УДК 004.056.53:004.75

А. Р. Магденко
А. В. Грицак

КОМПЛЕКСНИЙ АЛГОРИТМ ПРОФІЛЮВАННЯ КІБЕРЗЛОЧИНЦІВ МЕТОДАМИ ГРАФОВОЇ ТА БЛОКЧЕЙН-АНАЛІТИКИ

Вінницький національний технічний університет

Анотація

У представлених тезах досліджується та обґрунтовується інноваційна алгоритмічна модель автоматизованого збору, багаторівневої нормалізації та глибокого крос-платформного аналізу цифрових слідів, спрямована на комплексну деанонімізацію суб'єктів сучасних кіберзагроз. Розроблено гібридний аналітичний конвеєр, що інтегрує методи обробки природної мови (зокрема, розпізнавання іменованих сутностей та векторне представлення слів) для семантичного розбору неструктурованого тексту з тіньових форумів та закритих месенджерів. Окрему увагу приділено застосуванню апарату теорії зважених мультиграфів для математичного обчислення транзитивних зв'язків між ізольованими комунікаційними профілями, а також імплементації евристичних алгоритмів кластеризації UTXO-транзакцій у блокчейн-мережах для деобфускації фінансових потоків. Практичне застосування запропонованої математичної моделі забезпечує перехід від фрагментованих індикаторів компрометації до формування юридично значущих комплексних профілів кіберзлочинців. Це створює науково обґрунтований базис для подальшого легального блокування незаконних віртуальних активів на централізованих криптовалютних біржах та ініціювання кримінальних проваджень у межах міжнародної правоохоронної співпраці.

Ключові слова: кібербезпека, деанонімізація, теорія графів, блокчейн-аналітика, OSINT, обробка природної мови, кіберзлочинність, кластеризація транзакцій, тіньова економіка.

Abstracts

The presented abstracts explore and substantiate an innovative algorithmic model for automated collection, multi-level normalisation, and deep cross-platform analysis of digital traces, aimed at comprehensive de-anonymisation of modern cyber threats. A hybrid analytical pipeline has been developed that integrates natural language processing methods (in particular, named entity recognition and vector representation of words) for semantic parsing of unstructured text from shadow forums and closed messengers. Particular attention is paid to the application of weighted multigraph theory for the mathematical calculation of transitive links between isolated communication profiles, as well as the implementation of heuristic algorithms for clustering UTXO transactions in blockchain networks for the deobfuscation of financial flows. The practical application of the proposed mathematical model provides a transition from fragmented indicators of compromise to the formation of legally significant comprehensive profiles of cybercriminals. This creates a scientifically sound basis for the further legal blocking of illegal virtual assets on centralised cryptocurrency exchanges and the initiation of criminal proceedings within the framework of international law enforcement cooperation.

Keywords: *cybersecurity, de-anonymisation, graph theory, blockchain analytics, OSINT, natural language processing, cybercrime, transaction clustering, shadow economy.*

Вступ

Актуальність даного дослідження зумовлена стрімкою еволюцією транснаціональної кіберзлочинності та її переходом до високоорганізованої моделі надання тіньових послуг. Сучасна екосистема кіберзагроз характеризується безпрецедентним рівнем технологічної сегментації та інфраструктурної обфускації. Зловмисники масово використовують стратегію жорсткого розподілу операційних процесів між різними рівнями мережі Інтернет з метою унеможливлення відстеження їхньої діяльності єдиним аналітичним центром. Зокрема, використання публічних месенджерів з наскрізним шифруванням, таких як Telegram, відіграє роль так званої «верхньої частини воронки продажів» для рекрутингу та координації, тоді як бекенд-інфраструктура, панелі управління шкідливим програмним забезпеченням та системи ескроу ізолюються в оверлейних мережах типу Tor. Одночасно фінансовий кровообіг цієї екосистеми забезпечується виключно через децентралізовані криптографічні протоколи із застосуванням багаторівневих ланцюжків відмивання коштів та сервісів мікшування. Класичні парадигми цифрової криміналістики та інструменти розвідки на основі відкритих джерел (OSINT), які спираються на ручний або напівавтоматизований аналіз ізольованих

масивів даних, демонструють критичну неефективність в умовах такої крос-платформної фрагментації. Об'єктом даного дослідження виступає процес розслідування інцидентів інформаційної безпеки в гетерогенних мережевих середовищах. Предметом дослідження є алгоритмічні та математичні моделі агрегації, кореляції та аналізу розрізнених цифрових слідів. Головною метою роботи є презентація розробленого комплексного гібридного алгоритму, який шляхом синергії лінгвістичного аналізу, стохастичного графового моделювання та фінансових блокчейн-евристик дозволяє автоматизовано долати технологічні бар'єри між платформами, синтезуючи ізольовані цифрові артефакти у єдиний, доказовий профіль кіберзагрози.

