DOI 10.34185/1562-9945-5-162-2026-17
УДК 004.056.53:004.89

П.В. Віжевський, О.С. Савенко

ПОВЕДІНКОВЕ ПРОФІЛЮВАННЯ КОРИСТУВАЧІВ ТА ВИЯВЛЕННЯ АНОМАЛІЙ НА ОСНОВІ ГЕНЕТИЧНОГО АЛГОРИТМУ

Анотація. У сучасних корпоративних середовищах поведінка користувачів і характеристики доступу до даних постійно змінюються, що ускладнює виявлення інсайдерських інцидентів і викликає нерівномірні хвилі хибних тривог у UEBA-системах. Через статичні пороги та надмірну розмірність профілі швидко втрачають актуальність за умов дрейфу концепції, а нестабільний потік тривог підвищує експлуатаційну вартість детектора. Метою роботи є розроблення методу поведінкового профілювання, який виявляє аномалії за відхиленнями від контекстуальної норми і є практично придатним завдяки контролю хибних спрацювань, стабільності сигналів і інтерпретованості. Запропоновано потокову формалізацію профілю з експоненційним забуванням та генетичний алгоритм для оптимізації конфігурації детектора (підмножини ознак, ваг, порога реагування і параметра адаптації) з урахуванням штрафів за хибні тривоги, волатильність алертів і складність моделі. Експерименти на CLUE-LDS із контрольованою зміною профілю показали помірне зростання F1 за одночасного зменшення хибнопозитивних тривог, зниження волатильності потоку тривог, зменшення середнього добового навантаження і скорочення складності до третини початкового набору ознак. Таким чином еволюційна конфігурація детектора задовільняє практичні експлуатаційні вимоги.
Ключові слова: поведінкове профілювання, UEBA, інсайдерські загрози, генетичний алгоритм, відбір ознак, дрейф концепції, хибні тривоги, стабільність тривог, потокові дані, корпоративна безпека.

Постановка проблеми. Системи корпоративної безпеки дедалі частіше функціонують в середовищах, де повсякденна активність користувачів не має стабільного профілю: змінюються ролі, команди, інструменти, канали взаємодії, а також способи доступу до даних. На цьому тлі інсайдерські інциденти особливо складні для виявлення, бо зловмисники маскуються під легітимних користувачів, а відхилення від норми зазвичай невеликі, розтягнуті в часі та розподілені між кількома типами подій. Через це детектори або пропускають повільні сценарії зловживань, або генерують постійні хибні тривоги, які потрібно фільтрувати вручну [1].

Поведінкове профілювання (UEBA) дозволяє не покладатися на набір статичних правил, однак воно приносить власний набір експлуатаційних проблем. Дані майже завжди незбалансовані: справжні інциденти рідкісні, а підозрілі епізоди трапляються

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

часто. Контекст має значну вагу у визначенні зловмисних дій, той самий обсяг копіювань або доступів може бути нормою для однієї ролі й тривожним сигналом для іншої. Для того щоб розслідування тривоги не перетворювалась на сліпий пошук, системі потрібна інтерпретованість, яка дозволить сформувати аналітику гіпотезу, що саме пішло не так [2].

Навіть якщо модель добре підганяється під історичний період, зміна процесів або політик доступу та мінливість поведінкових патернів швидко робить її пороги тривог неефективними. В такій ситуації якість знижується не тільки в точності, а й у стабільності сигналів. Нестабільність тривог не дозволяє використовувати систему у великих організаціях, бо її вартість визначається не лише помилками, а й коливанням потоку сповіщень та часом аналітика на перевірку [3].

Виникає потреба в методі, який цілеспрямовано оптимізує практичні властивості детектора. Важливим є механізм, що дозволяє підбирати компактний набір поведінкових ознак і їх ваги під задані обмеження, зменшуючи хибні спрацювання та підтримуючи керовану складність моделі. Еволюційні методи, зокрема генетичні алгоритми, природно підходять до такої задачі як інструмент пошуку компромісу в просторі конфігурацій між точністю, ціною та інтепретованістю [4].

Аналіз останніх досліджень і публікацій. Сучасні оглядові роботи з інсайдерської детекції узгоджуються в тому, що ключова складність полягає не в браку алгоритмів як таких, а в їхній експлуатаційній придатності на реальних даних і в реальних процесах реагування [1]. Якість алгоритму напряму пов’язана із вартістю помилок: в інсайдерських сценаріях зниження хибних тривог часто не менш цінне, ніж невелике зростання повноти, бо саме хибні тривоги визначають навантаження на аналітиків.

