КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.29.933
УДК 004.056.3+004.056.5

Фесьоха Віталій Вікторович
доктор філософії, доцент, докторант
Військовий інститут телекомунікацій та інформатизації
імені Героїв Крут, Київ, Україна
ORCID ID: 0000-0001-6612-1970
vitaliifesokha@gmail.com

Субач Ігор Юрійович
доктор технічних наук, професор, завідувач кафедри
Інститут спеціального зв'язку та захисту інформації
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
ORCID ID: 0000-0002-9344-713X
igor_subach@ukr.net

МЕТОД ВИЯВЛЕННЯ ЗАКОНОМІРНОСТЕЙ ЕВОЛЮЦІЇ ТЕХНІК КІБЕРАТАК НА ОСНОВІ ТОПОЛОГІЧНОГО АНАЛІЗУ ДАНИХ

Анотація. У контексті підвищення кіберстійкості спеціальних інформаційно-комунікаційних систем (ІКС) розглядається завдання дослідження еволюції кібератак, зумовленого зростанням їхньої динамічності та непередбачуваності. Показано, що існуючі підходи (темпоральні графи, оцінка поширеності технік, аналіз ланцюжків, детекція змін, порівняння версій таксономій) здебільшого фіксують статистичні та послідовні аспекти і не виявляють прихованих інваріантів і структурних зав'язків між техніками кібератак. Запропоновано метод на основі топологічного аналізу даних, який моделює техніки у спільному інтерпретованому просторі структурних, графових і семантичних ознак (без нейромережевої компресії) та використовує зважену косинусну відстань для побудови симпліційних комплексів В'єторіса–Ріпса і динамічних персистентних діаграм з узгодженим міжверсійним зіставленням класів гомологій. Передбачено нормування порогів та уніфікацію ознак для міжверсійної порівнянності; критерії і параметри фіксуються, що забезпечує відтворюваність результатів і незалежність висновків від вибору масштабу. Метод забезпечує виявлення двох типів закономірностей: топологічних тенденцій (зміни зв'язності, фрагментації та циклічності простору технік у часі) і траєкторних інваріантів – ланцюгів класів гомологій, що без розривів відстежуються через версії за формальними критеріями покриття та персистентності. Демонстраційний аналіз таксономії MITRE ATT&CK (версії 14.1–17.0) виявив режимність змін із виразними перебудовами на мажорних релізах та тенденцію до фрагментації простору технік; інтегральні інваріанти інтерпретуються як стабільні межі/контури між кластерами технік і можуть бути використані для прогнозування змін та планування оновлень аналітики, реалізуючи принцип еволюційності кіберстійкості.

Ключові слова: кіберстійкість; штучний інтелект; інформаційно-комунікаційна система; закономірності; еволюція кібератак; топологічний аналіз даних.

ВСТУП

Останнім часом характер кібератак істотно трансформувався, що проявляється зростанням динамічності та непередбачуваності змін у їхніх тактиках і техніках реалізації [1]. Внаслідок цього попередні уявлення про загрози досить швидко втрачають актуальність, а засновані на них рішення щодо забезпечення кіберстійкості об'єктів критичної інформаційної інфраструктури стають менш ефективними.

Постановка проблеми. На практиці протидію кібератакам, що еволюціонують, здійснюють в межах безперервного циклу спостереження, аналізу та реагування, де

провідну роль відіграє проактивний захист ІКС, орієнтований на їх попередження ще до етапу реалізації [2]. Однак відсутність регламентованих процедур ідентифікації та інтеграції шаблонів розвитку кібератак у контури моніторингу, аналізу та планування кіберзахисту ІКС значно ускладнює завчасне розпізнавання змін у структурі та векторах кібератак, переходах між техніками і послідовностями ескалації. Відтак недостатня увага до тривалих тенденцій і закономірностей розвитку призводить до несвоєчасного оновлення моделей виявлення кібератак, помилкового визначення пріоритетів, нераціонального розподілу ресурсів, а також зниження результативності заходів із забезпечення кіберстійкості ІКС.

Зазначене актуалізує необхідність проведення подальших наукових досліджень, спрямованих на системне виявлення та фіксацію неочевидних тенденцій і структурних зсувів у практиках реалізації кібератак, відокремлення стійких еволюційних закономірностей від короткочасних флуктуацій та подання результатів у формі, придатній для підтримки прийняття рішень.

Аналіз останніх досліджень і публікацій. Основою більшості досліджень підходів до реалізації кібератак є формалізовані таксономії їхніх технік і тактик, наприклад, набори даних MITRE ATT&CK [3], суміжні класифікатори шаблонів кібератак, такі як CAPEC [4], а також каталоги відповідних контрзаходів, наприклад, D3FEND [5]. Огляд цієї групи наукових праць виділяє кілька основних напрямів досліджень [6]–[21]:

- ***побудова темпоральних графів кібератак за матеріалами аналітичних звітів*** [6]–[9] — перетворення звітів на формальні подієво-хронологічні структури, такі як послідовності та графи кроків кібератак, що співвідносяться з [3] для автоматичного відтворення хронології кібератак з тексту, а також отримання узгоджених ланцюжків/графів подій із позначенням фаз і станів системи;
- ***дослідження поширеності технік у різні періоди*** [10]–[12] — кількісне вимірювання хронологічної спостережуваності технік з офіційних наборів даних про кібератаки та виокремлення стабільного ядра поведінки противника з подальшою трансляцією отриманих ознак у практичні пріоритети;
- ***послідовний аналіз шаблонів кібератак*** [13], [14] — взаємодоповнювані підходи до аналізу «ланцюжків» у межах таксономії [3], де генеративний підхід моделює послідовності марковськими моделями (тактики — приховані стани, техніки — спостереження), тоді як описовий ґрунтується на процесному аналізі журналів подій із подальшим узгодженням їх із таксономією [3], що дає змогу і передбачати можливі наступні кроки та відтворювати фактичні шаблони;
- ***виявлення точок змін у практиках зловмисників*** [15]–[17] — поєднує аналіз часових рядів агрегованих показників згідно з [3], що дає змогу фіксувати переломні моменти в поширені технік кібератак із виявленням концептуального дрейфу, що у свою чергу пояснює окремі нові або модифіковані зразки;
- ***порівняння версій таксономій*** [18]–[21] — порівняння редакцій офіційних класифікаторів кібератак для автоматизованого вирівнювання ідентифікаторів і підрахунку доданих/змінених/видалених об’єктів з урахуванням структурних змін таксономії (нові/об’єднані/поділені та перейменовані сутності), що забезпечує коректні хронологічні зіставлення.

Наведений огляд існуючих напрямків дослідження розвитку кібератак свідчить про значний науковий прогрес у їх моделюванні та класифікації. Однак, переважна більшість методів орієнтована на статистичний та послідовний аналіз, що накладає суттєві обмеження на розуміння повноти картини реалізації кібератак. Описуються очевидні

аспекти кібератак (частотність технік і типові переходи між ними), тоді як поза увагою залишаються приховані інваріанти та взаємодії, зокрема такі, які не усвідомлюються навіть самими зловмисниками. Як наслідок, існуючі методи виявляються недостатніми для виявлення глибинних структурних зав'язків, топології взаємозалежностей між техніками кібератак та якісних трансформацій у тактиці противника.

