DOI 10.15276/imms.v15.no3.377
УДК 004.056.53:162.5:159.955

Informatics and Mathematical Methods in Simulation
Vol.15 (2025), No. 3, pp. 377-385

МОДЕЛЮВАННЯ ДИНАМІКИ ЛОГІЧНИХ ПАТЕРНІВ КІБЕРЗЛОЧИНЦІВ

О.Я. Ковальчук[1], Р.І. Іваницький[2], Л.В. Бабала[1]

[1]Західноукраїнський національний університет
11, Львівська, Тернопіль, 46009, Україна
[2]Тернопільський національний педагогічний університет імені Володимира Гнатюка
2, М. Кривоноса, Тернопіль, 46027, Україна
Emails: olhakov@gmail.com, romikiv@ukr.net, ludaduma7@gmail.com

У роботі досліджено вплив когнітивних упереджень на поведінку кіберзлочинців у контексті розширених постійних загроз. Проаналізовано чотири ключові когнітивні упередження: нехтування базовою частотою, упередження підтвердження, аверсію до втрат та помилку невідшкодованих витрат. Встановлено, що нехтування базовою частотою призводить до переоціювання значущості окремих індикаторів цілей; упередження підтвердження спонукає до селективного аналізу інформації; аверсія до втрат визначає вибір між агресивними та прихованими методами кібератак; помилка невідшкодованих витрат зумовлює прив'язаність до субоптимальних злочинних стратегій. Розроблено математичний апарат для формалізації когнітивних упереджень через параметричні моделі з відповідними коефіцієнтами для упередження підтвердження, аверсії до втрат та невідшкодованих витрат. Запропоновано адаптивну когнітивно-темпоральну модель (АКТМ), що інтегрує множинні упередження у єдину динамічну систему з механізмами навчання та адаптації. Модель враховує темпоральну еволюцію когнітивного профілю кіберзлочинців та забезпечує адаптивний перерозподіл уваги між упередженнями залежно від конкретних умов кібератаки. Визначено практичні застосування моделі для розробки персоналізованих систем виявлення загроз, оптимізації стратегій активної оборони та створення адаптивних пасток-систем для хибного націлювання зловмисників. Обґрунтовано можливість використання когнітивних упереджень для підвищення ефективності захисних механізмів через провокування прогнозованих помилок у рішеннях кіберзлочинців.

Ключові слова: математичне моделювання, логічні патерни, когнітивний профіль, кібербезпека, розширені постійні загрози, прийняття рішень, кіберзлочинність.

Вступ. Феномен розширених постійних загроз (Advanced Persistent Threats, APT) – складних, довготривалих та цілеспрямованих атак на інформаційні системи (ІС) належить до найбільш небезпечних проявів сучасної кіберзлочинності. Його специфіка полягає у поєднанні високотехнологічних засобів атаки з тривалим і прихованим проникненням до ІС. Середній час перебування APT у мережі без виявлення може сягати понад пів року, що створює передумови для масштабних втрат і стратегічних загроз [1]. На відміну від класичних кіберінцидентів, APT не мають на меті швидке завдання шкоди. Їхня мета – систематичне накопичення даних, контроль за інформаційними потоками та вплив на критично важливі ресурси.

APT здатні змінювати баланс сил на геополітичному та економічному рівнях. Так, у 2010 році атака Stuxnet вперше продемонструвала можливість фізичного знищення промислового обладнання шляхом маніпуляцій програмним забезпеченням систем управління технологічними процесами [2]. Операція Aurora, спрямована проти корпорацій Google та Adobe, виявила масштаби державного кібершпигунства та необхідність переосмислення підходів до інформаційної безпеки [3]. Північнокорейське угруповання APT38 у 2015–2016 рр. здійснило серію атак через мережу SWIFT, викравши сотні мільйонів доларів, що продемонструвало новий рівень фінансово мотивованого кіберзлочину [4]. Особливу складність становить те, що APT поєднують технічні та когнітивні компоненти. Технічний рівень виявляється у використанні zero-day вразливостей, кастомізованого шкідливого програмного забезпечення та методів

приховування слідів [5]. Когнітивний рівень полягає у застосуванні соціальної інженерії, маніпуляції інформаційними потоками та використанні людських вразливостей. Саме цей симбіоз зумовлює ефективність APT і робить їх об'єктом міждисциплінарних досліджень. Когнітивне моделювання може стати дієвим інструментом у вивченні APT та забезпечити можливість аналізувати динаміку взаємодії між технічними та людськими факторами, будувати сценарії розвитку атак і прогнозувати поведінку як зловмисників, так і потенційних жертв. Так нещодавні дослідження у сфері агентного моделювання показують можливість відтворення APT-кампаній з урахуванням когнітивних упереджень користувачів, що значно підвищує точність прогнозу щодо їхнього впливу на систему [6]. Когнітивне моделювання відкриває перспективи формування нових підходів до забезпечення кібербезпеки, де ключову роль відіграє розуміння не лише технічних вразливостей, а й людського фактору.

