцінити ризики, пов'язані з Bitcoin операціями. На даний час використовується правоохоронними органами під час розслідування кіберзлочинності[23].

Загалом, на ринку не існує ідеального інструменту для блокчейн-аналізу. Кожен із перелічених сервісів має свої переваги та недоліки. Наразі, повний аналіз вимагає

поєднання даних як з самого блокчейну, так із зовнішніми даними, отриманими за допомогою блокчейн-аналітики, вікі або публічних форумів.

1.8 Аналіз досліджень у сфері анонімності Bitcoin транзакцій та методів їх кластеризацій

На сьогодні існує багато досліджень у сфері кластеризацій адрес та анонімності Bitcoin мережі. Роботи базуються на застосуванні різноманітних евристичних підходів або статистичних методів. Існують методи кластеризації Bitcoin адрес шляхом перебору транзакцій, аналізуючи їх трафік та доповнюючи його зовнішньою інформацією[24]. Ще в одній роботі[25] була проаналізована система, в якій кластеризація була побудовані на основі тестових транзакцій до сервісів. Експериментальним шляхом, автор взаємодіяв з різними криптобіржами, аналізував рух коштів та ідентифікував і згрупував дані гаманців у кластери[26], також він показав що можна виявити тип сервісу[27], шляхом аналізу даних його поведінки. Збираючи транзакції в режимі реального часу[28], розроблена модель системи, що дозволяє кластеризувати адреси гаманців та прив’язати їх до IP-адреси користувача. Також розроблена система[29], яка на основі аналізу графів блокчейн мереж, паралельно використовує дані з інтернет-форума для кластеризації адрес.

Багато дослідників наголошували на недоліку у псевдо анонімності Bitcoin[30], вивчали альтернативні криптовалюти. У зв'язку з цим була побудована альтернатива Bitcoin – блокчейн мережа Zerocash[31], а також наведені теоретичні напрацювання підвищення приватності в Bitcoin мережах[32].

1.9 Постановка задачі

Після проведення огляду ринку та актуальних методів аналізу транзакцій криптовалют можна зробити висновок, що більшість сервісів мають застарілий інтерфейс, які повільні та складні у використанні звичайному користувачу. Кожен із сервісів будує власну методику кластеризації транзакцій згідно з відповідними показниками. Під час аналізу було виявлено декілька досліджень, які для більш ефективних методів кластеризації використовували алгоритми машинного навчання без вчителя. Проаналізувавши отримані дані, була сформована гіпотеза щодо можливостей використання комбінації оригінальних публічних даних блокчейну, згрупованих і кластеризованих даних наявних систем, публічної інформації та використання алгоритмів машинного навчання під наглядом, для покращення швидкості роботи системи та розв'язання задачі кластеризації ще невизначених об’єктів блокчейн мереж.

Метою роботи є дослідження актуальних підходів аналізу транзакцій криптовалют для перевірки їх законності. Враховуючи складність задачі та важливу складову системи – швидкість роботи, була розглянута можливість використання графових баз даних. Було прийнято рішення побудувати платформу та дослідити підходи до аналізу транзакцій криптовалют з використанням цих баз. У результаті роботи потрібно:

– базуючись на рекомендаціях FATF розробити класифікацію сервісів, операції з якими можуть сигналізувати потенційну загрозу;

– навчити алгоритм машинного навчання ідентифікувати потенційно небезпечні транзакції, на базі кластеризованої інформації.

– використовувати відкриті джерела даних і додати їх в алгоритм класифікатора;

– поєднати функціонал системи з інтуїтивно зрозумілим користувацьким інтерфейсом.

Для розробки системи були обрані такі технології:

– Neo4j (графова база даних);
– Node.js(API сервіс);
– Angular;
– Python та Scikit;
– Redis (розподілене сховище пар ключ-значення);
– Docker;
– S3 (сервіс-сховище даних);
– EC2.

Для дослідження будуть використані публічні дані блокчейну, що базуються на даних з Bitcoin Node, а також дані сервісів Walletexplorer та Chainalysis. Датасет складеться з більш ніж 1,5 млрд транзакцій та має розмір більше ніж 700 ГБ.

Під час дослідження необхідно проаналізувати способи розв’язування задачі кластеризації адрес криптовалютних гаманців за допомогою використання алгоритмів машинного навчання, знайти оптимальний по швидкості алгоритм пошуку ланцюгів даних.

В результаті дослідження буде розроблений зручний у використанні веб-додаток, де користувач матиме можливість ввести геш транзакції та отримати повноцінний звіт про потенційну загрозу та «чистоту» транзакції.

2 АНАЛІЗ МЕТОДІВ ДЛЯ ДОСЛІДЖЕННЯ

2.1 Огляд базових концептів роботи блокчейн мереж та їх анонімності

Анонімність – одна з найважливіших особливостей технології блокчейн. Сьогодні кожен може створити власний криптогаманець та використовувати його для надсилання або отримання коштів, не залишаючи свої персональні дані.

Вірогідно ідентифікувати особу за адресою криптовалютного гаманця без проходження KYC процедури неможливо, однак, завдяки аналізу блокчейн мереж, криптогаманці можуть бути згруповані залежно від їх поведінки[33].

Для того, щоб здійснити транзакцію у Bitcoin мережі, користувач повинен мати криптогаманець. Найбільш фундаментальним об’єктом в блокчейні є адреса криптогаманця. Найпоширеніша його форма складається з пари відкритого та приватних ключів. Відкритий ключ використовується для ідентифікації цієї адреси в ланцюзі блоків, наприклад для отримання Bitcoin. У той час як приватний ключ ECDSA, сформований із випадкового числа довжиною 256 біт, використовується для криптографічного підпису транзакції.

Другий найпоширеніший тип адреси – P2SH, де ключем для переказу коштів є не геш ключа, а геш сценарію. Це дозволяє проводити більш складні транзакції, де потрібно знати кілька ключів, пароль або що-небудь, щоб задовольнити виконання сценарію. Загальна транзакція складається з чотирьох основних елементів:

– геш транзакції;
– адреса відправника;
– адреса отримувача;
– сума.

Кожна транзакція в новому блоці обов’язково перевіряється майнерами, щоб переконатися, що жодні монети не витрачаються двічі. Транзакції нового блоку обробляються в єдиний геш, яке є коренем дерева Меркла[34, 35].

Така бінарна деревоподібна структура містить лише транзакції в листках. Схема гешування (див. рис. 2.1),