Потрібно враховувати часові патерни та контекстні залежності реальної роботи організації, комбінуючи різні типи сигналів, включно з поведінковими та текстовими [5]. Таким чином, для побудови поведінкового опису користувача використовують кілька типів наборів ознак (зведення активності, тематичні характеристики листування, мережі комунікацій) [6]. Комплексний погляд на одного користувача часто дає відчутну практичну користь, але розширення простору опису підвищує ризик того, що модель почне реагувати на шум або на короткочасні зміни процесів, продукуючи зайві сигнали. Це створює потребу в оптимізаційних методах, які здатні знаходити компактні підмножини ознак і визначати їх вагу.

Мультиступеневі пайплайни поєднують грубий етап звуження області інтересу та детальний аналіз потенційно підозрілих сегментів. Це зменшує обчислювальне навантаження й водночас допомагає уникати ситуації, коли система намагається глибоко аналізувати кожну дрібну подію в потоці. Це реалізується як послідовність рішень, що дозволяє керувати компромісом між точністю та обчислюваною складністю [4].

Дослідження, що сфокусовані на оптимізаційний методах, використовують еволюційні алгоритми як засіб для відбору ознак і налаштування порогів. Застосування генетичних алгоритмів до поведінково-орієнтованого виявлення аномалій демонструють навчання класифікатора, що ідентифікує поведінкові відхилення, і еволюційний пошук

здатний ефективно обирати комбінації ознак важливих для детекції [7]. Глобальний пошук в обмеженому просторі ознак усуває потребу у великих масивах міток.

У мережевих сценаріях генетичний алгоритм також використовується для налаштування підсистем збору та аналізу трафіку. В контексті програмно-визначених мереж (SDN) оптимізація ймовірностей вибірки трафіку дозволяє підвищити ефективність систем виявлення вторгнень, зменшуючи навантаження на контролер і не погіршуючи якість детекції [8].

Незалежно від вибору моделі визначення аномалій, використання генетичного алгоритму для відбору найбільш інформативних ознак знижує частки хибних спрацювань у порівнянні з базовими конфігураціями. Зокрема при побудові ансамбля дерев рішень Random Forest та XGBoost така синергія забезпечує покращення точності та зменшення кількості хибних тривог [9].

Детекція аномалій в Інтернеті речей демонструє спільні проблеми з профілюванням користувачів, брак даних із підтвердженим правдивими даними, різнорідність потоків подій, дрейф концепції та потребу в адаптивних методах, які можуть працювати без постійного нагляду аналітиків [10]. Еволюційні алгоритми розглядаються як один з інструментів, здатних коригувати роботу детектора в умовах зміни нормальної поведінки користувачів, контролюючи складність і кількість хибних тривог.

За умов дрейфу концепції оптимізувати в корпоративних системах доводиться не лише точність розпізнавання, а й операційні наслідки – керованість реакції, стійкість сигналів, витрати на обробку тривог, бюджет ресурсів [11]. Оскільки поведінкова норма в UEBA змінюється природно через організаційні процеси, без процедури контрольованого налаштування система швидко починає втрачати точність.

В сучасних стратегіях DLP, увага закономірно зміщується від одиничного алгоритму до цілісної архітектурної схеми: джерела сигналів, інтеграція з процесами реагування, обмеження на дані й розмітку, співіснування правил і моделей, а також питання підтримання якості в умовах змін [12]. І вартість хибних тривог визначається не як другорядна метрика, а системна характеристика, яка визначає придатність рішення до реальної експлуатації.

Дрейф концепції часто проявляється не як разовий збій моделі, а як послідовність переходів між режимами, де ознаки змін можуть бути слабкими, запізнілими або маскованими сезонністю й контекстом [13]. Для поведінкового профілювання детектор має або адаптуватися, або підтримувати стабільну політику консервативної реакції – і в обох випадках потрібні формальні механізми керування балансом чутливості та хибнопозитивних тривог.