Це дослідження продовжує цикл робіт з питань розробки інноваційних підходів до запобігання, виявлення та мінімізації наслідків кібератак [7–9].

Мета роботи. Метою даної роботи є розробка адаптивної когнітивно-темпоральної моделі логічних патернів кіберзлочинців у контексті APT на основі математичної формалізації когнітивних упереджень для підвищення ефективності прогнозування тактик атак та оптимізації захисних стратегій.

Для досягнення мети дослідження поставлено та вирішено такі завдання:

- ідентифікувати та формалізувати ключові когнітивні упередження (логічні патерни), що впливають на процеси прийняття рішень операторами APT під час довготривалих кібератак;
- розробити математичний апарат для моделювання нехтування базовою частотою, упередження підтвердження, аверсії до втрат та помилки невідшкодованих витрат у контексті кібероперацій;
- створити адаптивну когнітивно-темпоральну модель, що інтегрує множинні упередження у єдину динамічну систему з механізмами навчання та адаптації;
- обґрунтувати механізми темпоральної еволюції когнітивного профілю зловмисників під впливом зовнішніх умов і накопиченого досвіду;

Когнітивні упередження в APT: ризики ігнорування статистичної поширеності в кібероперації. Нехтування базовою частотою (base rate neglect) є загальновідомим когнітивним упередженням, яке полягає у схильності людини недооцінювати або повністю ігнорувати статистичну поширеність (базову частоту) події при оцінюванні її ймовірності у конкретній ситуації [11]. Люди схильні надавати перевагу специфічній інформації про частковий випадок, нехтуючи при цьому загальною статистичною інформацією [12]. У контексті кібербезпеки, зокрема при аналізі поведінки операторів APT, це упередження може суттєво впливати на процеси прийняття рішень під час проведення атак. Професіонали з безпеки є настільки ж вразливими до когнітивних упереджень, як і пересічні люди, внаслідок чого їхні рішення можуть виявитися менш оптимальними, ефективними або результативними [13]. Враховуючи зростаючу складність цільових атак, інструменти виявлення APT швидко впроваджуються в індустрії кібербезпеки. За оцінками експертів, глобальні продажі рішень для захисту від APT зростуть з 6,9 мільярда доларів у 2022 році до приблизно 15,2 мільярда доларів до 2026 року [14], що підкреслює критичну важливість розуміння поведінкових аспектів таких атак.

При проведенні APT-атак оператори часто стикаються з необхідністю прийняття рішень щодо цільових файлів або облікових записів на основі неповної інформації. Нехтування базовою частотою може проявлятися у переоцінці значущості певних ключових слів або назв файлів, які здаються релевантними або цінними. Наприклад, якщо зловмисник зустрічає файл з назвою, що містить слова «конфіденційно» або «секретно», він може помилково припустити, що такі файли завжди містять цінну інформацію, ігноруючи при цьому статистичну ймовірність того, що більшість файлів з такими назвами можуть бути звичайними документами. Традиційні системи виявлення

вторгнень (Itrusion Detection Systems, IDS) страждають від низької точності виявлення, високого рівня помилкових спрацювань та складності ідентифікації невідомих атак [15], що робить розуміння когнітивних упереджень зловмисників потенційно корисним для розробки більш ефективних захисних механізмів.

Упередження підтвердження в операціях APT. Упередження підтвердження, що визначається як схильність обирати варіанти, які підтримують власні переконання з пулу інформації, впливає на більшість рішень злочинців, що скоюють APT [16]. Це когнітивне упередження проявляється у їхній поведінці через надмірну увагу до доказів, що підтверджують їхні первинні гіпотези, при одночасному ігноруванні суперечливих свідчень.

Когнітивні моделі можуть ефективно відображати поведінку кіберзлочинців з різними рівнями аверсії (негативної реакції на певний тригер) до втрат та упередження підтвердження, досягаючи точності класифікації принаймні 0,83 у прогнозуванні когнітивних вразливостей [17]. Коли оператор APT виявляє файл з обліковими даними сервера та формує гіпотезу про існування цільового сервера з важливими файлами, навіть після численних невдалих спроб автентифікації він може продовжувати дотримуватися початкової гіпотези, інтерпретуючи невдачі як тимчасові технічні проблеми, а не як спростування своїх припущень.