Результати дослідження

Проблематика технологічної ізоляції цифрових слідів. Фундаментальною проблемою, з якою сьогодні стикаються підрозділи кіберполіції та фахівці з Threat Intelligence, є аналітичний парадокс, викликаний асиметрією між обсягом доступних даних та здатністю їх корелювати. Захист сучасної тіньової інфраструктури базується не лише на анонімності мережевого рівня, але й на активній протидії автоматизованому збору інформації (Anti-Scraping Evasion). Легальні криптобіржі, тіньові маркетплейси та веб-клієнти месенджерів масово впроваджують системи аналізу поведінкових аномалій та фільтрації на рівні транспортних протоколів. Це означає, що спроби прямого вивантаження даних блокуються сучасними Web Application Firewalls (WAF), які аналізують TLS-фінгерпринти та параметри браузерних рушіїв. Внаслідок цього слідчі змушені працювати з фрагментарними, неповними «зрізами» інформації. Проте навіть у випадку успішного подолання бар'єрів доступу, галузь стикається з проблемою інформаційних «бункерів» (data silos). Парсери соціальних мереж та блокчейн-експлорери існують у паралельних просторах: лінгвістичний контекст переговорів у Telegram ніяк не пов'язаний із потоком нерозтрачених виходів (UTXO) у мережі Bitcoin [1].

Ситуація критично ускладнюється явищем комбінаторного вибуху під час спроб відстеження фінансових потоків. Зловмисники автоматизують процеси відмивання капіталу за допомогою алгоритмічного смурфінгу (формування ланцюжків відшаровування – Peel Chains), дроблячи великі суми вкрадених активів на тисячі мікротранзакцій для обходу AML-моніторингу на централізованих біржах. Крім того, активне використання смарт-контрактів з нульовим розголошенням (наприклад, Tornado Cash) та протоколів мікшування повністю руйнує детерміновану криптографічну лінійність транзакційного графа. У таких умовах спроби прямого, ручного зіставлення ізольованих індикаторів компрометації (IoC) – наприклад, пошук однакових псевдонімів або часткових збігів криптовалютних адрес – не мають жодної математичної доказової бази та призводять до критичного рівня хибнопозитивних спрацьовувань. Галузь кібербезпеки терміново потребує переходу від лінійного моніторингу до створення автономних систем крос-платформної ідентифікації. Необхідні нові архітектурні рішення, здатні автоматично розраховувати ймовірнісні (м'які) зв'язки між сутностями з різних світів, спираючись на непрямі ознаки: стилометрію, часові патерни активності та стохастичний аналіз обсягів ліквідності, перетворюючи розрізнений цифровий шум на структуровану розвідувальну інформацію [2].

Для вирішення фундаментальної проблеми технологічної сегментації цифрових слідів у даному дослідженні розроблено та обґрунтовано комплексну архітектурну модель конвеєра даних (Data Pipeline), яка функціонує на стику методів обробки природної мови (NLP), теорії графів та блокчейн-аналітики. Першим етапом роботи запропонованого алгоритму є багаторівнева нормалізація та екстракція іменованих сутностей (Entity Extraction) із гетерогенного масиву неструктурованого тексту, зібраного з тіньових форумів та закритих Telegram-каналів. Для подолання умисної лінгвістичної обфускації та специфічного транснаціонального кібер-сленгу алгоритм використовує попередньо натреновані великі мовні моделі архітектури Transformer (зокрема, BERT та RoBERTa), які здійснюють семантичний розбір тексту на основі механізму внутрішньої уваги (Self-Attention). Це дозволяє системі контекстуально ідентифікувати ролі зловмисників, назви шкідливого програмного забезпечення та фінансові реквізити. Паралельно застосовується детермінований модуль на базі регулярних виразів (RegEx) із вбудованою криптографічною валідацією контрольних сум, що гарантує безпомилкове вилучення PGP-ключів та криптовалютних адрес різних стандартів кодування (Base58, Bech32). Для вирішення проблеми гомогліфічних атак та варіативності написання нікнеймів імплементовано математичний алгоритм нечіткого пошуку рядків, який обчислює редакційну відстань Левенштейна у комбінації з фонетичною нормалізацією. У результаті роботи цього етапу система генерує криптографічно чистий, стандартизований простір ізольованих індикаторів компрометації [3].