У зв'язку з цим, для подолання зазначених обмежень щодо дослідження природи розвитку кібератак запропоновано застосування топологічного аналізу даних (ТАД), здатного досліджувати глибинну структуру даних для виявлення глобальних, інваріантних форм і зв'язків у високорозмірних і зашумлених даних [22]. Ефективність ТАД вже доведена в суміжних тематиках з кібербезпеки, зокрема в задачах виявлення аномалій та шкідливих програм, аналізу графів взаємодій та обробки неструктурованих даних журналів активності [23]–[25]. До того ж, застосування ТАД надасть змогу реалізувати принцип еволюційності, запропонований в [2] для підвищення кіберстійкості спеціальних ІКС, що значно підсилить їхній проактивний захист.

Метою статті є розробка методу виявлення закономірностей еволюції технік кібератак на основі ТАД.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для фіксації еволюційних закономірностей кібератак доцільно здійснювати їх аналіз на рівні технік, а не тактик, оскільки техніки є операційними, безпосередньо фіксуються в телеметрії (відображають поведінку зловмисної активності у вигляді конкретних прийомів реалізації) та прямо пов'язані з джерелами даних і контрзаходами, тоді як тактики надто агреговані, що значною мірою згладжує тонкі перебудови послідовностей. У даному контексті виявлення багатомасштабних, стійких до шуму структур у темпоральних мережах і послідовностях даних здійснюємо засобами ТАД, спираючись на таксономію [3] як стандартизовану основу для уніфікованого подання, інтерпретації та міжверсійного зіставлення технік кібератак.

Вихідні дані:

$V = \{v_1, \dots, v_i, \dots, v_n\}$ — множина версій технік кібератак з [3] у форматі STIX 2.1;

$T(v_i) = \{t_{1i}, \dots, t_{ji}, \dots, t_{mi}\}$ — множина технік кібератак у $v_i \in V$.

Обмеження і допущення:

- аналізується ядро опису технік кібератак (ідентифікатор, назва, опис, платформи, передумови/привілеї, пов'язані джерела даних), поля Detection і Mitigation виключено з ознакового простору з метою уникнення внесення захисних процедур в еволюційні патерни технік;
- відповідність технік між v_i та v_{i+1} здійснюється за Mitre_id;
- обрана множина версій технік кібератак 14.1–17.1: 14.1–17.0 для аналізу;
- еволюційною закономірністю (траєкторним інваріантом) r вважаємо ланцюг гомологічних класів однієї розмірності, що без розривів відстежується у v_i (для локальних на більш ніж 2 послідовних v_i з персистентністю не нижче 75%, для глобальних — із покриттям усіх мажорних релізів і персистентністю не нижче 75%).

Необхідно: виявити наступні закономірності еволюції t_{ji}:

- топологічні тенденції: оцінити й описати зміни форми простору технік через версії, зокрема режимність оновлень, еволюцію зв'язності (H_0: пороги

злиття, частка найбільшої компоненти) і циклічності (H_1: поява/зникнення та довжина стійких циклів);
- траєкторні інваріанти: виявити в H_0 та H_1: ланцюги класів, що відповідають локальним і глобальним закономірностям).

Метод виявлення закономірностей еволюції технік кібератак на основі ТАД передбачає реалізацію наступних етапів:

1. Підготовка вихідних даних

Оскільки методи ТАД працюють виключно з числовими просторами [26] то усі вихідні атрибути перетворюються на числа. З метою реалізації зазначеного передбачено формування блокового представлення ознак технік кібератак:

1. Структурний блок акумулює інтерпретовані атрибути самої техніки: належність до тактик (kill-chain phases) і платформ, статус підтехніки, зв'язок із батьківською технікою. Усі категоріальні поля кодуються як one-hot або multi-hot вектори з бінарними координатами {0, 1}; назва кожної координати однозначно ідентифікує сутність (наприклад, s::tactic.Execution, s::platform.Windows).
2. Графовий блок відображає положення техніки у межах кожної версії. Будується ненаправлений граф: вершини — техніки (включно з підтехніками); ребра — відношення і зв'язки за спільною тактикою або платформою.
3. Семантичний блок отримується з текстових полів найменувань, описів та псевдонімів: на об'єднанні всіх версій складається спільний словник, після чого тексти векторизуються за TF-IDF з l_2-нормуванням (кожна ознака — конкретний термін словника).

Вибір таких представлень ознак обумовлено інтерпретацією і відтворюваністю результатів ТАД. Так, структурні та графові ознаки описують прямий зв'язок «координата → атрибут/зв'язок» у [3], що дає змогу після виявлення топологічних структур назвати конкретні тактики, платформи чи ребра, які сформували спостережувану топологію. Для семантичних ознак використано лінійний TF–IDF у спільному словнику, що на відміну від нейромережевих векторів забезпечує легкий контроль над блоком ознак.

На цій основі значення блоків ознак зводяться до спільного простору (X, ρ), як зважені відстані для порівнянного внеску у топологію даних за принципом медіанного вирівнювання. Після стандартизації значень (центрування та масштабування за стандартним відхиленням) відстань між двома точками i, j визначаємо згідно (1).

$$d(i,j) = \alpha\delta_E(i,j) + \beta\delta_S(i,j) + \gamma\delta_G(i,j), \quad \alpha,\beta,\gamma \geq 0, \alpha+\beta+\gamma = 1, \tag{1}$$

де $\delta_E, \delta_S, \delta_G$ — косинусна відстань для семантичних, структурних і графових ознак, α, β, γ — вагові коефіцієнти.

Завершальним кроком етапу є побудова попарної матриці відстаней між усіма точками на основі косинусної відстані, оскільки вона ефективно поєднує індикаторні і TF-IDF-координати та є стійкою до різної розмірності блоків. Результатом підготовки вихідних даних є матриця відстаней $D = \left[d_{ij}\right]_{N\times N}$ разом із супровідними таблицями метаданих та каталогом ознак.

Таким чином, після проведених розрахунків техніку кібератак можна представити аналітичним виразом (2).

$$t_i = \left\{id, parent\ (\perp), S_{t_i}, G_{t_i}, E_{t_i}\right\}. \tag{2}$$

2. Побудова симпліційного комплексу

Отримана на попередньому етапі матриця D подається на вхід алгоритмам побудови симпліційного комплексу В'єторіса–Ріпса як способу апроксимації форми даних про техніки кібератак у метричному просторі X, індукованому D [27] (3).

ISSN 2663 - 4023

$$VR_\varepsilon = \left\{\sigma \subseteq X : \max_{a<b} p\left(x_{i_a}, x_{i_b}\right) \leq \varepsilon\right\}, \quad (3)$$

де $p\left(x_i, x_j\right)$ — відстань між точками x_i і x_j (елемент матриці відстаней D), ε — поріг (масштаб), при якому будується комплекс, a, b — локальні індекси вершин симплекса.

Таким чином, побудова симпліційного комплексу полягає у побудові ε-графу для фіксованого $\varepsilon > 0$: вершини x_i і x_j з'єднуються ребром, якщо $D_{ij} \leq \varepsilon$. Будь-яка підмножина вершин, усі попарні відстані між якими $\leq \varepsilon$, утворює симплекс (вершина, ребро, трикутник). На рис. 1 проілюстровано побудову симпліційного комплексу із утворених симплексів на множині технік кібератак версії 14.1, де кожна точка — техніка.