Проблему дрейфу добре видно серед систем виявлення вторгнень, бо зміна трафіку, протоколів, інфраструктури та інструментів призводить до зміни інформативності ознак і деградації моделей. Надмірна залежність від великого, слабо контрольованого простору ознак підвищує ризик нестійких рішень. Модель починає описувати шум, що на практиці створює хвилі хибних тривог і надмірне навантаження на аналітиків [14]. В свою чергу, еволюційна оптимізація в завдяки керованому відбору підмножини ознак і

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

налаштування їх ваг, спрямована на стабілізацію поведінки детектора, а не на максимізацію абстрактної метрики.

Роботи з динамічної адаптації концептів в онлайнових аномалійних сценаріях показують, що практичний виграш часто досягається завдяки механізмам переузгодження моделі з потоком, але за ціною складніших процедур оновлення й контролю якості, що підкреслює необхідність в механізмах які дозволяють оновлювати уявлення про норму без неконтрольованої втрати стабільності [15].

Постановка завдання. Мета роботи полягає у побудові методу поведінкового профілювання користувачів, який виявляє аномалії на основі відхилень від контекстуальної норми та водночас залишається придатним до практичної експлуатації. Метод повинен забезпечувати мінімізацію кількості хибних тривог, зрозумілу логіку спрацювання і можливість адаптації до поступових змін поведінки та способів комунікації користувачів .

Для досягнення мети необхідно формально визначити поведінковий профіль як сукупність статистичних характеристик активності користувача у часових вікнах та задати міру відхилення поточної поведінки від базового коректного профілю. Потрібно розробити процедуру автоматизованого відбору ознак і корекції їх ваг, через те що надмірна розмірність погіршує узагальнення та ускладнює інтерпретацію. Також важливою є можливість контрольовано зміщувати баланс між чутливістю та хибними спрацюваннями механізму прийняття рішення про аномалію, що критично для практичного використання методу в реальних системах корпоративної безпеки.

Узгоджений механізм пошуку конфігурації детектора під експлуатаційні вимоги забезпечується використанням генетичного алгоритма для визначення складу ознак, їхніх ваг і порогів реагування. Головна увага приділяється компромісу між якістю виявлення, стабільністю сигналів і інтерпретованістю, тобто на ті характеристики, які визначають реальну цінність UEBA-детектора для DLP-середовищ

Експериментальна частина роботи має перевірити, у яких саме умовах та за якими показниками така еволюційно налаштована конфігурація дає практичний виграш.

Виклад основного матеріалу. Потік спостережень корпоративного середовища розглядається як сукупність елементарних подій, кожна з яких фіксує дію користувача в часі разом із контекстом.

$$e = (u, t, a, r, s), \tag{1}$$

де u – ідентифікатор користувача; t – часова мітка; a – тип дії; r – ресурс/об’єкт доступу; s – контекстні атрибути (зокрема робоча станція, IP, роль, канал).

Для побудови профілю поведінка агрегується у часові вікна тривалістю $\Delta > 0$. Множина подій користувача u у вікні з індексом k визначається як

$$E_{u,k} = \{ e : e = (u, t, a, r, s) \wedge k\Delta \leq t < (k+1)\Delta \}, \tag{2}$$

де $E_{u,k}$ – підмножина подій користувача u, часові мітки яких належать інтервалу $[k\Delta, (k+1)\Delta)$; Δ – тривалість вікна; k – індекс вікна.

На основі $E_{u,k}$ формується вектор ознак поведінки.

$$x_u(k) = \left(x_{u1}(k), x_{u2}(k), \dots, x_{um}(k)\right), \quad (3)$$

де $x_u(k) \in \mathbb{R}^m$ – вектор ознак користувача u у вікні k; m – кількість ознак; $x_{uj}(k)$ – значення j-ї ознаки. Конкретний склад ознак відповідає доступним журналам (автентифікація, файлові дії, мережеві з'єднання, доступ до сервісів) і повинен бути сумісним із UEBA-логами, на яких виконується оцінювання.

Щоб врахувати дрейф поведінки, для кожної ознаки підтримується «норма» у вигляді експоненційно зваженого середнього з коефіцієнтом забування $\rho \in (0,1)$.

$$\mu_{uj} = \rho\, \mu_{uj}(k-1) + (1-\rho)x_{uj}(k), \quad (4)$$

де $\mu_{uj}(k)$ – адаптивне середнє значення j-ї ознаки для користувача uу вікні k; ρ– коефіцієнт забування; $x_{uj}(k)$ – поточне значення ознаки.