Наступним, концептуально найважливішим етапом розробленого алгоритму є перехід від лінійного аналізу до просторового графового моделювання. Усі вилучені та верифіковані сутності (комунікаційні

профілі, фінансові гаманці, технічні індикатори) проєктуються у вигляді вершин (вузлів) у структурі багатовимірного зваженого мультиграфа G = (V, E, W). Оскільки прямого зв'язку між екосистемами DarkWeb та публічними месенджерами не існує, алгоритм генерує імовірнісні (м'які) ребра на основі непрямих поведінкових кореляцій. Для цього застосовується математична модель стилометрії, яка трансформує тексти повідомлень з різних платформ у багатовимірні вектори та розраховує косинусну подібність (Cosine Similarity) між ними, виявляючи єдиного автора за специфічними граматичними аномаліями. Паралельно алгоритм часової кореляції аналізує мітки часу (timestamps) активності профілів, вибудовуючи хронобіологічні теплові карти та розраховуючи статистичну кореляцію Пірсона для ідентифікації синхронних «вікон тиші» (циркадних ритмів). Кожному виявленому збігу присвоюється відповідна математична вага (від 0.0 до 1.0). Для ухвалення фінального рішення щодо злиття двох ізольованих профілів в єдину особу, графова база даних автоматично агрегує ваги всіх паралельних ребер за формулою ймовірності об'єднання незалежних подій $W_{total} = 1 - \Pi_{i=1}^{n}(1 - w_i)$. Цей підхід дозволяє перетворити сукупність слабких непрямих доказів на високий композитний індекс довіри, математично доводячи крос-платформну ідентичність зловмисника [4].

Заключним вектором роботи гібридного алгоритму є деобфускація фінансових інфраструктур за допомогою глибокого аналізу орієнтованих ациклічних графів (DAG) у блокчейн-мережах. З метою деанонімізації багатоланцюжкових транзакцій система імплементує комплекс детермінованих евристик, базовою з яких є евристика спільного використання входів (Common Input Ownership Heuristic, CIOH) та алгоритм ідентифікації тіньової адреси для здачі. Їхня синергія дозволяє автоматично стягувати сотні тисяч одноразових UTXO-адрес у єдиний фінансовий кластер злочинного угруповання. Для протидії алгоритмічному смурфінгу та відстеження ланцюжків відшаровування (Peel Chains), які зловмисники використовують для виведення коштів на легальні біржі (VASP), розроблений модуль використовує модифікований пошук у глибину (DFS), налаштований на розпізнавання асиметричних патернів розподілу вартості. У випадках використання кіберзлочинцями протоколів мікшування, де лінійність графа руйнується, алгоритм здійснює динамічне перемикання на стохастичний аналіз потоків (Stochastic Flow Analysis), шукаючи кореляції обсягів ліквідності у заданих часових вікнах для відновлення розірваних маршрутів. У фіналі роботи конвеєра результати графового моделювання комунікацій та результати блокчейн-кластеризації автоматизовано зливаються у новий концептуальний об'єкт – Мета-Вузол. Цей об'єкт являє собою комплексний цифровий паспорт кіберзагрози, що містить математично доведену ієрархію соціальних зв'язків, точні фінансові обороти у фіатному еквіваленті та геолокаційні маркери, формуючи вичерпну доказову базу для проведення оперативно-розшукових заходів [5].