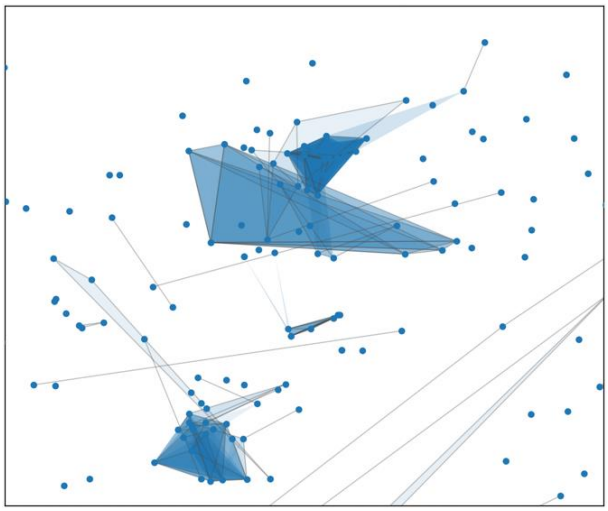

Рис. 1. Побудова симпліційного комплексу на множині технік кібератак версії 14.1

3. Фільтрація

На цьому етапі з матриці відстаней D утворюється впорядкована за порогом ε послідовність вкладених комплексів $VR_\varepsilon(X)$. Для фіксованого ε у комплекс входить підмножина вершин σ тоді й лише тоді, коли всі попарні відстані між ними не перевищують ε, тобто комплекс зростає: якщо $\varepsilon_1 \leq \varepsilon_2$, то $VR_{\varepsilon_1}(X) \subseteq VR_{\varepsilon_2}(X)$.

Таким чином, відбувається параметризація шкали через квантилі відстаней у кожній версії, що забезпечує порівняння технік кібератак між версіями при відносно однаковій щільності комплексу. На рис. 2 проілюстровано утворення послідовності вкладених комплексів $VR_\varepsilon(X)$ на множині перетворених даних технік кібератак версії 14.1 при $\varepsilon = 0$, $\varepsilon \approx 0.1$, $\varepsilon \approx 0.2$, $\varepsilon = 0.3$.

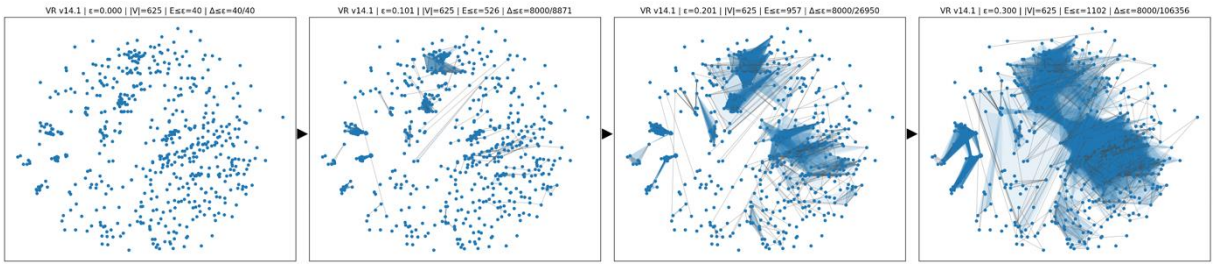

Рис. 2. Утворення послідовності вкладених комплексів $VR_\varepsilon(X)$ на множині даних технік кібератак версії 14.1 при $\varepsilon = 0$, $\varepsilon \approx 0.1$, $\varepsilon \approx 0.2$, $\varepsilon = 0.3$

4. Обчислення персистентних гомологій

Даний етап призначено для відстеження гомологічних класів (H_0, H_1, H_2) крізь масштаби: при зростанні порога ε на етапі фільтрації K_ε індукуються нові симплекси, внаслідок чого класи гомології з'являються (народжуються) та зникають (помирають). Саме ці інтервали життя і становлять персистентну гомологію.

Класи гомології (практичний зміст для технік кібератак):

1. H_0 — компоненти зв'язності: відображають скупчення технік (кластеризацію) при певній щільності; зникнення класу відповідає злиттю двох кластерів (поява/зникнення топологічних островів і їхні злиття).
2. H_1 — цикли (нетривіальні контури, петлі): їх поява означає альтернативні шляхи/варіанти технік кібератак навколо порожнього центру; зникнення — момент, коли контур заповнюється симплексами, що пояснює сімейства близьких варіантів із відсутньою комбінацією ознак у центрі.
3. H_2 — порожнини (двовимірні порожні області, що заповнюються тетраедрами): інтерпретуються як стійкі поєднання великих підмножин ознак, які в базовій конфігурації, як правило, не обчислюється.

Для кожного рівня ε обчислюється гомологія $H_k(K_\varepsilon)$ над полем $\mathbb{F}_2$. Далі відстежується еволюція H_k у міру зростання ε і для кожного класу фіксується пара «народження — зникнення» (b_i, d_i) на основі (4).

$$(H_k(K_\varepsilon), \iota) \cong \bigoplus_i I_{[b_i, d_i]}, \tag{4}$$

де K_ε — комплекс на рівні порогу ε, ι — індуковані включення, які задають відображення $H_k(K_\varepsilon) \to H_k(K_{\varepsilon'})$, $I_{[b_i,d_i]}$ — інтервальний модуль.

Сукупність зазначених пар утворює діаграму персистентності $P_k = \{(b_i, d_i)\}$ і відповідний штрих-код, а кількість «живих» класів у момент ε дорівнює числу Бетті $\beta_k(\varepsilon)$. На рис. 3 представлено діаграму гомологічного класу H_0 на 60-ти техніках кібератак версії 17.0 з найбільшим життям, де кожна горизонтальна смуга — окремий компонент зв'язності (H_0, кластер) у комплексі В'єторіса–Ріпса.

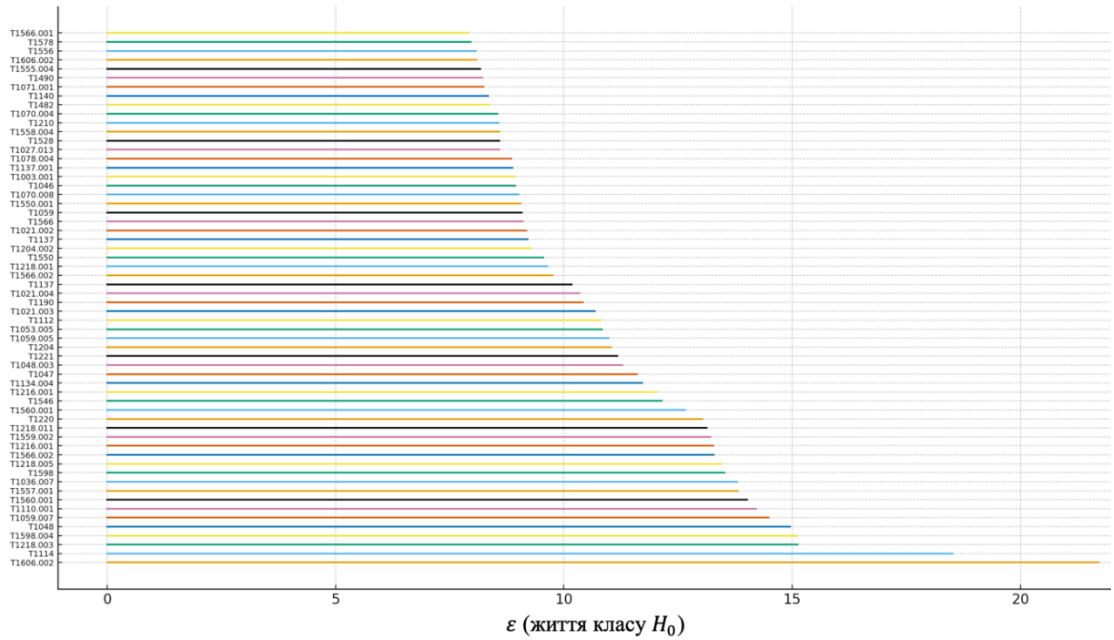

Рис. 3. Діаграма гомологічного класу H_0 на 60-ти техніках кібератак версії 17.0

У комплексі В'єторіса–Ріпса усі вершини присутні вже за $\varepsilon = 0$, тому для H_0 класи народжуються при $b_i = 0$ і зникають у момент d_i, коли відповідний компонент зв'язності зливається зі старшим (вага ребра, що з'єднало компоненти). Для H_1 народження відбувається на найменшому ε, за якого з'являється нетривіальний цикл, тоді як зникнення відбувається тоді, коли контур заповнюється симплексами. Кожному класу відповідає інтервал $[b_i, d_i]$, у фільтрації, а його довжина $d_i - b_i$ є персистентністю, що використовується як міра значущості.