Масштаб ознак нормалізується через адаптивну дисперсію.

$$v_{uj}(k) = \rho\, v_{uj}(k-1) + (1-\rho)\left(x_{uj}(k) - \mu_{uj}(k)\right)^2, \quad (5)$$

де $v_{uj}(k)$ – експоненційно зважена дисперсія j-ї ознаки; $(x_{uj}(k) - \mu_{uj}(k))^2$ – квадратичне відхилення..

Відхилення поточної поведінки від норми задається нормалізованою величиною.

$$\mathrm{d_{uj}}(k) = \frac{\left|x_{uj}(k) - \mu_{uj}(k)\right|}{\sqrt{v_{uj}(k)} + \varepsilon}, \quad (6)$$

де $d_{uj}(k) \geq 0$ – нормалізована міра відхилення; $\varepsilon > 0$ – мала константа для чисельної стійкості.

Аномальність поведінки у вікні описується зваженою сумою відхилень, але лише за тими ознаками, які активовані конфігурацією детектора. Це забезпечує керовану складність і підґрунтя для інтерпретації.

$$S_u(k) = \sum_{j=1}^{m} b_j w_j d_{uj}(k), \quad (7)$$

де $S_u(k)$– скалярний скор аномальності для користувача u у вікні k; $b_j \in \{0,1\}$– індикатор включення j-ї ознаки; $w_j \geq 0$ – вага ознаки.

Рішення про тривогу задається порогом τ.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

$$y_u(k) = I(S_u(k) \geq \tau), \quad (8)$$

де $y_u(k) \in \{0,1\}$ – вихід детектора (1 – тривога, 0 – норма); $I(\cdot)$ – індикаторна функція; τ – поріг спрацювання.

Конфігурація детектора кодується як хромосома, що одночасно визначає склад ознак, їх ваги та параметри адаптації.

$$C = (B, W, \tau, \rho), \quad (9)$$

де C – хромосома; $B = (b_1, \dots, b_m)$ – бінарна маска ознак; $W = (w_1, \dots, w_m)$ – вектор ваг; τ – поріг тривоги; ρ – коефіцієнт забування в профілі.

Складність моделі вимірюється часткою активних ознак:

$$Complexity(C) = \left(\frac{1}{m}\right)\sum_{j=1}^{m} b_j, \quad (10)$$

де $Complexity(C) \in [0,1]$ – нормована складність; m – кількість ознак.

Операційна стабільність сигналів фіксується через мінливість потоку тривог. Нехай U – множина користувачів, а $A(k)$ – кількість тривог у вікні k:

$$A(k) = \sum_{u \in U} y_u(k). \quad (11)$$

Для інтервалу оцінювання довжини T вводиться коефіцієнт варіації потоку тривог як міра нестабільності:

$$CV_{alert}(C) = \frac{\sigma(A)}{\bar{A} + \varepsilon}, \quad (12)$$

де $CV_{alert}(C)$ – відносна волатильність потоку тривог; $\bar{A}$ і $\sigma(A)$ – відповідно середнє та стандартне відхилення послідовності $\{A(1), \dots, A(T)\}$; $\varepsilon > 0$ – мала константа.

Функція пристосованості задає компроміс між якістю виявлення та експлуатаційними штрафами. Оскільки постановка роботи не передбачає «тотальної переваги» за всіма метриками, цільова функція прямо карає хибні тривоги, нестабільність сигналів і надмірну складність.

$$Fitness(C) = F1(C) - \alpha\, FPR(C) - \beta\, CV_{alert}(C) - \gamma\, Complexity(C), \quad (13)$$

де $Fitness(C)$ – значення пристосованості; $F1(C)$ – F1-міра, обчислена на валідаційному інтервалі для конфігурації C; $FPR(C)$ – частка хибнопозитивних спрацювань; $\alpha, \beta, \gamma \geq 0$ – коефіцієнти ваг штрафів, що задають політику компромісу залежно від вартості реагування в організації.

Інтерпретованість забезпечується тим, що скор є лінійною композицією внесків окремих ознак. Для кожної ознаки визначається внесок у конкретне спрацювання:

$$contrib_{uj}(k) = b_j w_j d_{uj}(k). \quad (14)$$

Експерименти виконано на публічному наборі журналів CLUE-LDS [2], який містить події реальних кількох тисяч анонімізованих користувачів. Для відтворюваного оцінювання аномальні події додаються штучно: після певного моменту події, що приписуються користувачу u, замінюються на статистично більше схожі на інший профіль. Така постановка добре узгоджується з прикладною UEBA-логікою: на практиці аналітика цікавить стійке зміщення звичного профілю, яке підвищує ризик інциденту, а не аномальна подія як така [1].