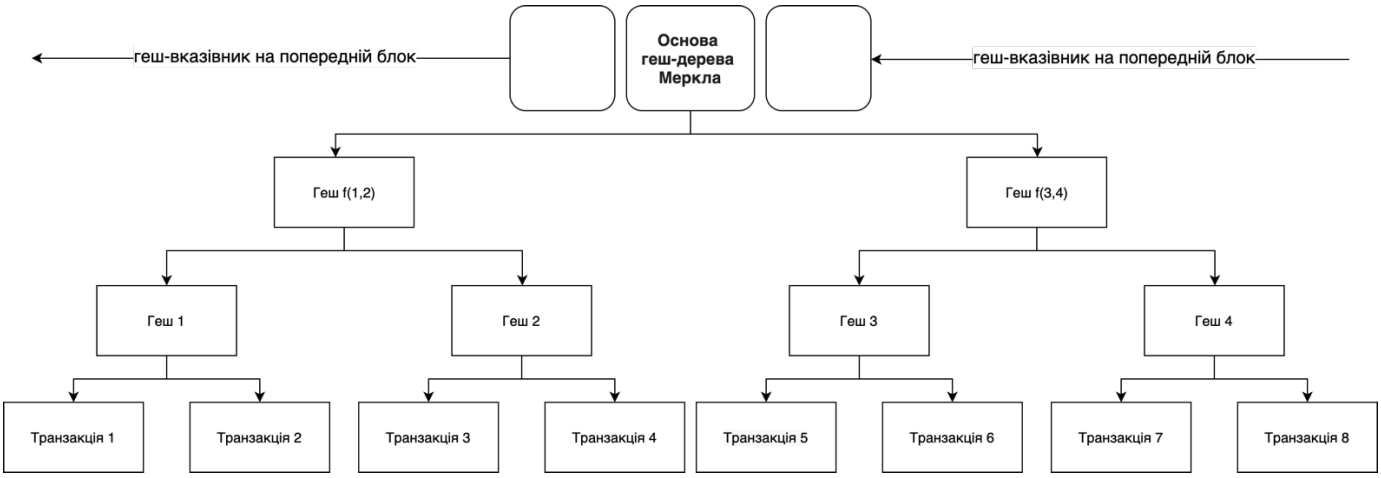

Рисунок 2.1 – Структура Bitcoin мережі на основі дерева Меркла

До кожного нового блоку додається геш попереднього блоку. Узагальнено процес транзакції виглядає наступним чином:

– знімається повна сума з гаманця відправника;

– кошти відправляються на адресу одержувача;

– формується решта (віднімається комісія);

– решта відправляється на адресу гаманця відправника;

– транзакція відправляється в мережу, де формується в блоки разом з іншими транзакціями мережі;

– коли всі вузли мережі підтвердять операцію за алгоритмом консенсусу, транзакція буде публічно доступною.

Необхідно зауважити, що відправник може зазначити іншу (відмінну від початкової) адресу для повернення решти, це є один зі способів ще більше анонімізувати транзакцію. Кожна транзакція повинна бути підтверджена відправником з використанням приватного ключа, котрий був згенерований під час створення гаманця.

Як наголошувалось раніше, переваги використання блокчейн мереж – анонімність, але особливість архітектури Bitcoin блокчейн мереж полягають у тому, що коли адреса гаманця пов'язується з реальним власником, існує можливість розкрити та ідентифікувати усі операції, здійснені ним, без можливості видалення історії транзакцій.

2.2 Огляд методів кластеризації криптовалютних гаманців

Кластер – скупчення гаманців, власниками яких є один користувач. На сьогодні кластери прийнято ідентифікувати за деякими характеристиками.

Кластеризація на основі патернів отримання коштів: більшість бірж автоматично генерують гаманці клієнтам, після отримання коштів на такий гаманець, система автоматично перенаправляє їх на основний гаманець біржі. Прослідкувавши поведінку транзакцій, можливо виявити основні гаманці біржі, а також пул генерованих гаманців для клієнтів.

Кластеризація на основі сторонньої публічної інформації: у цьому типі кластеризація відбувається за допомогою інформації, що збирається ззовні блокчейн мережі. Джерела даних різноманітні: судові документи, партнерська інформація, публічна інформація бірж, та зібрані дані вручну.

Кластеризація на основі спільної вхідної адреси: полягає в групуванні адрес, які належать одній особі. Оскільки транзакції у блокчейні анонімні, майже неможливо зробити висновок, кому належить криптогаманець або кому спрямована транзакція. Ідея цієї кластеризації полягає у використанні інформації, що зберігається в об’єктах транзакцій, для побудови кластерів за адресою. Згідно з припущенням приватний ключ, що використовується для підписання транзакцій, повинен знаходитись в руках тієї самої особи (відправника).

Якщо розглянути тип транзакції, коли гроші відправляються з двох криптогаманців на одну адресу (див. рис. 2.2), можливо зробити висновок, що приватний ключ обох адрес, з яких відправляється решта коштів, повинен знаходитись під контролем однієї особи.

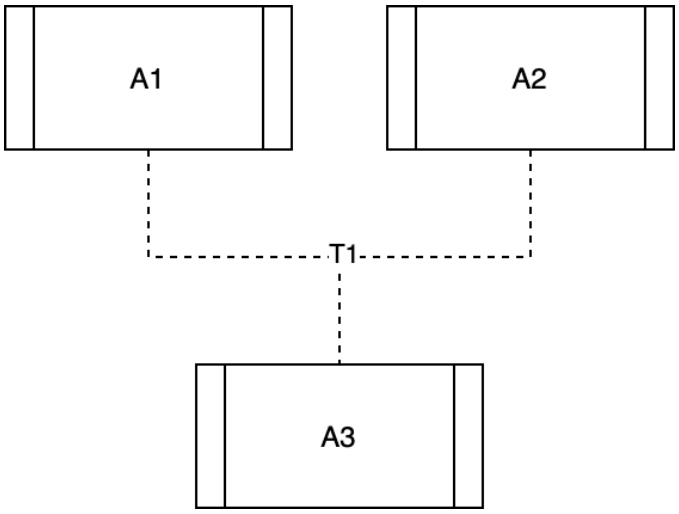

Рисунок 2.2 – Кластеризація на основі спільної вхідної адреси

Адреси A_1 і A_2 здійснюють переказ коштів на адресу A_3 транзакцією T_1. Оскільки обидві адреси A_1 і A_2 є вхідними даними в цій транзакції, приватний ключ обох перебуває в руках особи X.

Тепер можна проаналізувати дані у блокчейні та знайти усі транзакції, в яких брала участь адреса A_1. Зрештою, з раніше не пов’язаних транзакцій, можемо сформувати кластер усіх адрес, що контролюються однією особою.

Кластеризація на основі спільної адреси відправлення решти – коли відправник хоче зробити транзакцію, в якій використовується лише частина грошей, можливо зробити припущення, що решта повинна бути надіслана на нову адресу, яка також під контролем відправника.

Припустимо, що особа X_1 хоче передати 4 Bitcoin особі X_2. Для здійснення цього переказу X_1 використовуватиме 5 Bitcoin, доступні йому за адресою A3. Мережа повинна надіслати решту, на раніше вказану адресу відправником.

Ця адреса (A_5) і буде використовуватися для групування усіх транзакцій пов'язаних з відправником (див. рис. 2.3), оскільки можливе ствердження, що приватний ключ обох адрес перебуває в руках однієї особи X_1.