На рис. 4 представлено фрагмент дендрограми послідовності злиття технік кібератак версії 17.0, де висота злиття на дендрограмі — зникнення відповідного класу.

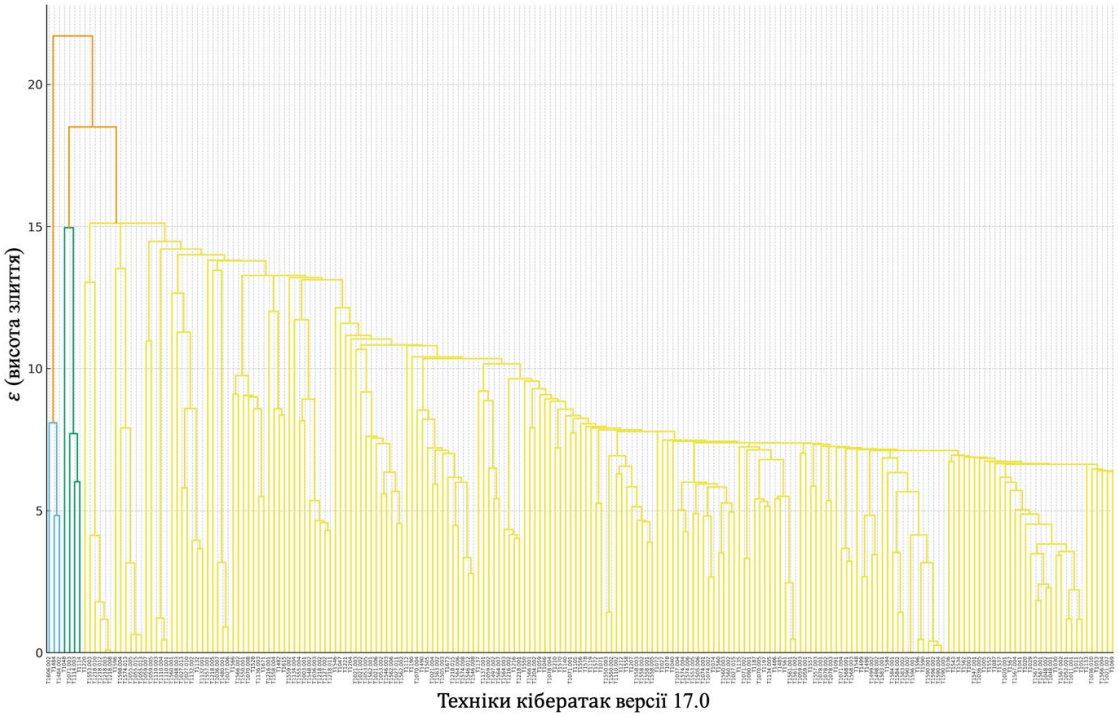

Рис. 4. Фрагмент дендрограми послідовності злиття технік кібератак версії 17.0

5. *Формування динамічних персистентних діаграм і визначення топологічних тенденцій*

Емпіричну динаміку технік кібератак розглядаємо як послідовність форм простору ознак у версіях $v_1, ..., v_n$. Для кожної версії v_j обчислюється персистентна діаграма $P_k(v_j) = \{(b_i^j, d_i^j)\}$, тобто множина пар «поява — зникнення» для класів k-го виміру. Для порівняння рівнів ε між версіями використовуємо нормовану шкалу: $\varepsilon_{v_j}(q) = Q_{v_j}(q)$, де Q_{v_j} — q квантиль розподілу попарних відстаней у версії v_j.

Між сусідніми версіями $v_j \to v_{j+1}$ будується оптимальне зіставлення точок діаграм як бієкція $m_j: \tilde{P}_j \leftrightarrows \tilde{P}_{j+1}$, що мінімізує найбільше зміщення точки за зіставлення: (5):

$$m_j \in \arg\min_{\pi} \max_{x \in \tilde{P}_j} d_\infty\left(x, \pi(x)\right),\ d_\infty\big((b,d),(b',d')\big) = max\{|b-b'|, |d-d'|\}, \quad (5)$$

де $x \in \tilde{P}_j$ — точка інтервалу життя класу в розширеній діаграмі для v_j за координатами, π — бієкція, m_j — обране оптимальне зіставлення між $\tilde{P}_j$ і $\tilde{P}_{j+1}$, $\arg\min_{\pi}$ — множина всіх бієкцій π, що мінімізують вказану функцію втрат, $\max_{x \in \tilde{P}_j}$ —

вартість зіставлення, що оцінюється за найбільшим зміщенням серед усіх точок $\tilde{P}_j$, d_∞ — метрика на площині «поява — зникнення».

Звідси, динамічна персистентна діаграма V k-го виміру — множина траєкторій, отриманих послідовним застосуванням m_j (6):

$$V_k = \{(x_1, \dots, x_n): x_1 \in \tilde{P}_1,\ x_{j+1} = m_j(x_j) \text{ для всіх } j\}. \quad (6)$$

На рис. 5, 6 представлено динамічні персистентні діаграми, на яких для кожної версії технік кібератак уздовж ε відстежуються класи гомологій у нормованій шкалі [0,1]. По H_0 траєкторія відображає межу компонент зв'язності (коли кластери зливаються), по H_1 — контури (цикли), які існують, доки діри не заповняться.