Математично це можна представити через параметр $\theta_c \in [0, 1]$, що позначає частоту виявлення підтверджуючих доказів серед усіх дій з перевірки файлів облікових даних. За умови $\theta_c >> 0{,}5$ можна стверджувати про наявність високого рівня упередження підтвердження у поведінці зловмисника [18].

Аверсія до втрат є когнітивною вразливістю, що призводить до негативної емоційної реакції на втрати, навіть у тому разі, коли переваги переважають над ризиками [19]. Кіберзлочинці, схильні до уникнення втрат, віддають перевагу малоризиковим методам збору даних. На початковому етапі ідентифікації сервісів вони обмежуються скануванням найбільш поширених портів, відкладаючи перевірку інших на пізніші етапи та здійснюючи її приховано. Подібна тактика імітує звичайну мережеву активність, що знижує ймовірність виявлення з боку правоохоронних органів.

Відповідно до теорії перспектив, асиметричне сприйняття зловмисниками втрат або здобутків $\omega \in R$ може бути представлене функцією суб'єктивної корисності $u(\omega, \lambda_l)$, в якій λ_l R+ позначає коефіцієнт, що контролює аверсію до втрат [20].

У процесі виявлення сервісів аверсію до втрат можна представити рівностями (1)–(2):

$$\Psi(\alpha, \beta, \lambda_l) = \frac{1}{1 + e^{-\tau(\upsilon(\alpha,\lambda_l) - \upsilon(\beta,\lambda_l))}}, \quad (1)$$

$$\upsilon(\omega, \lambda_l) = \begin{cases} \omega^{\delta} & \text{якщо } \omega \geq 0 \\ -\lambda_l(-\omega)^{\delta} & \text{якщо } \omega < 0 \end{cases}, \quad (2)$$

де параметри $\alpha \in R$ та $\beta \in R$ представляють оцінені втрати або виграші агресивного виявлення сервісів та прихованого виявлення сервісів відповідно.

Агресивне виявлення сервісів (α):

- виграші: швидке отримання повної інформації про відкриті порти та сервіси; економія часу; можливість швидко перейти до наступного етапу атаки;
- втрати: високий ризик виявлення системами захисту; можливе блокування IP-адреси; втрата стелс-режиму; ризик припинення всієї операції.

Приховане виявлення сервісів (β):

- виграші: низький ризик виявлення; збереження анонімності; можливість тривалої присутності в мережі без підозр

– втрати: значно більше часу на збір інформації; неповні дані про цілі; можливість втратити вікно можливостей; ризик того, що вразливості будуть виправлені до завершення сканування.

$\Psi(\alpha, \beta, \lambda_l) \in [0, 1]$ – це ймовірність застосування агресивного виявлення сервісів. Якщо $\alpha > \beta$ (виграші агресивного підходу перевищують приховані), то $\Psi(\alpha, \beta, \lambda_l) \to 1$ (висока ймовірність агресивної поведінки). Якщо $\alpha < \beta$ (втрати агресивного підходу занадто великі), то $\Psi(\alpha, \beta, \lambda_l) \to 0$ (перевага прихованому підходу).

Параметр λ_l посилює негативне сприйняття втрат: чим більше λ_l, тим більше зловмисник уникає агресивних дій навіть при потенційно високих виграшах. Це дозволяє моделювати різні типи зловмисників: від обережних (високе λ_l) до ризикових (низьке λ_l). $\delta \in \mathrm{R}$ представляє параметр, що контролює кривизну $\upsilon(\omega, \lambda_l)$.

$\tau \in \mathrm{R}$ визначає, наскільки «стабільним» або «непередбачуваним» є кіберзлочинець у своїх рішеннях; наскільки «рішуче» він приймає рішення між агресивним та прихованим скануванням.

Високе τ (висока чутливість):

– кіберзлочинець приймає чіткі, категоричні рішення;
– навіть невелика різниця між $\upsilon(\alpha, \lambda_l)$ та $\upsilon(\beta, \lambda_l)$ призводить до однозначного вибору;
– Ψ швидко наближається до 0 або 1;
– поведінка зловмисника «або повністю агресивно, або повністю приховано».

Низьке τ (низька чутливість):

– кіберзлочинець приймає більш різноманітні «розмиті» рішення;
– навіть значна різниця між варіантами може призвести до проміжних значень Ψ;
– більше невизначеності в поведінці зловмисника;
– поведінка: змішаний підхід, часткове сканування.