Базові методи відбиралися так, щоб покривати типові класи UEBA-рішень: (i) статистичний скоринг на основі нормалізованих відхилень без оптимізації ваг; (ii) однокласові та ансамблеві моделі для виявлення аномалій у векторному представленні; (iii) двоступеневий підхід із грубим відбором кандидатів і деталізацією лише для підозрілих фрагментів, що відображає ідею «coarse-to-fine» у поведінковій аналітиці [3]. Окремо контролювався вплив **саме еволюційної оптимізації**: як референс розглядалася конфігурація з тим самим набором ознак, але з фіксованими вагами та порогом, підібраним традиційною валідацією без пошуку по підмножинах.

Запропонований метод оптимізував хромосому $C = (B, W, \tau, \rho)$, де B задає підмножину активних ознак, W – їх ваги, τ – поріг тривоги, ρ – параметр забування. Генетичний алгоритм налаштовувався в режимі помірного бюджету пошуку: популяція 40–60 особин, 25–35 поколінь, турнірний відбір малого розміру, імовірність кросоверу близько 0.8 та мутації на рівні 0.03–0.08.

Дані агрегувалися у добові вікна $\Delta = 1$. Загалом використано 2 400 користувачів та 210 діб спостережень, що дало 504 000 вікон (u, k). Аномальні епізоди вводилися в 0.4% вікон (2 016 позитивних прикладів), що зберігає характерну для UEBA дисбалансність і робить точність особливо чутливою до хибних тривог [1]. Оцінювання виконано на тестовому відрізку після налаштування порогів і параметрів на попередньому часовому інтервалі.

Таблиця 1

Порівняння запропонованого методу з базовими

Метод	$F1$	FPR	$\bar{A}$	CV_{alert}	Complexity
Статистичний скоринг (i)	0.058	0.028	84	0.62	1.00
Однокласова модель у векторному просторі (ii)	0.063	0.025	76	0.60	1.00
Двоступеневий coarse-to-fine пайплайн (iii)	0.075	0.019	58	0.52	0.85
ГА-оптимізація B, W, τ, ρ	0.093	0.013	41	0.39	0.35

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

У таблиці 1 наведено результати порівняння з базовими методами. Порівняно з двоетапним coarse-to-fine пайплайном без еволюційної оптимізації $F1$ зростає лише з 0.075 до 0.093, що є незначним покращенням. Натомість у тих показниках, які визначають реальну вартість експлуатації UEBA-детектора, перевага більш виражена. Частка хибнопозитивних спрацювань зменшується з 0.019 до 0.013, тобто приблизно на 32% у відносному вимірі, а волатильність потоку тривог за коефіцієнтом варіації падає з 0.52 до 0.39, що відповідає близько 25% відносного зниження. Середня кількість тривог на добу зменшується з 58 до 41, отже навантаження на аналітиків стає нижчим і більш прогнозованим.

Для базових підходів $Complexity$ рівна одиниці, що означає використання всього простору ознак або близького до нього підпростору. У запропонованому методі $Complexity$ становить 0.35, тобто активною залишається приблизно третина ознак. Зменшення розмірності поведінкового профілю спрощує пояснення спрацювань і знижує ризик того, що система почне реагувати на короткочасні або контекстно нестійкі сигнали, продукуючи зайві тривоги.

Отримані результати пояснюються структурою оптимізації. Еволюційний пошук по масці B усуває ознаки, які виявляються надто варіативними або слабо пов'язаними з аномальними подіями. На практиці саме такі нестійкі ознаки часто є джерелом хвиль хибних тривог при зміні ролей, інструментів чи процесів. Зменшення FPR та $\bar{A}$ демонструє що відбір компактнішого підпростору ознак знижує шанс інтерпретації короткочасних організаційних коливань як інцидентів.

Оптимізація ваг W та порогу τ дозволяє зменшити CV_{alert}, тобто кількість тривог у часових відрізках коливається менше, а навантаження на аналітиків рівномірнішим. Для великих організацій зменшення різких піків у потоці сигналів покращує процес реагування. Скорочення $Complexity$ знижує обчислювальне навантаження, що особливо корисно в системах із великим числом користувачів. Компактна конфігурація підсилює пояснюваність, це, в свою чергу, зменшує час на формування гіпотези щодо природи тривоги аналітиком.