Оцінюючи розроблену алгоритмічну модель, доцільно наголосити на концептуальному зсуві, який пропонується для сучасної галузі кібербезпеки. Традиційно цифрова криміналістика звикла реагувати на інциденти постфактум, збираючи розрізнені артефакти, коли кошти вже виведені, а інфраструктура знищена. Проте справжня наукова та практична цінність даного дослідження полягає у зміщенні фокусу з реактивного розслідування на проактивну деанонімізацію. Фундаментальною метою проєктування системи стало максимальне усунення людського фактора з процесу пошуку неочевидних крос-платформних зв'язків. В епоху масового використання криптографічних міксерів та оверлейних мереж ручний аналіз тисяч транзакційних маршрутів та стилометричних патернів втрачає будь-яку ефективність. Інтеграція модулів обробки природної мови, зважених графових баз даних та фінансових евристик кластеризації у єдиний безперервний конвеєр створює надійний алгоритмічний міст між абсолютно різними вимірами цифрових слідів. Завдяки такому підходу математично доведено, що ілюзію анонімності в тіньовому сегменті мережі можна зруйнувати автоматизовано, перетворюючи розрізнений інформаційний шум на юридично значущу доказову базу. Очікується, що саме такі гібридні, багатовимірні аналітичні системи стануть новим стандартом для Threat Intelligence, оскільки вони здатні ефективно нівелювати головну зброю зловмисників – технологічну сегментацію.

Висновки

Підсумовуючи результати проведеного дослідження та проєктування гібридної аналітичної системи, можна констатувати успішне створення фундаментально нової алгоритмічної парадигми для подолання інфраструктурної обфускації у сучасних комп'ютерних мережах. Головним науково-технічним здобутком роботи є практичне доведення того факту, що проблема крос-платформної фрагментації цифрових слідів може бути ефективно вирішена виключно шляхом синергетичного злиття детермінованих імовірнісних методів машинного навчання, графового моделювання та блокчейн-аналітики. Розроблений алгоритм успішно трансформує концептуально несумісні масиви інформації – неструктуровану семантику природної мови з тіньових форумів та криптографічно

жорсткі UTXO-транзакції – у єдиний, математично вимірюваний багатовимірний простір ознак. Ключовою інновацією запропонованої моделі є відмова від пошуку прямих ідентифікаційних збігів на користь агрегації «м'яких» поведінкових аномалій. Завдяки імплементації стохастичного аналізу потоків, оцінки стилометричних векторів та розрахунку композитних індексів довіри на основі теорії зважених мультиграфів, алгоритм набув властивості самостійно генерувати транзитивні зв'язки навіть в умовах навмисного переривання інформаційних ланцюжків протоколами мікшування. Впровадження механізмів динамічного відсікання хибних гілок (Pruning) та розрахунку коефіцієнта «забрудненості» вузлів дозволило успішно нівелювати загрозу комбінаторного вибуху, забезпечивши високу обчислювальну ефективність системи при обробці великих даних (Big Data). У результаті, безперервна робота цього алгоритмічного конвеєра математично доводить вразливість концепції абсолютної анонімності в гетерогенних мережах: навіть за умови ідеального дотримання операційної безпеки (OPSEC) окремо на кожній платформі, загальний поведінковий та фінансовий патерн зловмисника неминуче кристалізується в єдиний комплексний Мета-Вузол. Сформована таким чином архітектура не лише автоматизує процес деанонімізації суб'єктів кіберзагроз, але й формує надійний, масштабований науковий фундамент для подальшого розвитку інтелектуальних систем проактивного моніторингу та захисту інформації в умовах глобалізованої тіньової економіки.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Проблематика технологічної ізоляції цифрових слідів. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/IOCTA%202023%20-%20EN.pdf (дата звернення: 09.03.2026).

2. Peel chains, laundering, AML, мікшування. URL: https://www.chainalysis.com/blog/2024-crypto-crime-report-introduction/ (дата звернення: 09.03.2026).

3. NLP, BERT/RoBERTa. URL: https://arxiv.org/abs/1810.04805 (дата звернення: 09.03.2026).

4. Графові моделі, мережевий аналіз. URL: https://www.cambridge.org/core/journals/network-science/all-issues (дата звернення: 09.03.2026).

5. Блокчейн-аналітика, кластеризація адрес, UTXO. URL: https://www.researchgate.net/publication/349335348_Detecting_and_Quantifying_Wash_Trading_on_Decentralized_Cryptocurrency_Exchanges (дата звернення: 09.03.2026).

Магденко Анастасія Романівна – студентка групи 1КІТС-22б, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: anastasiiamahdenko@gmail.com

Науковий керівник: ***Грицак Анатолій Васильович –*** доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com

Mahdenko Anastasiia R. – student of group 1KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: anastasiiamahdenko@gmail.com

Supervisor*:* ***Hrytsak Anatoly V. –*** Associate Professor of the Department of Management and Security of Information Systems Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com