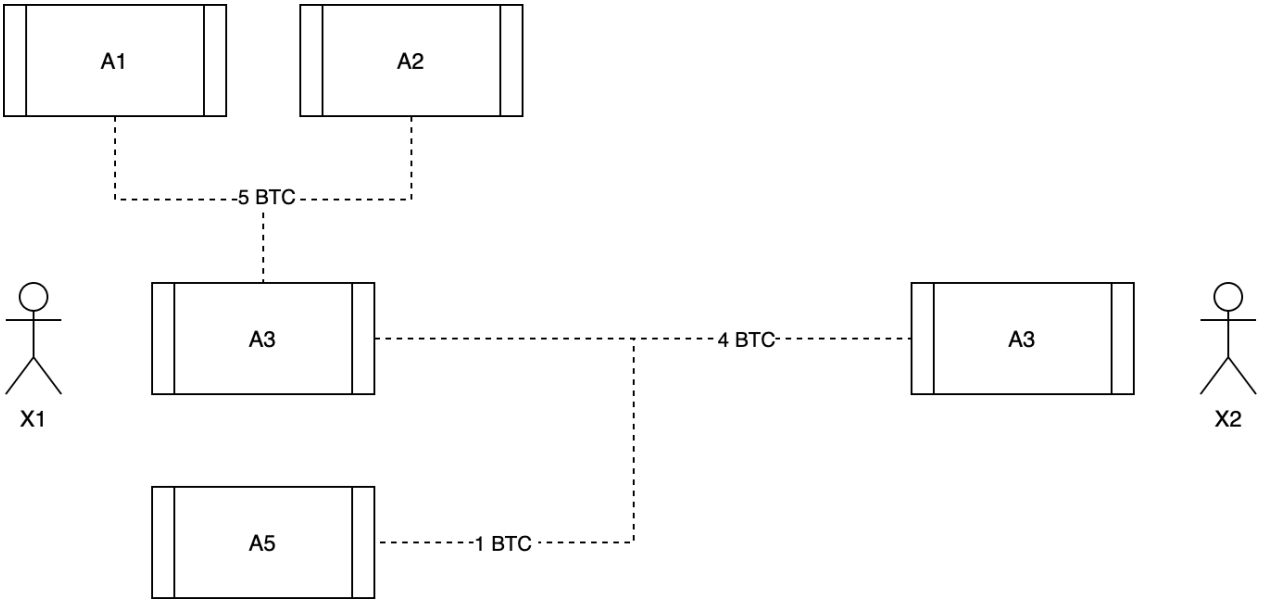

Рисунок 2.3 – Кластеризація на основі спільної адреси відправлення решти

Ще однією інформацією, яка може бути корисною, є те, на якій відстані розташовані дві адреси в блокчейні, також можливо враховувати кількість транзакцій між ними. Ці дані можуть бути використані для об'єднання двох раніше не пов'язаних між собою адрес.

Функціонал простий, потрібно вказати лише початкову адресу та кінцеву адресу. Третім параметром є обмеження глибини, яке залишається за вибором користувача. Система виконує пошук у ширину, переглядаючи транзакції, в яких брала участь кожна з адрес (див. рис. 2.4).

Алгоритм починається з пошуку всіх транзакцій, в яких брала участь адреса A_1, і перебирає їх та додає до черги, поки не дійде до заданої кінцевої адреси. Після знаходження кінцевої адреси, шлях, за яким проходять адреси, буде еквівалентний пошуку в глибину.

Необхідно зазначити, що існує багато сценаріїв, коли на одну адресу відправляють кошти з суміжних вузлів. Наприклад, існує дві транзакції, T_1 і T_2, в яких T_1 перераховує кошти між A_3 – A_7, а T_2 між A5 – A7.

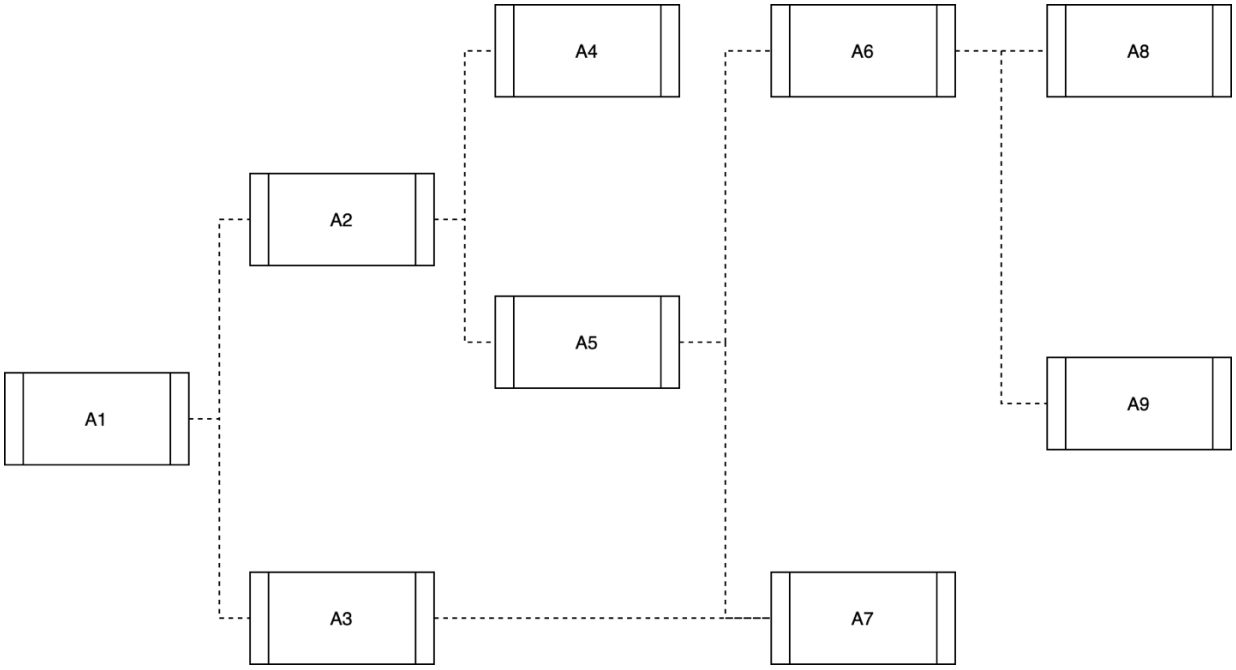

Рисунок 2.4 – Кластеризація на основі відстані між попередніми транзакціями

Оскільки дерево обходить дерево за історією транзакцій у хронологічному вигляді, алгоритм спочатку бачить транзакцію T_1, і ніколи не усвідомлює, що існує інший (довший) шлях, якби він йшов іншою гілкою. Оскільки алгоритм повинен повертати найкоротший шлях між двома адресами, було вирішено, що це не проблема.

Ключовою проблемою, яку слід згадати, є транзакція, в якій кілька десятків адрес використовуються як вихідні дані. У цій ситуації, оскільки створюється багато гілок, алгоритм недостатньо розумний, щоб відкинути шляхи, які нікуди не потраплять.

2.3 Огляд методів оцінки ризику та категоризації кластерів

Оскільки використання криптовалют стає все більш розповсюдженим, постачальники послуг роботи з криптовалютами повинні боротися з загрозами відмивання грошей. Група розробки фінансових заходів боротьби з відмиванням грошей (FATF) у 2020 році опублікували звіт про індикатори, що свідчать про процес відмивання грошей та фінансування тероризму. На основі цього звіту постачальники послуг роботи з криптовалютами, фінансові органи та криптовалютні біржі повинні розробити та впровадити програми, що протидіють відмиванню грошей.

Звіт містить наступні показники, які необхідно відстежувати:

– тип транзакції;

– дані відправника та одержувачі;

– джерело коштів;

– географічні ризики;

Якщо клієнт намагається виконати наступні операції, вони можуть свідчити про активний процес відмивання грошей та фінансування тероризму:

– аномально багата кількість криптовалютних транзакцій невеликими коштами;

– серія транзакцій аномально великими коштами за короткий проміжок часу;

– аномально багато перетворень депозиту коштів на різні типи криптовалют;

– виведення коштів на криптовалютні гаманці, які були визнані вкраденими;

– внесення аномально великого депозиту, який потім продається або знімається повністю в той самий день (або незабаром після цього);

– внесення коштів невеликими транзакціями з не пов'язаних між собою гаманців, які негайно переносяться на інший гаманець;

– внесення або переведення коштів на гаманці, які пов’язані з Dark market.

FATF рекомендує кожній організації створити власну систему оцінки ризику транзакцій криптовалют, завдяки якої можлива класифікація загроз.

Оцінка ризику – це показник, який допомагає оцінити ймовірність пов’язання адреси з незаконною діяльністю. Значення може коливатися від 0% до 100%, де 0% означає, що адреса безпечна, а 100% означає, що адреса причетна до незаконної діяльності.