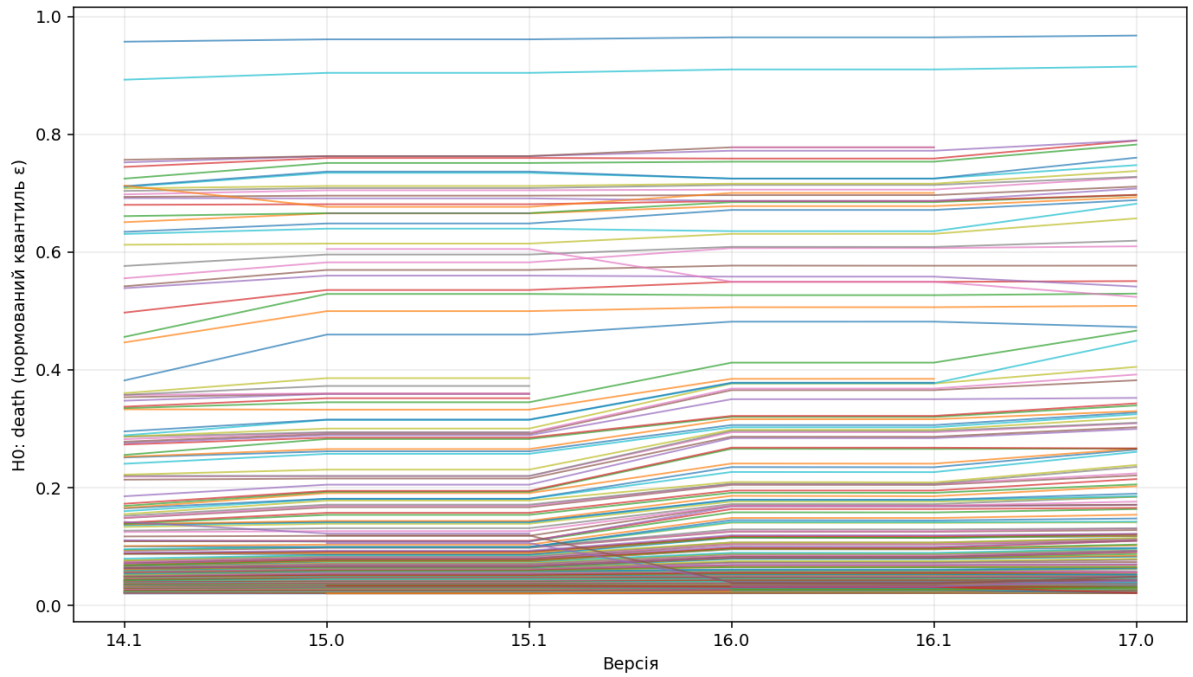

Рис. 5. Динамічна персистентна діаграма H_0 між версіями 14.1 – 17.0 технік кібератак

Всього 238 траєкторій. З діаграми видно такі топологічні тенденції:

1. Низькопорогова консолідація. Більшість злиттів відбувається при малих ε (нижня частина шкали): значна частина ліній зникає досить рано (у вузькому нижньому інтервалі ε), що означає швидке стягування дрібних груп технік у кілька більших кластерів.
2. Підйоми на мажорних релізах. На переходах 14.1 → 15.0, 15.1 → 16.0 та 16.1 → 17.0 помітно, що на значній частині траєкторій підвищується рівень ε (тягнуться вище), тобто межі між кластерами стають стійкішими (для їх злиття потрібно більший поріг). Це узгоджується домінуючими подіями, коли один гомологічний клас у v_j має кілька нащадків у версіїv_{j+1} (1067 для H_0), що вказує на фрагментацію простору, коли великі групи технік розділяються на кілька тем із більш чіткими межами.
3. Довгі інваріантні межі. Невелика частина ліній проходить крізь усі версії технік і зберігає високу персистентність, що формує опорні роздільні лінії

простору, які не руйнуються при оновленнях. Саме вони є претендентами на глобальні закономірності H_0.

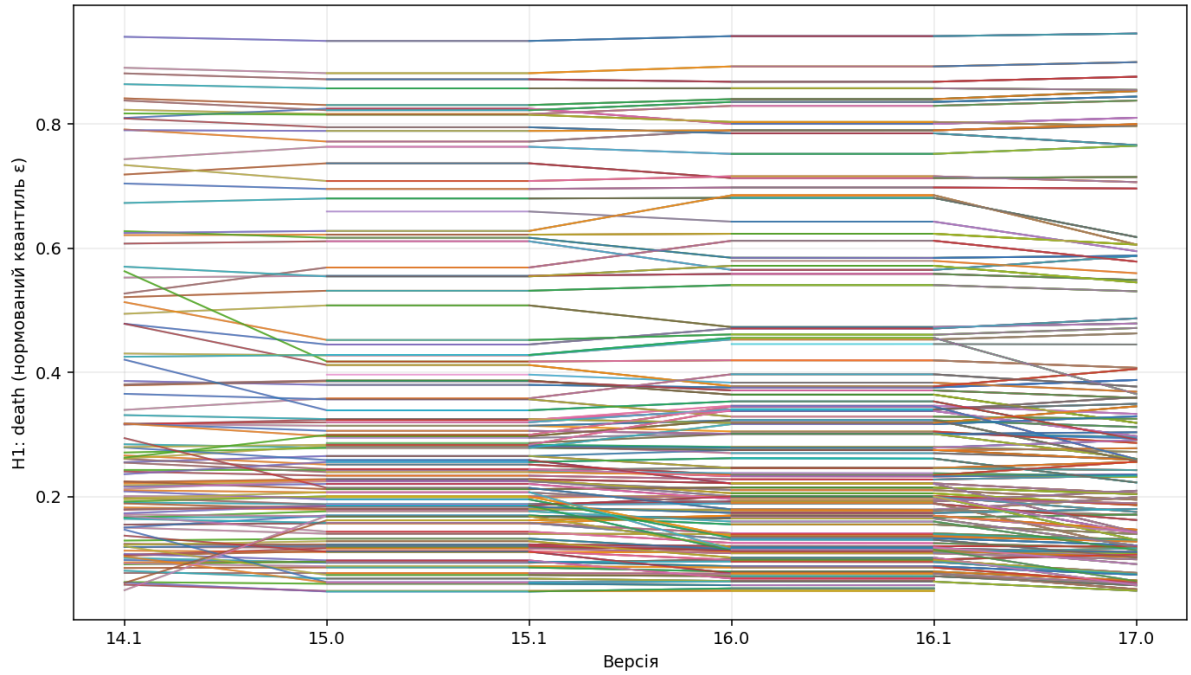

Рис. 6. Динамічна персистентна діаграма H_1 між версіями 14.1 – 17.0 технік кібератак

Траєкторій — 719. З діаграми прослідковуються наступні топологічні тенденції:

1. Короткоіснуючі контури. Переважають короткі траєкторії, тобто більшість циклів виникає й зникає при невеликих значеннях $\boldsymbol{\varepsilon}$. Це означає, що сімейства варіантів конфігурацій технік короткоживучі: контури локально з'являються, але швидко зникають.
2. Імпульси на мажорних релізах. На «.0» випусках сплески появ циклів. Сукупно між версіями домінують події, коли один гомологічний клас у v_j має кілька нащадків у версії v_{j+1} (усього 2364), що відповідає перебудові локальних варіантів, тоді як на «.1» частіше спостерігається розгалуження (менше довгоживучих контурів).
3. Відсутність довгих глобальних контурів. Обрані критерії безперервності не пропускають глобальних кандидатів H_1, тобто ці структури епізодичні та чутливі до змін описів/зв'язків.

6. Визначення траєкторних інваріантів

На попередньому етапі побудовано динамічні персистентні діаграми з 238 траєкторіями H_0 та 719 траєкторіями H_1.

З метою перетворення отриманих результатів в траєкторні інваріанти необхідно визначити порядок зіставлення гомологічних класів між сусідніми версіями. Так, класичне представлення динамічної персистентної діаграми опирається на відповідність один-до-одного (безрозривне простежування тієї самої сутності) [26], [27], що не дозволяє в повній мірі дослідити ланцюги різноманітних подій у таксономії технік кібератак, оскільки на практиці масово трапляються їх розгалуження та/або злиття. У

зв'язку з цим, запропоновано розширене зіставлення класів між версіями на основі доповнення відношень один-до-багатьох (розгалуження) і багато-до-одного (злиття).

Позначимо через $P_{v_j}^k$ множину точок діаграми персистентності розмірності k для версії v_j (кожна точка — клас із нормованою персистентністю [0,1]). Для точки $x \in P_{v_j}^k$ визначимо множину можливих нащадків у v_{j+1} (7).

$$S_{v_j}(x) = \{y \in P_{v_{j+1}}^k : d(x,y) \le \tau\}, \quad (7)$$

де $d(x,y)$ — відстань зіставлення точок на діаграмах, τ — малий поріг близькості.