Водночас варто наголосити що сценарій аномалій формується контрольованим додаванням змін профілю і абсолютні значення метрик відображають саме цей клас відхилень і не гарантують аналогічних результатів для всіх типів зловмисної активності.

Висновки. Для UEBA в корпоративних середовищах ключовою є експлуатаційна придатність у присутності дрейфу поведінки, дисбалансу та контекстної неоднорідності. Це перетворює задачу на пошук керованого компромісу між якістю виявлення, відсіканням шумів і складністю моделі.

Запропонований підхід формалізує поведінкове профілювання як потоковий механізм, де пішення приймається на основі зваженого скору $S_u(k)$. Новизна реалізована в тому, що конфігурація детектора $C = (B, W, \tau, \rho)$ не задається вручну й не фіксуєть-

ся раз і назавжди, а підбирається еволюційно з урахуванням штрафів за хибні тривоги, нестабільність сигналів і надмірну розмірність.

Експеримент на CLUE-LDS із контрольованим введенням аномальних подій ппродемонстрував найбільш відчутний ефект генетичного методу саме в показниках, які визначають операційну корисність. Згідно з таблицею 1, приріст за $F1$ є помірним, але частка хибнопозитивних спрацювань зменшується з 0.019 до 0.013, волатильність потоку тривог CV_{alert} – з 0.52 до 0.39, а середнє добове навантаження $\bar{A}$ – з 58 до 41. Паралельно істотно знижується складність конфігурації до приблизно третини початкового простору ознак, що забезпечує пояснюваність. Підхід демонструє перевагу як інструмент зменшення шуму та вирівнювання потоку тривог при збереженні рівня якості виявлення.

Принцип формування датасету і проведення експерименту не гарантує аналогічного ефекту для коротких високоризикових дій або складних багатоетапних кампаній.Практичним продовженням роботи є перевірка методу на ширшій палітрі сценаріїв, а також інтеграція обчислювальних обмежень та ресурсу аналітиків на реагування у вигляді явних SLA-порогів у функції пристосованості, щоб забезпечити відтворюваний перехід від експериментального налаштування до промислової експлуатації.

ЛІТЕРАТУРА

1. Kamatchi K., Uma E. Insights into user behavioral-based insider threat detection: systematic review // *International Journal of Information Security*. 2025. Vol. 24. Art. 88. DOI: 10.1007/s10207-025-01002-6.

2. Landauer M., Skopik F., Höld G., Wurzenberger M. A User and Entity Behavior Analytics Log Data Set for Anomaly Detection in Cloud Computing // *2022 IEEE International Conference on Big Data (Big Data): 6th International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2022)*. Osaka, Japan, 17–20 Dec. 2022. P. 4285–4294. DOI: 10.1109/BigData55660.2022.10020672.

3. Alzaabi F. R., Mehmood A. A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods // *IEEE Access*. 2024. Vol. 12. P. 30907–30927. DOI: 10.1109/ACCESS.2024.3369906.

4. Kim J., Park M., Kim H., Cho S., Kang P. Insider threat detection based on user behavior modeling and anomaly detection algorithms // *Applied Sciences*. 2019. Vol. 9, No. 19. Art. 4018. DOI: 10.3390/app9194018.

5. Feng W., Cao Y., Chen Y., Wang Y., Hu N., Jia Y., Gu Z. Multi Granularity User Anomalous Behavior Detection // *Applied Sciences*. 2025. Vol. 15, No. 1. Art. 128. DOI: 10.3390/app15010128.

6. Mohammed A. S., Kanka V., Selvaraj A. Advanced behavioral analytics for user and entity behavior anomaly detection in hybrid cloud environments // *Cybersecurity and Network Defense Research*. 2022. [Електронний ресурс]. Режим доступу: https://thesciencebrigade.com/ccndri/advanced-behavioral-analytics-for-user-and-entity-behavior-anomaly-detection-in-hybrid-cloud-environments/ (дата звернення: 16.12.2025).