Нижче наведені основні категорії кластерів (за стандартною класифікацією):

– адреси бірж: організацій, що дозволяють своїм клієнтам торгувати криптовалютами;

– генератори гаманців: довірені організації, які пропонують послуги зберігання криптовалюти;

– торгівельні організації;

– майнінг: сервера, які беруть участь в обчисленні гешів;

– міксери: організації, які надають послуги для змішування коштів з різних джерел, щоб зробити їх відстеження важчим або майже неможливим. В основному використовується для відмивання грошей;

– адреси сервісів азартних ігор: організацій, які пропонують азартні послуги;

– адреси, які пов’язані з шахрайством;

– тор: торгові майданчики, які в першу чергу сприяють торгівлі незаконними товарами, такими як наркотики, викрадені кредитні картки, паспорти тощо.

2.4 Огляд використання графових баз даних

Зберігання даних у реляційних СУБД має свої переваги, але коли є необхідність зберігати мільйони зв'язків між даними такі бази не підходять. Для вирішення цих проблем можна змінити СУБД.

Neo4j – графова база даних з відкритим вихідним кодом, розроблена на мові Java, з підтримкою транзакцій (ACID). Дані зберігаються у власному форматі збереження вузлів та ребер (спеціалізовані сховища графів). Основні області застосування: соціальні мережі, системи надання рекомендацій, картографічні системи.

Особливості Neo4j:

– має вузли і зв'язки (відносини);

– вузли можна розглянути як документи, що містять властивості у вигляді пар ключ-значення;

– одному вузлу можна призначити декілька міток.

Відносини також можуть містити властивості. Це дозволяє встановлювати додаткові метадані в графічні алгоритми, додавати додаткову семантику зв'язку.

Для запитів для використання використовується Cypher – декларативна мова запитів до графа. Синтаксис схожий на синтаксис SQL. Підтримуються операції створення, вибору, оновлення, видалення даних.

2.5 Огляд області використання алгоритмів машинного навчання з вчителем

При навчанні з учителем машина навчається на прикладах. Оператор забезпечує алгоритм машинного навчання набором відомих даних, які містять необхідні вхідні та вихідні значення. Алгоритм повинен встановити, як даними входів отримати дані виходів. Сам оператор знає рішення поставленого завдання; алгоритм виявляє закономірності в даних, вчиться на основі спостережень і робить прогнози. Ці прогнози потім коригуються оператором. Процес триває до тих пір, поки алгоритм не досягне високого рівня точності / продуктивності. При збільшенні даних для аналізу

росте здатність алгоритму приймати рішення на основі цих даних, а також точність цих рішень.

За допомогою алгоритмів машинного навчання з вчителем можна розв’язати проблему кластеризації. Кластеризація передбачає групування наборів схожих даних (на основі певних критеріїв). Це корисно для сегментації даних на кілька груп і проведенні аналізу на основі кожного набору даних окремо для пошуку закономірностей.

Існує велика кількість алгоритмів навчання з учителем, кожен з яких має свої сильні й слабкі сторони.

До числа алгоритмів навчання з учителем для вирішення задач класифікації відносяться:

– метод k-найближчих сусідів;
– випадковий ліс;
– зайві дерева;
– бутстреп агрегація;
– посилення градієнту.

Слід також зауважити, що не існує єдиного алгоритму, який найкраще підходить для всіх завдань аналізу. Навіть найдосвідченіші фахівці з аналізу даних не зможуть сказати вам, який алгоритм буде працювати краще, не проекспериментувавши різними їх видами для розв'язання встановленої задачі.

3 ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ

3.1 Проектування архітектури ПЗ

Проектування програмного продукту було проведено з використанням діаграм мови UML[36]. Була створена діаграма розгортання (див. рис.3.1).

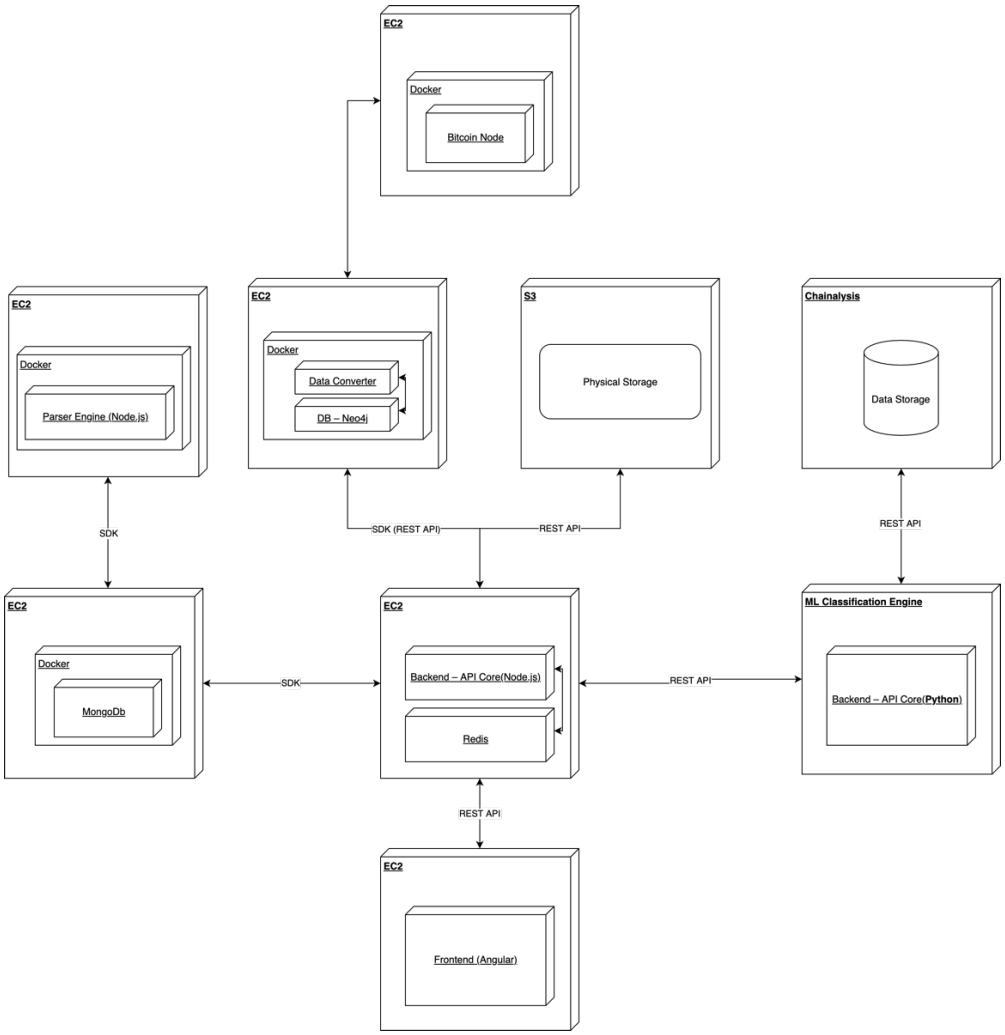

Рисунок 3.1 – Діаграма розгортання

Пропонується реалізувати систему, яка комбінує та використовує дані кластерізованих адрес, відстежує актуальні дані транзакцій Bitcoin мережі та публічну інформацію, завдяки чому здатна згенерувати звіт про «чистоту» транзакції.

Для використання актуальних даних транзакцій Bitcoin мережі, а також збільшення швидкості взаємодії з даними необхідно побудувати додаткову систему парсингу транзакцій до Neo4j бази (див. рис.3.2).