Тоді події класифікуємо наступним чином:

- $S_{v_j}(x) = 1 \; – \; 1 \to 1$: безрозривне відстежування того самого класу (*продовження/посилення/послаблення)*;
- $S_{v_j}(x) = 1 – 1 \to 0$: клас зникає;
- $S_{v_j}(x) = 1 – 1 \to k$: від одного класу в v_j відходить кілька нащадків у v_{j+1};
- $S_{v_j}(x) = 1 – k \to 1$: злиття;
- $S_{v_j}(x) = 1 – 0 \to 1$: народження.

Таким чином, традиційний підхід відповідає лише $1 \to 1$-шляху, а запропонований підхід працює з подійним графом між версіями, у якому $1 \to 1$-дуги формують траєкторії, а $1 \to k/k \to 1$ фіксують фрагментацію/агрегацію топологічного простору.

Траєкторія-кандидат $\mathcal{T}$ на траєкторний інваріант — максимальний нерозривний ланцюг класів між сусідніми версіями. Для $\mathcal{T}$ обчислюємо:

- $len(\mathcal{T})$ — кількість переходів між версіями;
- покриття $cov(\mathcal{T}) =$ кількість покритих v_j/V;
- медіанна персистентність $pers(\mathcal{T}) \in [0,1]$,
- індикатор проходження всіх «.0»: $\text{major}(\mathcal{T}) \in \{0,1\}$.

Кандидат $\mathcal{T}$ вважається еволюційною закономірністю тоді й лише тоді, коли виконується:

локальна закономірність: $len(\mathcal{T}) \ge 2 \; \wedge \; pers(\mathcal{T}) \ge 0.75$;

глобальна закономірність: $cov(\mathcal{T}) = 1 \; \wedge \; \text{major}(\mathcal{T}) = 1 \; \wedge \; pers(\mathcal{T}) \ge 0.75$.

В аналізованих даних між версіями домінують розгалуження: H_0 — 1067, H_1 — 2364, що узгоджується з візуальною фрагментацією на «.0»-релізах.

Результати відбору закономірностей:

1. З H_0-діаграми за наведеними критеріями відібрано 10 глобальних та 6 локальних закономірностей. Глобальні (стійкі через 14.1→17.0) мають повне покриття версій, проходять усі «.0» і демонструють стабільний 1↔1-тренд із високою $pers(\mathcal{T}) \approx 0.74$. Локальні — довгі безрозривні інтервали зі стабільною або помірно змінною персистентністю.
2. Для H_1 при аналогічних порогах кандидатів не відібрано: цикли здебільшого короткоживучі, а треки перериваються подіями розгалуження/злиття, що підтверджує імпульсний характер контурів.

Приклади виявлених глобальних закономірностей:

1. Інваріантна межа на середньому порозі при $v = 17.0$, $\varepsilon \approx 0.767$, техніки T1588.004, T1585.001 залишаються відокремленими аж до високого ε. Межа повторюється у всіх версіях, що формує опорну границю простору. Висновок: межу доцільно фіксувати як стабільний кластерний розріз: правила/евристики для технік доцільно підтримувати окремо в релізах «.0».

2. Стабільне розгалуження двох сімейств розвідки (v = 17.0, $\varepsilon \approx 0.731$), техніки T1593, T1597. Кластери, що знаходяться на T1593 та T1597, не знижуються до $\varepsilon \approx 0.73$ і так поводяться в усіх версіях, що формує надійний інваріант. Висновок: підходить як контрольний розріз для моніторингу. Якщо у наступному «.0» техніки обєднаються при істотно меншому ε — це сигнал перебудови домену розвідки.
3. Жорстка межа між великою групою та ядровою технікою (v = 17.0, $\varepsilon \approx 0.962$). Велика група: T1588.007, T1583.005, T1586.003, T1586.002, T1584.005, T1584.006, T1608.002, T1584.008, T1583.004, T1584.004, ядрова техніка T1586.001. Жорстка стіна тримається між ними у всіх версіях, що формує сильний глобальний інваріант. Висновок: будь-яке зближення T1586.001 до групи великої групи в новій «.0» дорівнює події, що варта перегляду правил.

Приклади виявлених локальних закономірностей (на рівні версій):

1. Локальна копія (v = 17.0, $\varepsilon \approx 0.767$), техніки T1588.004, T1585.001.Підсилює попередні висновки щодо інваріантної межі: проявляється і на локальному рівні. Висновок: доцільно використовувати як додатковий засіб валідації.
2. Локальна копія (v = 17.0, $\varepsilon \approx 0.731$), техніки T1593, T1597. Підтверджує стабільність розрізу у домені розвідки на локальному шарі.
3. Локальна копія (v = 17.0, $\varepsilon \approx 0.962$), велика група технік: T1588.007, T1583.005, T1586.003, T1586.002, T1584.005, T1584.006, T1608.002 і ядрова техніка T1586.001. Дублює жорстку межу й спрощує операційне спостереження за нею у версіях.

Таким чином, запропоновано метод виявлення закономірностей еволюції технік кібератак на основі топологічного аналізу даних, який моделює їхню хронологічну динаміку у спільному просторі структурних, графових і семантичних ознак та виокремлює топологічні тенденції і траєкторні інваріанти, що реалізує принцип еволюційності кіберстійкості.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

В умовах розбудови національної системи кібербезпеки України критично важливим є набуття спроможності швидко адаптуватися до внутрішніх і зовнішніх загроз та оперативно відновлювати стале функціонування об’єктів критичної інформаційної інфраструктури. За таких умов традиційні превентивні, реактивні та постінцидентні підходи виявляються недостатніми, а деструктивні впливи на критично важливі ІКС часто залишаються непоміченими до настання реальних наслідків.

У статті запропоновано метод виявлення закономірностей еволюції технік кібератак на основі топологічного аналізу даних із інтерпретованим багатоблоковим простором ознак (структурні, графові, семантичні), зваженою косинусною відстанню, побудовою комплексів В’єторіса–Ріпса та динамічних персистентних діаграм з узгодженим міжверсійним зіставленням класів. Метод формує два результати: топологічні тенденції (динаміка зв’язності, фрагментації та циклічності простору технік) та траєкторні інваріанти (стійкі ланцюги гомологій), що формалізують і фіксують еволюцію.

Демонстрація на MITRE ATT&CK (версії 14.1–17.0) виявила режимність змін із виразними перебудовами на мажорних релізах і тенденцію до фрагментації простору

технік; стабільні інваріантні межі/контури інтерпретуються як опорні межі між кластерами технік. Практично це надає ранні індикатори (зсув порогів злиття, руйнування інваріантів), допомагає пріоритезувати оновлення правил, планувати рев'ю аналітики під мажорні випуски таксономій і підтримує принцип еволюційності кіберстійкості. Перевага підходу — відтворюваність і пояснюваність (без нейромережевої компресії), а також узагальненість на будь-які версійні домени знань.

Таким чином, запропонований метод значно підсилює системне розуміння динаміки технік кібератак і створює основу для проактивних рішень та прогнозування змін у критично важливих ІКС.