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

7. Seenivasan S. R., Ganaga Durga M. GA trained classification for behavior based anomaly detection in the MANETS // *International Journal of Applied Engineering Research*. 2015. Vol. 10, No. 11. P. 28811–28827.
8. Zhao X., Su H., Sun Z. An intrusion detection system based on genetic algorithm for software defined networks // *Mathematics*. 2022. Vol. 10, No. 21. Art. 3941. DOI: 10.3390/math10213941.
9. Seyedi B., Postolache O. Securing IoT communications via anomaly traffic detection: Synergy of genetic algorithm and ensemble method // *Sensors*. 2025. Vol. 25, No. 13. Art. 4098. DOI: 10.3390/s25134098.
10. Chatterjee A., Ahmed B. S. IoT anomaly detection methods and applications: A survey // *Internet of Things*. 2022. Vol. 19. Art. 100568. DOI: 10.1016/j.iot.2022.100568.
11. Віжевський П. В., Савенко О. С. Еволюційна адаптація політик DLP за умов дрейфу концепції у потокових даних // *Центральноукраїнський науковий вісник. Технічні науки.* 2025. Вип. 12(43), ч. II. С. 9–19. DOI: 10.32515/2664-262X.2025.12(43).2.9-19.
12. Sachenko A., Vizhevskyi P., Savenko O., Ostroverkhov V., Maslyyak B. Modern strategies for data leak detection and prevention in corporate networks // *MoDaST 2025: Modern Data Science Technologies Doctoral Consortium.* Lviv, Ukraine, 15 June 2025. P. 275–292. URL: https://ceur-ws.org/Vol-4005/paper19.pd
13. Hinder F., Vaquet V., Hammer B. One or Two Things We Know About Concept Drift – A Survey on Monitoring in Evolving Environments. Part A: Detecting Concept Drift // *Frontiers in Artificial Intelligence.* 2024. Art. 1330257. DOI: 10.3389/frai.2024.1330257.
14. Shyaa A., Zulkernine M., Abouelela O., Miri A. Comprehensive survey: Concept drift and feature dynamics in intrusion detection systems // *Engineering Applications of Artificial Intelligence.* 2024. Vol. 132. Art. 109143. DOI: 10.1016/j.engappai.2024.109143.
15. Zhu S., Liu Z., Bansal N., Han J., Shah N., Papalexakis E., Faloutsos C. METER: A dynamic concept adaptation framework for online anomaly detection // *Proceedings of the VLDB Endowment.* 2023. Vol. 17, no. 4. P. 794–807. DOI: 10.14778/3636218.3636233.

REFERENCES

1. Kamatchi, K., & Uma, E. (2025). Insights into user behavioral-based insider threat detection: Systematic review. *International Journal of Information Security, 24*, Article 88. https://doi.org/10.1007/s10207-025-01002-6.
2. Landauer, M., Skopik, F., Höld, G., & Wurzenberger, M. (2022). A user and entity behavior analytics log data set for anomaly detection in cloud computing. In *2022 IEEE International Conference on Big Data (Big Data): 6th International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2022)* (pp. 4285–4294). IEEE. https://doi.org/10.1109/BigData55660.2022.10020672.
3. Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. *IEEE Access, 12*, 30907–30927. https://doi.org/10.1109/ACCESS.2024.3369906.
4. Kim, J., Park, M., Kim, H., Cho, S., & Kang, P. (2019). Insider threat detection based on user behavior modeling and anomaly detection algorithms. *Applied Sciences, 9*(19), 4018. https://doi.org/10.3390/app9194018.