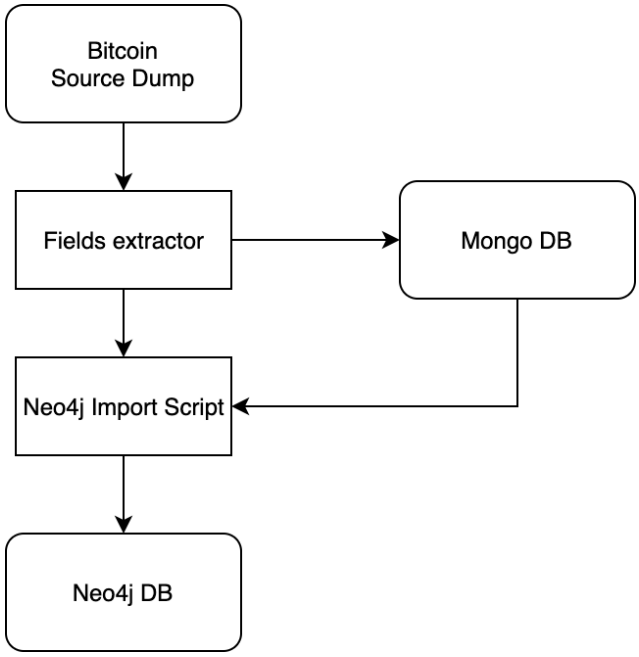

Рисунок 3.2 – Процес міграції даних до Neo4j бази

Під час міграції даних до Neo4j бази були скопійовані наступні дані:

– геш транзакції;

– адреса відправника;

– адреса отримувача;

– кількість надісланих коштів;

– дата та час транзакції;

– геш блоку транзакції;

Для реалізації системи було вирішено використовувати клієнт-серверну архітектуру. Цей вибір був зроблений з наступних причин:

– так як усі обчислення виконуються на сервері, то вимоги до клієнта знижуються;

– так як між між сервером та клієнтом передаються невеликі дані, це дозволяє розвантажити мережу;

– перевага у вигляді відсутності дублювання коду програми-сервера програмою-клієнтом;

На самому примітивному рівні абстракції додаток, орієнтований на роботу з сервером, складається з наступних архітектурних шарів:

- ядро додатка, яке включає в себе компоненти системи, не доступні для взаємодії з користувачем;
- графічний інтерфейс;
- компоненти повторного використання: бібліотеки, візуальні компоненти й інше.

Найважливішою умовою побудови додатку є відділення ядра системи від графічного інтерфейсу системи, настільки, що б одне, могло успішно функціонувати без іншого[37].

Для серверної частини була обрана RESTful архітектура. Існує шість обов’язкових вимог для побудови розподілених REST-додатків по Филдингу:

– модель клієнт-сервер;

– відсутність станів;

– кешування;

– одноманітність інтерфейсів;

– слої;

– код за вимогою.

Під час реалізації серверної частини було використано наступні патерни проектування:

– фабрика;

– проксі;

– абстракція;

– прототип;

– міст;
– фасад;
– декоратор.

Загальна клієнт-серверна архітектура наведена на рис.3.3.

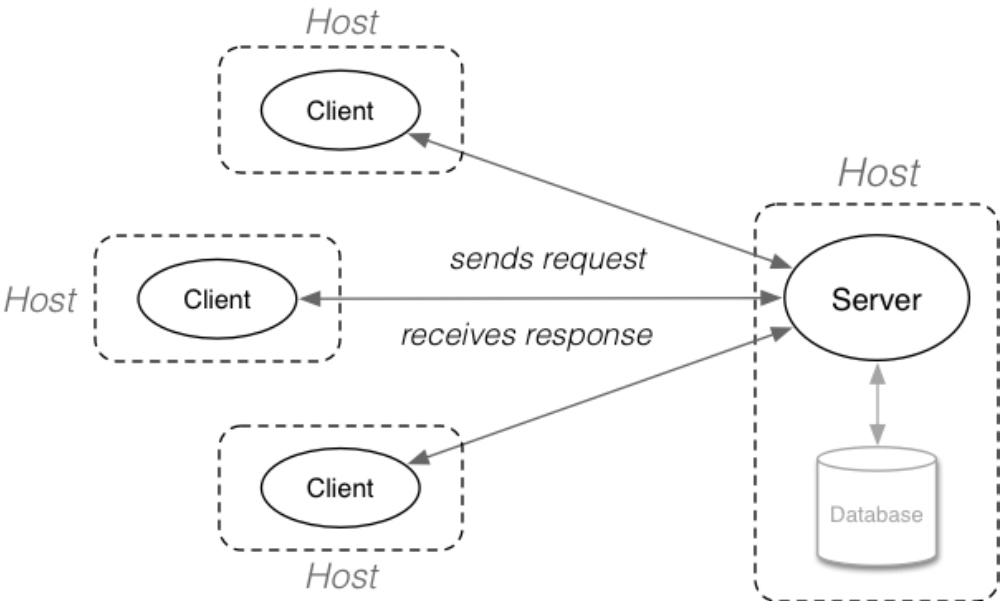

Рисунок 3.3 – Опис клієнт-серверної архітектури

Фабрика дозволяє відокремити створення об'єкта від його реалізації. Фактично обгортання фабрикою створення нового екземпляру та дає більшу гнучкість і керованість цього дійства. За допомогою фабрики можна створити новий екземпляр, використовуючи замикання, прототип і оператор new, метод Object.create(), або навіть повернути екземпляр з певним станом. Завдяки замикань ми використовували фабрику як механізм інкапсуляції.

Проксі – це об'єкт, керуючий доступом до іншого об'єкту, який називається суб'єктом[38]. Проксі і суб'єкт мають ідентичні інтерфейси, і це дозволяє прозоро підміняти одного іншим. Ми використовували проксі під час перевірки допустимості вхідних даних перед відправкою на клієнт, а також під час реалізації кешування - проксі зберігає внутрішній кеш, викликаючи суб'єкта, тільки якщо викликані дані ще відсутні в кеші.

3.2 Огляд бази даних від компанії Chainalysis

Як вже згадувалося раніше, для навчання нейронної мережі буде використовуватися набір даних, що був наданий компанією Chainalysis, яка спеціалізується на аналізі Bitcoin блокчейн мережі. Набір даних містить корисну інформацію про кожну транзакцію, яку надіслав користувач системи, наприклад: час відправлення транзакції, кількість коштів, адреса отримувача. Набір даних також містить характеристики кожного кластера і в деяких випадках категорії вже є визначеними.

Як показано в таблиці 3.1, загальний набір даних, використаний у цьому дослідженні містить приблизно 200 мільйонів транзакції, а це 321 унікальних кластерів. Кількість транзакцій на кластер суттєво коливається, від малої кількості до декількох мільйонів операції.

Таблиця 3.1 – Статистика вхідних даних

Характеристика	Кількість
Загальна кількість транзакцій	189108465
Кількість унікальних кластерів	321
Середня кількість транзакцій на кластер	456445,52
Найменша кількість транзакцій у кластері	9
Найбільша кількість транзакцій у кластері	27049997

Крім того, таблиця 3.2 ілюструє кількість транзакцій для кожної категорії більш детальніше: середнє значення, медіана, мінімальна та максимальна кількість

транзакцій. Для кожної транзакції є свої атрибути. Ця інформація потрібна для аналізу поведінки кластера.

Таблиця 3.2 – Кількість транзакцій за категорією кластеру

Категорія	$Тр_{сер}$	$Тр_{мед}$	$Тр_{мін}$	$Тр_{макс}$
Адреси бірж	564227	45424	7	37948999
Азартні ігри	448366	28934	73	24095779
Генератори гаманців	11881254	112562	11531	8657104
Торгівельні організації	579738	233738	2521	3255842
Майнінг	897631	98367	725	12852924
Міксери	6780273	211123	231	23827594
Інші	94535	23185	589	760392
Вимагальники	3336	2893	74	9149
Шахраї	32388	29946	838	79068
Тор	392145	28985	81	4215288

Для навчання алгоритмів ми будемо використовувати наступні вхідні значення:

– дата транзакції;
– кількість відправлених коштів;
– кількість отриманих коштів;
– категорія, до якої належить транзакція;
– кількість відправлених коштів до кластера;
– кількість отриманих коштів до кластера;
– загальна кількість транзакцій в кластері;
– різниця між кількістю відправлених та отриманих коштів до кластера.