Перспективним напрямком подальших наукових досліджень є розробка моделі прогнозування векторів розвитку технік кібератак за результатами топологічного аналізу їх еволюції.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Фесьоха, В. (2024). Особливості протистояння оборонного та наступального штучного інтелектів у кіберпросторі. *International Science Journal of Engineering & Agriculture*, *3*(4), 105–114. https://doi.org/10.46299/j.isjea.20240304.11.
2. Фесьоха, В., & Субач, І. (2025). КОНЦЕПТУАЛЬНА ОСНОВА ПІДВИЩЕННЯ КІБЕРСТІЙКОСТІ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМ В УМОВАХ ЕВОЛЮЦІЇ КІБЕРЗАГРОЗ. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, *4*(28), 511–528. https://doi.org/10.28925/2663-4023.2025.28.856.
3. The MITRE Corporation. *MITRE ATT&CK®*. Retrieved September 15, 2025, from https://attack.mitre.org.
4. The MITRE Corporation. *CAPEC™: Common attack pattern enumeration and classification*. Retrieved September 15, 2025, from https://capec.mitre.org/.
5. The MITRE Corporation. *MITRE D3FEND™*. Retrieved September 15, 2025, from https://d3fend.mitre.org/.
6. Rahman, M. R., Wroblewski, B., Matthews, Q., Morgan, B., Menzies, T., & Williams, L. (2024). *Mining temporal attack patterns from cyberthreat intelligence reports* [Preprint]. arXiv. https://arxiv.org/abs/2401.01883.
7. Zhang, Y., Du, T., Ma, Y., Wang, X., Xie, Y., Yang, G., Lu, Y., & Chang, E.-C. (2025). AttacKG+: Boosting attack graph construction with Large Language Models. *Computers & Security*, *150*, 104220. https://doi.org/10.1016/j.cose.2024.104220.
8. Nadeem, A., Verwer, S. E., Moskal, S., & Yang, S. J. (2022). Alert-driven attack graph generation using S-PDFA. *IEEE Transactions on Dependable and Secure Computing*, *19*(2), 731–746.
9. Ahmadou, F., Ghaffarzadegan, S., Nour, B., Pourzandi, M., Debbabi, M., & Assi, C. (2025). Automated attack testflow extraction from cyber threat report using BERT for contextual analysis. *arXiv*. https://arxiv.org/abs/2507.07244.
10. Abo-alian, A., Youssef, M., & Badr, N. L. (2025). A data-driven approach to prioritize MITRE ATT&CK techniques for Active Directory adversary emulation. *Scientific Reports*, *15*, 27776. https://doi.org/10.1038/s41598-025-12948-x.
11. Center for Threat-Informed Defense. (2021). *Sightings Ecosystem: A data-driven analysis of ATT&CK in the wild* (Report No. CT0039). MITRE Engenuity. https://ctid.mitre-engenuity.org/.
12. Fetterman, R., & Chacon, T. (2024, October 10). *Macro-ATT&CK 2024: A five-year perspective*. Splunk Blog. https://www.splunk.com/en_us/blog/security/macro-att-ck-2024-a-five-year-perspective.html.
13. Rodríguez, M., Betarte, G., & Calegari, D. (2024). A process mining-based method for attacker profiling using the MITRE ATT&CK taxonomy. *Journal of Internet Services and Applications*, *15*(1), Article 1. https://doi.org/10.5753/jisa.2023.3902.
14. Choi, S., Yun, J.-H., & Min, B.-G. (2021). Probabilistic attack sequence generation and execution based on MITRE ATT&CK for ICS datasets. In *Proceedings of the Cyber Security Experimentation and Test Workshop (CSET '21)* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3474718.3474722.
15. Maffia, L., Nisi, D., Kotzias, P., Lagorio, G., Aonzo, S., & Balzarotti, D. (2021). *Longitudinal study of the prevalence of malware evasive techniques*. arXiv. https://arxiv.org/abs/2112.11289.

16. Yang, L., Guo, W., Hao, Q., Ciptadi, A., Ahmadzadeh, A., Xing, X., & Wang, G. (2021). CADE: Detecting and explaining concept drift samples for security applications. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security '21)*. USENIX Association.
17. Landauer, M., Skopik, F., Stojanović, B., Flatscher, A., & Ullrich, T. (2025). A review of time-series analysis for cyber security analytics: From intrusion detection to attack prediction. *International Journal of Information Security, 24*(3). https://doi.org/10.1007/s10207-024-00921-0.
18. CTI Butler. (2024, November 4). *An analysis of the changes in ATT&CK version 16.0*. CTI Butler Blog. https://www.ctibutler.com/blog/analysis_mitre_attack_16/.
19. Pennington, A., & Ajmo, J. (2022, April 25). *ATT&CK goes to v11: Structured detections, beta sub-techniques for mobile, and ICS joins the band.* MITRE ATT&CK Blog (Medium). https://medium.com/mitre-attack/att-ck-goes-to-v11-599a9112a025.
20. Pennington, A., & Burns, J. (2020, October 27). *Bringing PRE into Enterprise*. MITRE ATT&CK Blog (Medium). https://medium.com/mitre-attack/the-retirement-of-pre-attack-4b73ffecd3d3.
21. MITRE. (2025, April). *Updates: MITRE ATT&CK* (Version 17). Retrieved September 9, 2025, from https://attack.mitre.org/resources/updates/.
22. Kindelan, R., Frías, J., Cerda, M., & Hitschfeld, N. (2021). *Classification based on Topological Data Analysis*. arXiv. https://arxiv.org/abs/2102.03709.
23. Davies, T. (2022). *A review of topological data analysis for cybersecurity*. arXiv. https://arxiv.org/abs/2202.08037.
24. Tidjon, L. N., & Khomh, F. (2022). *Reliable malware analysis and detection using topology data analysis*. arXiv. https://arxiv.org/abs/2211.01535.
25. Bihl, T. J., Gutierrez, R. J., Bauer, K. W., Boehmke, B. C., & Saie, C. (2020). Topological data analysis for enhancing embedded analytics for enterprise cyber log analysis and forensics. *Proceedings of the 53rd Hawaii International Conference on System Sciences,* 1937–1946. https://doi.org/10.24251/HICSS.2020.238.
26. Carlsson, E., Carlsson, G., & de Silva, V. (2006). An algebraic topological method for feature identification. *International Journal of Computational Geometry & Applications, 16*(4), 291–314. https://doi.org/10.1142/S021819590600204X.
27. Attali, D., Lieutier, A., & Salinas, D. (2013). Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, *46*(4), 448–465. https://doi.org/10.1016/j.comgeo.2012.02.009.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023

Vitalii Fesokha
PhD in Information systems and technologies, Associate Professor, Postdoctoral researcher
Kruty Heroes Military Institute of Telecommunications and Information
Technologies, Kyiv, Ukraine
ORCID ID: 0000-0001-6612-1970
vitaliifesokha@gmail.com

Ihor Subach
Doctor of Technical Science, Professor, Head of the Department
Institute of Special Communications and Information Protection
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
ORCID ID: 0000-0002-9344-713X
igor_subach@ukr.net

METHOD FOR DETECTING PATTERNS IN THE EVOLUTION OF CYBERATTACK TECHNIQUES BASED ON TOPOLOGICAL DATA ANALYSIS

Abstract. In the context of improving the cyber resilience of special information and communication systems (ICS), the task of studying the evolution of cyber attacks, caused by their increasing dynamism and unpredictability, is considered. It is shown that existing approaches (temporal graphs, assessment of technique prevalence, chain analysis, change detection, comparison of taxonomy versions) mostly capture statistical and sequential aspects and do not reveal hidden invariants and structural relationships between cyberattack techniques. A method based on topological data analysis is proposed, which models techniques in a common interpreted space of structural, graph and semantic features (without neural network compression) and uses weighted cosine distance to construct Vietoris–Rips simplicial complexes and dynamic persistent diagrams with consistent cross-version comparison of homology classes. Thresholds are standardized and characteristics are unified for inter-version comparability; criteria and parameters are fixed, ensuring reproducibility of results and independence of conclusions from the choice of scale. The method identifies two types of patterns: topological trends (changes in connectivity, fragmentation, and cyclicality of the technique space over time) and trajectory invariants—chains of homology classes that are tracked seamlessly across versions using formal coverage and persistence criteria. A demonstration analysis of the MITRE ATT&CK taxonomy (versions 14.1–17.0) revealed a pattern of change with significant restructuring in major releases and a tendency towards fragmentation of the technique space. Integral invariants are interpreted as stable boundaries/contours between clusters of techniques and can be used to predict changes and plan analytics updates, implementing the principle of cyber resilience evolution.