5. Feng, W., Cao, Y., Chen, Y., Wang, Y., Hu, N., Jia, Y., & Gu, Z. (2025). Multi granularity user anomalous behavior detection. *Applied Sciences, 15*(1), 128. https://doi.org/10.3390/app15010128.
6. Mohammed, A. S., Kanka, V., & Selvaraj, A. (2022). Advanced behavioral analytics for user and entity behavior anomaly detection in hybrid cloud environments. *Cybersecurity and Network Defense Research*. https://thesciencebrigade.com/ccndri/advanced-behavioral-analytics-for-user-and-entity-behavior-anomaly-detection-in-hybrid-cloud-environments/.
7. Seenivasan, S. R., & Ganaga Durga, M. (2015). GA trained classification for behavior based anomaly detection in the MANETS. *International Journal of Applied Engineering Research, 10*(11), 28811–28827. https://www.ripublication.com/ijaer10/ijaerv10n11_125.pdf.
8. Zhao, X., Su, H., & Sun, Z. (2022). An intrusion detection system based on genetic algorithm for software defined networks. *Mathematics, 10*(21), 3941. https://doi.org/10.3390/math10213941.
9. Seyedi, B., & Postolache, O. (2025). Securing IoT communications via anomaly traffic detection: Synergy of genetic algorithm and ensemble method. *Sensors, 25*(13), 4098. https://doi.org/10.3390/s25134098.
10. Chatterjee, A., & Ahmed, B. S. (2022). IoT anomaly detection methods and applications: A survey. *Internet of Things, 19*, 100568. https://doi.org/10.1016/j.iot.2022.100568.
11. Vizhevskyi, P. V., & Savenko, O. S. (2025). Evolutionary adaptation of DLP policies under concept drift in streaming data. *Central Ukrainian Scientific Bulletin. Technical Sciences*, 12(43), Part II, 9–19. https://doi.org/10.32515/2664-262X.2025.12(43).2.9-19.
12. Sachenko, A., Vizhevskyi, P., Savenko, O., Ostroverkhov, V., & Maslyyak, B. (2025). Modern strategies for data leak detection and prevention in corporate networks. In *Proceedings of the Modern Data Science Technologies Doctoral Consortium (MoDaST 2025)* (CEUR Workshop Proceedings, Vol. 4005, pp. 275–292). CEUR-WS.org. https://ceur-ws.org/Vol-4005/paper19.pdf.
13. Hinder, F., Vaquet, V., & Hammer, B. (2024). One or two things we know about concept drift—a survey on monitoring in evolving environments. Part A: Detecting concept drift. *Frontiers in Artificial Intelligence, 7*, Article 1330257.
https://doi.org/10.3389/frai.2024.1330257.
14. Shyaa, M. A., Ibrahim, N. F., Zainol, Z., Abdullah, R., Anbar, M., & Alzubaidi, L. (2024). Evolving cybersecurity frontiers: A comprehensive survey on concept drift and feature dynamics aware machine and deep learning in intrusion detection systems. *Engineering Applications of Artificial Intelligence, 137*, 109143.
https://doi.org/10.1016/j.engappai.2024.109143
15. Zhu, J., Cai, S., Deng, F., Ooi, B. C., & Zhang, W. (2023). METER: A dynamic concept adaptation framework for online anomaly detection. *Proceedings of the VLDB Endowment, 17*(4), 794–807. https://doi.org/10.14778/3636218.3636233.

Received 15.01.2026.
Accepted 19.01.2026.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Evolutionary adaptation of DLP policies under concept drift in streaming data

In modern streaming DLP systems deployed across cloud and hybrid environments, fixed policies degrade rapidly due to concept drift. Operators must simultaneously control the risk-weighted miss cost, limit the false-alarm burden, meet latency SLOs, and keep alert streams stable under tight memory and compute budgets. These competing objectives are not adequately balanced by traditional detectors or manual policy tuning.

We present an online evolutionary controller that casts policy adaptation as constrained multi-objective optimization. The method uses a chromosome encoding with drift-aware exploration–exploitation switching, an archive of vetted policies for warm starts, a compact active mixture, and guarded rollbacks for operational safety. On six streams (synthetic and real), the controller keeps the integrated cost within 0–3.5% of the best baseline (mean absolute gap ≈1.6%), sustains p95 latency below 100 ms, and reduces alert-rate volatility by 50–63% while maintaining comparable or lower false-alarm rates.

Two practical sensitivities emerge: the drift-gate threshold governing the exploration/exploitation balance, and short-lived compute bursts immediately after detected changes. Warm starts, a compact mixture, and mutation-budget guards mitigate these effects without sacrificing responsiveness.

Keywords: genetic algorithms, Data Loss Prevention, anomaly detection, concept drift, cloud security

Віжевський Петро Володимирович – асистент кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, Україна.
ORCID: https://orcid.org/0009-0009-4851-0839
Савенко Олег Станіславович – професор, д.т.н., професор кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, Україна.
ORCID: https://orcid.org/0000-0002-4104-745X

Vizhevskyi Petro – teaching asistant of Computer Engineering and Programming Department, Khmelnytskyi National University, Ukraine.
ORCID: https://orcid.org/0009-0009-4851-0839
Savenko Oleg– Professor, Doctor in Information Technology (Doctor of Technical Sciences), Professor of Computer Engineering and Programming Department, Khmelnytskyi National University, Ukraine.
ORCID: https://orcid.org/0000-0002-4104-745X