Фаза аналізу даних складається з трьох основних етапів[39, 40]. Першим кроком ми вибрали та навчили набір класифікаторів мультикласів з використанням параметрів за замовчуванням. Цей крок надав попередню оцінку тому, які алгоритми можуть бути використані для розв'язання проблем. У другому кроці, ми налаштуємо гіперпараметри для кожної моделі за допомогою перехресного пошуку. На третьому

етапі, ми оцінимо продуктивність кожного алгоритму після тренування кожного моделі з відповідним набором оптимальних параметрів.

Для аналізу даних Bitcoin транзакцій ми використовували наступні алгоритми машинного навчання, які є популярними для цього типу проблем[41]:

– випадковий ліс;
– бутстреп агрегація;
– посилення градієнту.

Під час роботи над алгоритмом ми знайшли, що наш набір даних містить класи міксерів та генераторів гаманців, які важко ідентифікувати з наданого набору даних. Такі класи залишаються неідентифікованими через їх характер діяльності. Наприклад, міксери переміщують надіслані кошти за визначеною адресою. Потім вони починають надсилати з цієї адреси дуже невеликі суми на інші сервіси, повертаючи решту на абсолютно новий гаманець. Цей процес повторюється до останньої монети. Це призводить до створення десятків або навіть сотні адрес гаманців, тому їх важко ідентифікувати та кластеризувати.

Для розв'язання цієї проблеми дисбалансу класів ми будемо використовувати метод CMOT[42], щоб збалансувати недостатньо представлені класи. Цей підхід будує зразки класів для вдосконалення незбалансованих наборів даних. Було показано, що при підвищенні чутливості класифікатора до класу меншин, а також внаслідок збільшення обсягу вибірки, модель прогнозування може досягти кращих показників[43]. Метод будемо застосовувати зі співвідношенням 0,075 для двох класів з найменшою кількістю спостережень.

3.3 Результати проведення експерименту

У результаті експерименту, з усіх алгоритмів найбільшу точність надав алгоритм посилення градієнту – 84%. Якість роботи алгоритмів наведені в таблиці 3.3.

Таблиця 3.3 – Результати роботи алгоритмів

Алгоритм	Точність	Влучність	Чутливість	F-міра
Випадковий ліс	0,84	0,81	0,81	0,78
Бутстреп агрегація	0,85	0,83	0,87	0,82
Посилення градієнту	0,88	0,84	0,89	0,86

Код навчання алгоритму та його параметри наведені нижче:

```
gradientBoostingClassifier(criterion='friedmanmse', learningrate=0.3, loss=
'deviance', maxdepth=13, maxleafnodes=None, minimpuritysplit=1e-09,
minsamplesleaf=24, minsamplessplit=4, minweightfractionleaf=0.0,
nestimators=39,randomstate=none, subsample=2.0, verbose=1, warmstart=False)
```

Важливо зазначити і потенційні обмеження набору даних. Зараз дані, які використовуються для навчання моделі, не враховують додаткові дані, що можуть бути отримані з Bitcoin блокчейн мережі. Це стосується комісії за транзакцію, яка пов’язана з пріоритетом транзакції, кількість підписів, що використовувались для підписання транзакції, відповідні IP-адреси або розмір транзакції. Крім того, кількість кластерів, що використовується для навчання моделі прогнозування обмежена тими, що вже були класифіковані Chainalysis. Якщо попрацювати над обмеженнями, які приведені вище, можливо значно покращити показники точності та влучності алгоритму посилення градієнту.

3.4 Реалізація програмного забезпечення

Для моделювання UI/UX дизайну системи було створено макет сайту у програмі Sketch. Sketch – це платний векторний графічний редактор інтерфейсів для Apple's macOS. Особливість Sketch полягає в його універсальності і можливості генерувати професійні дизайни в рази швидше і якісніше ніж у інших редакторів.

Після створення дизайну сайту, було створено прототип. Прототипи корисні тим, що допомагають сфокусуватись саме на сенсі сторінки та її функціях, а також явно продемонструвати способи взаємодії клієнта з системою, перед тим як перейти до кодування. Добре створений прототип є повноцінним каркасом сайту чи будь-якого іншої системи, що має користувацький інтерфейс[44].

Для створення прототипу сайту використовувалась спеціалізована програма – «Marvel».

Серверна частина була розроблена з використанням мови Node.js. Ця технологія надає можливість реалізовувати легко масштабовані сервери за короткий проміжок часу.

Основні бібліотеки серверної частини:

– hapi-js модуль – компонент-обертка над Expresss, що дозволяє з легкістю будувати REST API endpoints;

– S3 модуль – компонент, що використовують для взаємодії зі стороннім сервісом зберігання контенту користувача.

Для розв'язання задачі кластеризації ще невизначених об’єктів блокчейн мереж, була реалізована система з використання Python та Scikit-learn бібліотеки.

На стороні веб-клієнта використовується технологія Angular, що дозволяє будувати одно сторінкові сайти (SPA), які не потребують перезавантаження[45].

Основні модулі, що були використані на front-end:

– router – компонент, завдяки якому, користувач може переходити по сторінках сайту без необхідності перезавантаження;

– reactive forms - компонент, завдяки якому ми маємо змогу зберігати отримані дані з сервера на стороні клієнта та створювати динамічні форми.

Веб-додаток було спроектовано згідно з дизайном. Відкривши сторінку сайту у користувача є можливість відразу переглянути інформацію про систему та скористатися головною функцією – перевіркою транзакції чи криптовалютного гаманця (див. рис.3.4).

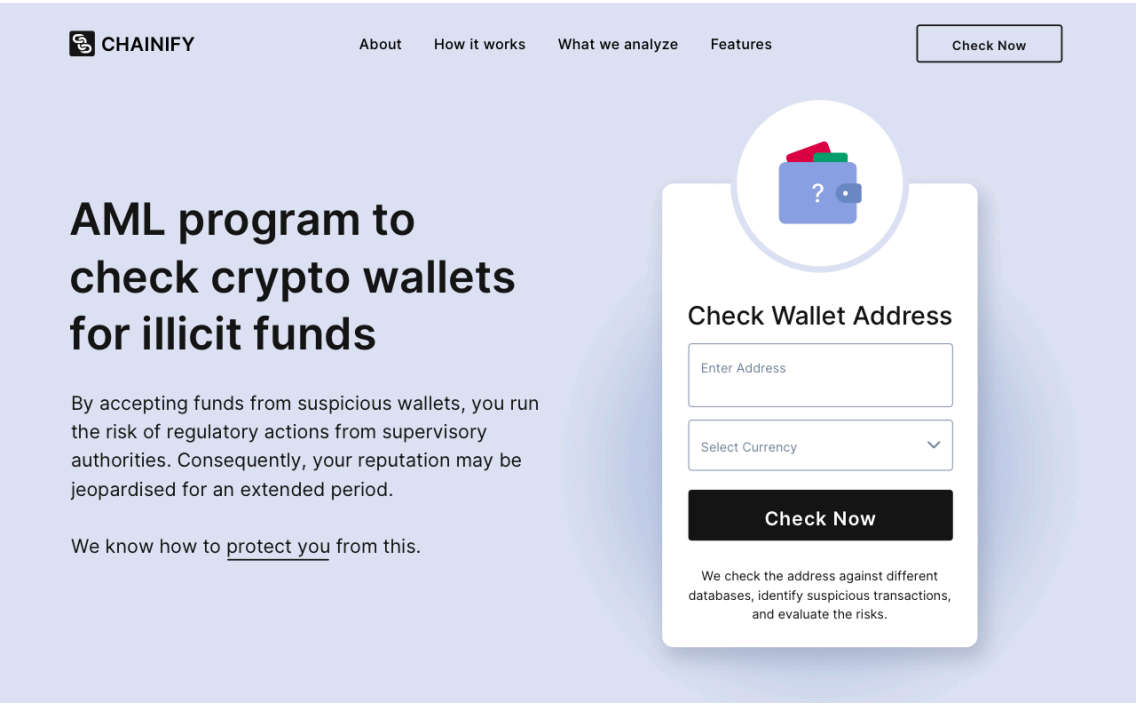

Рисунок 3.4 – Перша секція сайту

Користувач системи також має можливість переглянути інформацію про те, як працює система та що вона перевіряє (див. рис.3.5). До ризикових категорій належать наступні джерела:

– адреси, пов’язані з азартними іграми;
– адреси, пов’язані з міксерами;
– адреси, пов’язані з вимагальниками;
– адреси, пов’язані з шахраями;
– адреси, пов’язані з Tor мережами.

Рисунок 3.5 – Категорії ризиків, що аналізує система

Після вводу адреси криптовалютного гаманця, користувач повинен натиснути на кнопку пошуку інформації. Система почне аналізувати дані наступним чином:

– перевірка та аналіз історії транзакцій у Neo4j, на основі використання алгоритмів машинного навчання;

– перевірка історії транзакцій у Chainalysis;

– перевірка публічної інформації;

– генерація звіту щодо статусу транзакцій та ймовірної загрози.

Після виконання усіх наведених раніше шагів, користувач отримує звіт заданого формату з аналізом усіх потенційно небезпечних транзакцій (див. рис.3.6).

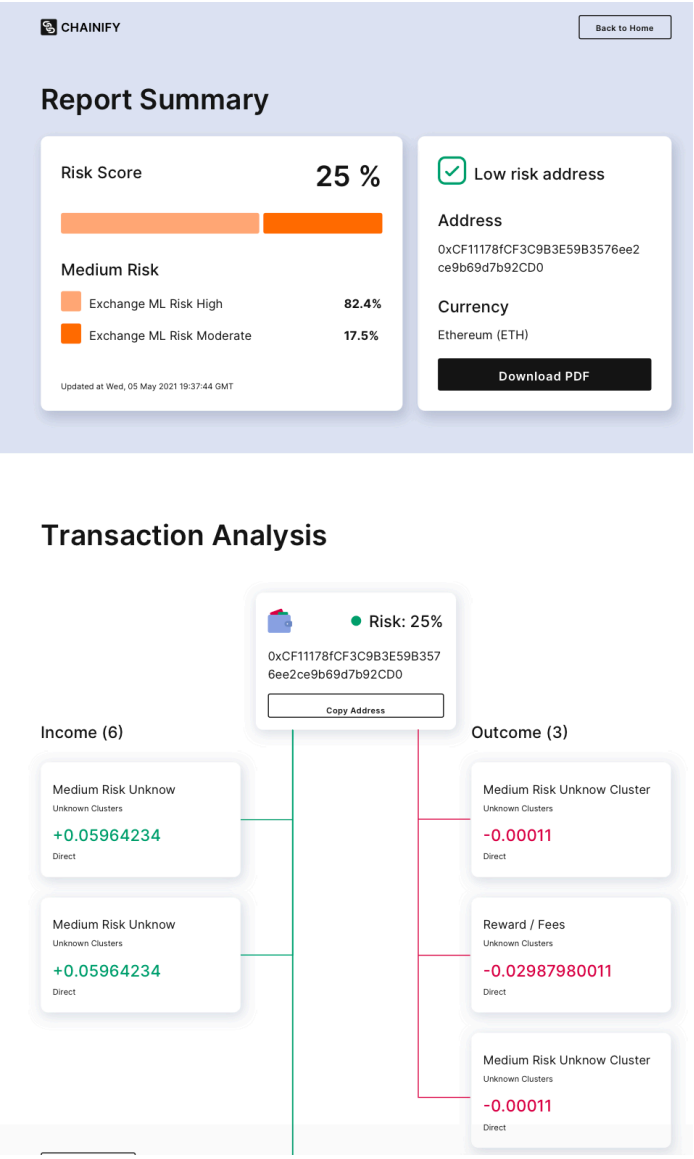

Рисунок 3.6 – Формат звіту системи

У наступних версіях системи планується розширення функціоналу, реалізація можливості детального вивчення графової структури визначеної транзакції через UI інтерфейс, а також створення платної підписки.

ВИСНОВКИ

Під час роботи був проведений аналіз предметної області. Розглянута важливість та актуальність поставленої теми, адже в останні роки криптовалюти стали невід'ємною частиною світової фінансової системи, але на фоні глобального зростання ринків криптовалют, зростає також кількість шахраїв та кібертерористичних угрупувань, дії яких потрібно відстежувати навіть звичайному користувачу.

У результаті роботи були розглянуті базові концепти роботи блокчейн мереж, сучасні методи кластеризації гаманців, розглянуті основні методи оцінки ризику та категоризації кластерів. Базуючись на рекомендаціях FATF була розроблена класифікація сервісів, операції з якими можуть сигналізувати потенційну загрозу. Був розроблений ML алгоритм, який дозволяє ідентифікувати потенційно небезпечні транзакції, а також розроблена система з інтуїтивно зрозумілим користувацьким інтерфейсом, яка використовує ML алгоритм та відкриті джерела даних. Система дозволяє ввести геш криптовалютного гаманця, перевірити історії транзакцій гаманця та ідентифікувати потенційну загрозу. Результати роботи були апробовані у науковій публікації. Слід пам'ятати, що не існує одного алгоритму, який найкраще підходить для всіх завдань аналізу, але алгоритм посилення градієнту продемонстрував найкращі результати у 84% коректного розпізнавання категорії, до якого належить транзакція.

Також необхідно зазначити, що на сьогоднішній день найбільш популярні методи кластеризації є кластеризація на основі сторонньої публічної інформації та на основі патернів отримання коштів, які також були використані в системі.

Наступним кроком до удосконалення результатів роботи системи буде аналіз методів покращення вхідних даних, які використовуються для навчання моделі, враховуючи додаткові дані, що можуть бути отримані з Bitcoin блокчейн мережі. Це

стосується комісії за транзакцію, яка пов’язана з пріоритетом транзакції, кількість підписів, що використовувались для підписання транзакції, відповідні IP-адреси або розмір транзакції, а також поведінкові патерни користувача. У наступних версіях системи планується додаткове розширення функціоналу, реалізація можливості детального вивчення графової структури визначеної транзакції через UI інтерфейс, а також розробка платної підписки.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Капіталізація криптовалют / URL: https://ru.investing.com/crypto/charts (дата звернення: 13.04.2021).

2. Аналітика шахрайств та порушення кібербезпеки, що пов'язані з участю криптовалютних операцій / URL: https://crystalblockchain.com/articles/security-breaches-fraud-involving-crypto-still-high-despite-tech-development (дата звернення: 13.04.2021).

3. Darknet / URL: https://en.wikipedia.org/wiki/Darknet (дата звернення: 13.04.2021).

4. Національний банк України – AML/CFT. Фінансовий моніторинг в Україні / URL: https://www.facebook.com/NationalBankOfUkraine/posts/2188796084667885/ (дата звернення: 13.04.2021).