Keywords: cyber resilience; artificial intelligence; information and communication system; patterns; evolution of cyber attacks; topological data analysis.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Fesokha,V. (2024). Features of the confrontation between defensive and offensive artificial intelligence in cyberspace.International Science Journal of Engineering & Agriculture,3(4),105–114.https://doi.org/10.46299/j.isjea.20240304.11.
2. Fesokha, V., & Subach, I. (2025). CONCEPTUAL FRAMEWORK FOR ENHANCING CYBER RESILIENCE OF INFORMATION AND COMMUNICATION SYSTEMS IN THE CONTEXT OF EVOLUTION OF CYBER THREATS. Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique», 4(28), 511–528. https://doi.org/10.28925/2663-4023.2025.28.856.
3. The MITRE Corporation. *MITRE ATT&CK*®. Retrieved September 15, 2025, from https://attack.mitre.org.
4. The MITRE Corporation. *CAPEC™: Common attack pattern enumeration and classification*. Retrieved September 15, 2025, from https://capec.mitre.org/.
5. The MITRE Corporation. *MITRE D3FEND™*. Retrieved September 15, 2025, from https://d3fend.mitre.org/.
6. Rahman, M. R., Wroblewski, B., Matthews, Q., Morgan, B., Menzies, T., & Williams, L. (2024). *Mining temporal attack patterns from cyberthreat intelligence reports* [Preprint]. arXiv. https://arxiv.org/abs/2401.01883.

7. Zhang, Y., Du, T., Ma, Y., Wang, X., Xie, Y., Yang, G., Lu, Y., & Chang, E.-C. (2025). AttacKG+: Boosting attack graph construction with Large Language Models. *Computers & Security, 150*, 104220. https://doi.org/10.1016/j.cose.2024.104220.
8. Nadeem, A., Verwer, S. E., Moskal, S., & Yang, S. J. (2022). Alert-driven attack graph generation using S-PDFA. *IEEE Transactions on Dependable and Secure Computing, 19*(2), 731–746.
9. Ahmadou, F., Ghaffarzadegan, S., Nour, B., Pourzandi, M., Debbabi, M., & Assi, C. (2025). Automated attack testflow extraction from cyber threat report using BERT for contextual analysis. *arXiv*. https://arxiv.org/abs/2507.07244.
10. Abo-alian, A., Youssef, M., & Badr, N. L. (2025). A data-driven approach to prioritize MITRE ATT&CK techniques for Active Directory adversary emulation. *Scientific Reports, 15*, 27776. https://doi.org/10.1038/s41598-025-12948-x.
11. Center for Threat-Informed Defense. (2021). *Sightings Ecosystem: A data-driven analysis of ATT&CK in the wild* (Report No. CT0039). MITRE Engenuity. https://ctid.mitre-engenuity.org/.
12. Fetterman, R., & Chacon, T. (2024, October 10). *Macro-ATT&CK 2024: A five-year perspective*. Splunk Blog. https://www.splunk.com/en_us/blog/security/macro-att-ck-2024-a-five-year-perspective.html.
13. Rodríguez, M., Betarte, G., & Calegari, D. (2024). A process mining-based method for attacker profiling using the MITRE ATT&CK taxonomy. *Journal of Internet Services and Applications, 15*(1), Article 1. https://doi.org/10.5753/jisa.2023.3902.
14. Choi, S., Yun, J.-H., & Min, B.-G. (2021). Probabilistic attack sequence generation and execution based on MITRE ATT&CK for ICS datasets. In *Proceedings of the Cyber Security Experimentation and Test Workshop (CSET '21)* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3474718.3474722.
15. Maffia, L., Nisi, D., Kotzias, P., Lagorio, G., Aonzo, S., & Balzarotti, D. (2021). *Longitudinal study of the prevalence of malware evasive techniques*. arXiv. https://arxiv.org/abs/2112.11289.
16. Yang, L., Guo, W., Hao, Q., Ciptadi, A., Ahmadzadeh, A., Xing, X., & Wang, G. (2021). CADE: Detecting and explaining concept drift samples for security applications. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security '21)*. USENIX Association.
17. Landauer, M., Skopik, F., Stojanović, B., Flatscher, A., & Ullrich, T. (2025). A review of time-series analysis for cyber security analytics: From intrusion detection to attack prediction. *International Journal of Information Security, 24*(3). https://doi.org/10.1007/s10207-024-00921-0.
18. CTI Butler. (2024, November 4). *An analysis of the changes in ATT&CK version 16.0*. CTI Butler Blog. https://www.ctibutler.com/blog/analysis_mitre_attack_16/.
19. Pennington, A., & Ajmo, J. (2022, April 25). *ATT&CK goes to v11: Structured detections, beta sub-techniques for mobile, and ICS joins the band*. MITRE ATT&CK Blog (Medium). https://medium.com/mitre-attack/att-ck-goes-to-v11-599a9112a025.
20. Pennington, A., & Burns, J. (2020, October 27). *Bringing PRE into Enterprise*. MITRE ATT&CK Blog (Medium). https://medium.com/mitre-attack/the-retirement-of-pre-attack-4b73ffecd3d3.
21. MITRE. (2025, April). *Updates: MITRE ATT&CK* (Version 17). Retrieved September 9, 2025, from https://attack.mitre.org/resources/updates/.
22. Kindelan, R., Frías, J., Cerda, M., & Hitschfeld, N. (2021). *Classification based on Topological Data Analysis*. arXiv. https://arxiv.org/abs/2102.03709.
23. Davies, T. (2022). *A review of topological data analysis for cybersecurity*. arXiv. https://arxiv.org/abs/2202.08037.
24. Tidjon, L. N., & Khomh, F. (2022). *Reliable malware analysis and detection using topology data analysis*. arXiv. https://arxiv.org/abs/2211.01535.
25. Bihl, T. J., Gutierrez, R. J., Bauer, K. W., Boehmke, B. C., & Saie, C. (2020). Topological data analysis for enhancing embedded analytics for enterprise cyber log analysis and forensics. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 1937–1946. https://doi.org/10.24251/HICSS.2020.238.
26. Carlsson, E., Carlsson, G., & de Silva, V. (2006). An algebraic topological method for feature identification. *International Journal of Computational Geometry & Applications, 16*(4), 291–314. https://doi.org/10.1142/S021819590600204X.
27. Attali, D., Lieutier, A., & Salinas, D. (2013). Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry, 46*(4), 448–465. https://doi.org/10.1016/j.comgeo.2012.02.009.