5. Огляд ринку криптовалют в Україні / URL: https://umn.ua/news/971 (дата звернення: 13.04.2021).

6. Статус ринку криптовалют в Україні / URL: https://www.slovoidilo.ua/2020/12/09/infografika/biznes/rynok-kryptovalyut-yak-ukrayini-ta-inshyx-krayinax-rehulyuyut-virtualni-aktyvy (дата звернення: 13.04.2021).

7. Crystal / URL: https://crystalblockchain.com/ (дата звернення: 13.04.2021).

8. H. Kuzuno and C. Karam, "Blockchain explorer: An analytical process and investigation environment for bitcoin," 2017 APWG Symposium on Electronic Crime Research (eCrime), Scottsdale, AZ, 2017, pp. 9-16.

9. K. Liao, Z. Zhao, A. Doupe and G. Ahn. "Behind Closed Doors:Measurement and Analysis of CryptoLocker Ransom in Bitcoin." in Electronic Crime Research, 2016 APWG Symposium on 2016 Jun 1, Toronto, ON. IEEE.

10. G. Ahn, A. Doupe, Z. Zhao and K. Liao. “Ransomware and Cryptocurrency: Partners in Crime” in T. Holt (ed.) Cybercrime Through an Interdisciplinary Lens. New York: Taylor & Francis, 2017, pp. 105- 126.

11. F. Reid and M. Harrigan. “Analysis of anonymity in the Bitcoin System. Security and Privacy Social Networks” New York: Springer, 2013, pp. 197-223.

12. Blockchain Inspector. “What is Blockchain Inspector.” [Online]. Available: http://www.blockchaininspector.com (дата звернення: 13.04.2021).

13. M. Moser and R. Bohme. “Trends, Tips, Tolls:A Longitudinal Study of Bitcoin transaction Fees” in International Conference on Financial Cryptography and Data Security, 2015 Jan 30. Springer, Berlin, Heidelberg.

14. Coindesk. “About.” [Online]. Available: https://www.coindesk.com/about (дата звернення: 13.04.2021).

15. M. Lischke and B. Fabian. “Analysing the Bitcoin Network:The First Four Year”. Future Internet, 2016, vol. 8, no. 1.

16. D. Ron, A. Shamir. “Quantitative analysis of the bitcoin transaction graph.” International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2013. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-642-39884-1_2 (дата звернення: 13.04.2021).

17. Blockexplorer.com. “About block explorer.” [Online]. Available: https://blockexplorer.com [Assessed: 1-May-2021].

18. G. Battista, V. Donato, M. Patrignani, M. Pizzonia, V. Roselli and R. Tamassia. “Bitconeview: Visualisation of Flows in the Bitcoin Transaction Graph” in 2015 IEEE Symposium on Visualization for Cyber Security, VizSec. IEEE Computer Society, 2015, pp. 1-8.

19. Blockchain. “About.” [Online]. Available: https://www.blockchain.com/about/index.html (дата звернення: 13.04.2021).

20. C. Kinkeldey, J. Fekete and P. Isenberg. “BitConduite: Visualising and Analysing Activity on the Bitcoin Network” Eurographics Conference on Visualization (EuroVis), Posters Track (2017), 2017, pp. 1–3.

21. Bitcoincharts. “About”. [Online]. Available: https://bitcoincharts.com/about (дата звернення: 13.04.2021).

22. Chainalysis. “About.” [Online]. Available: https://www.chainalysis.com/#about (дата звернення: 13.04.2021).

23. E. Cheng. “Dark web finds bitcoin increasingly more of a problem than a help, tries other digital currencies.” [Online]. Available: https://www.cnbc.com/2017/08/29/dark-web-finds-bitcoin-increasinglymore-of-a-problem-than-a-help-tries-other-digital-currencies.html (дата звернення: 13.04.2021).

24. F. Reid and M. Harrigan, “An analysis of anonymity in the bitcoin system,” in Security and privacy in social networks, pp. 197–223, Springer, 2013.

25. E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, “Evaluating user privacy in bitcoin,” in International Conference on Financial Cryptography and Data Security, pp. 34–51, Springer, 2013.

26. S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, “A fistful of bitcoins: characterizing payments among men with no names,” in Proceedings of the 2013 conference on Internet measurement conference, pp. 127–140, ACM, 2013.

27. D. Ron and A. Shamir, “Quantitative analysis of the full bitcoin transaction graph,” in International Conference on Financial Cryptography and Data Security, pp. 6–24, Springer, 2013.

28. P. Koshy, D. Koshy, and P. McDaniel, “An analysis of anonymity in bitcoin using p2p network traffic,” in International Conference on Financial Cryptography and Data Security, pp. 469–485, Springer, 2014.

29. M. Fleder, M. S. Kester, and S. Pillai, “Bitcoin transaction graph analysis,” arXiv preprint arXiv:1502.01657, 2015.

30. S. Barber, X. Boyen, E. Shi, and E. Uzun, “Bitter to better-how to make bitcoin a better currency,” in Financial cryptography and data security, pp. 399–414, Springer, 2012.

31. E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, “Zerocash: Decentralized anonymous payments from bitcoin,” in Security and Privacy (SP), 2014 IEEE Symposium on, IEEE, 2014.

32. S. Meiklejohn and C. Orlandi, “Privacy-enhancing overlays in bitcoin,” in International Conference on Financial Cryptography and Data Security, pp. 127–141, Springer, 2015.

33. Oleg Voytenko. Review of modern analytical services for blockchain // X International Scientific and Practical Conference. International Forum: Actual trends of modern scientific research . May 9-11. – 2021. – P. 157-164.

34. R. C. Merkle, “A digital signature based on a conventional encryption function,” in Advances in Cryptology — CRYPTO ’87: Proceedings. Springer Berlin Heidelberg, 1988, pp. 369–378.

35. O.G. Kachko, D. Televnyi. The kupyna hash function cryptanalysis with merkle tress signature schemes USA Telecommunications and Radio Engineering, 2019. – pages 683-689 DOI: 10.1615/ TelecomRadEng. v78.i8.40.

36. Буч Г. UML. Руководство пользователя / Г. Буч, – М.: ДМК Пресс; Издание 2-е, стер., 2014. – 432 с.

37. Kachko O., N. Bilous , Semerkov V. Research on methods for secure web applications development Information Technolo-gies in Innovation Busi-ness (ITIB), 7-9

38. Фаулер М. Архітектура програмних додатків. – Київ: Знання, 2007. – 315 с.

39. A. Yerokhin, A. Nechyporenko, O. Turuta, A. Babii. A new intelligence-based approach for rhinomanometric data processing 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), Kiev, 2016, pp. 198-201. doi: 10.1109/ELNANO.2016.7493047.sdf

40. Shubin, I. , Karmanenko, O., Gorbach, T., Umyarov, K. The methods of adaptation in computer-based training systems Shubin, I. , Karmanenko, O., Gorbach, T., Umyarov, K.

2015 Information Technologies in Innovation Business Conference, ITIB 2015 - Proceedings, 2015, с. 64-67, 7355054

41. R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in Proceedings of the 23rd international conference on Machine learning, pp. 161–168, ACM, 2006.

42. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.

43. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 463–484, 2012.

44. Гаевский, А.Ю. 100% самоучитель. Создание Web-страниц и Web-сайтов. HTML и JavaScript / А.Ю. Гаевский, В.А. Романовский. - М.: Триумф, 2014. - 464 с.

45. Дакетт, Джон HTML и CSS. Разработка и дизайн веб-сайтов (+ CD-ROM) / Джон Дакетт. - М.: Эксмо, 2013. - 480 с.