Помилка невідшкодованих витрат у виборі цілей розширеними стійкими загрозами. Ще однією важливою когнітивною упередженістю, що впливає на прийняття рішень кіберзлочинцями, є помилка невідшкодованих витрат (sunk cost fallacy). Цей психологічний феномен характеризує схильність до прийняття субоптимальних рішень на основі раніше витрачених ресурсів, а не майбутніх очікуваних прибутків [21]. У сенсі APT злочинці, які припускаються помилки невідшкодованих витрат, виявляють виражену схильність до продовження дій, які вже поглинули значні часові та обчислювальні ресурси, незалежно від їх поточної життєздатності чи стратегічної цінності.

Для формалізації цієї поведінкової моделі розглянемо ландшафт загроз, що складається з Ω доступних цілей, де кожна ціль $\mu \in \Omega = \{1, ..., \Omega\}$ представляє потенційну можливість досягнення. Сприйняту корисність цілі μ можна квантифікувати через функцію оцінювання $V(\mu) : \Omega \to \mathrm{R}$.

інійна модель корисності може бути представлена рівністю (3):

$$V(\mu) = \rho(\mu) + \mu_s k(\mu), \tag{3}$$

де $\rho(\mu) : \Omega \to \mathrm{R}$ – функція винагороди, що квантифікує очікувану користь від розподілу ресурсів на ціль μ; $k(\mu) : \Omega \to \mathrm{R}$ позначає кумулятивну функцію невідшкодованих витрат, що представляє ресурси, раніше витрачені на μ. Параметр $\mu_s \in \mathrm{R}_+$ служить масштабуючим коефіцієнтом, який регулює величину впливу невідшкодованих витрат на процеси прийняття рішень.

Ймовірність вибору цілі визначається шляхом застосування процедури softmax-нормалізації та описується рівністю (4):

$$p_s(\mu) = \frac{V(\mu)}{\sum_{j=1}^{\Omega} V(j)}, \tag{2}$$

де ймовірність вибору цілі μ пропорційна її сприйнятій корисності відносно сукупної корисності всіх доступних цілей.

Ця математична модель відображає основну характеристику помилки невідшкодованих витрат, що полягає у систематичному переоцінюванні кіберзлочинцями раніше використовуваних тактик, що потенційно призводить до тривалого залучення до скомпрометованих або субоптимальних векторів атак (методів реалізації кібератаки, що здатні дати результат, проте поступаються оптимальним за ефективністю досягнення цілей зловмисника) [22].

Адаптивна когнітивно-темпоральна модель динаміки логічних патернів кіберзлочинців. Аналіз окремих когнітивних упереджень, що впливають на поведінку операторів APT забезпечив можливість зробити висновок про необхідність розробки концептуальної комплексної моделі, яка б враховувала динамічну взаємодію різних психологічних факторів у часі та їх адаптивну зміну під впливом зовнішніх умов і накопиченого досвіду [17]. Традиційні підходи до моделювання поведінки кіберзлочинців розглядають когнітивні упередження як статичні характеристики [18, 23], що не відображає реальну динаміку довготривалих APT-кампаній, які можуть тривати місяцями або роками.

У цій роботі запропоновано інноваційну адаптивну когнітивно-темпоральну модель (АКТМ), що інтегрує множинні когнітивні упередження у єдину динамічну систему з механізмами навчання та адаптації. На відміну від статичних моделей, АКТМ моделює когнітивний профіль зловмисника як багатовимірний вектор упереджень, що еволюціонує протягом тривалої APT-кампанії під впливом результатів попередніх дій, зміни умов захищеності цільової мережі та накопичення операційного досвіду.

Математичний апарат моделі. Формально, когнітивний стан оператора APT у дискретний момент часу t можна представити у вигляді багатовимірного вектора

$$\Phi(t) = [\theta_c(t), \lambda_l(t), \mu_s(t), \xi_a(t)], \tag{5}$$

де $\theta_c(t) \in [0,1]$ – коефіцієнт упередження підтвердження в момент t;

$\lambda_l(t) \in R_+$ – параметр аверсії до втрат в момент t;

$\mu_s(t) \in R_+$ – коефіцієнт впливу невідшкодованих витрат в момент t;

$\xi_a(t) \in [0,1]$ – індекс адаптивної пам'яті (параметр, що характеризує здатність кіберзлочинця модифікувати свою поведінку на основі попереднього досвіду).

Параметр $\xi_a(t)$ відображає кумулятивний вплив попередніх результатів на поточні рішення та визначається рівністю (6):

$$\xi_a(t) = \frac{1}{t}\sum_{i=1}^{t} \gamma^{j-1} R_i, \tag{6}$$

де $\gamma \in (0,1)$ – коефіцієнт забування (discount factor), $R_i \in \{-1, 0, 1\}$ – результат дії в момент i (-1 – невдача, 0 – нейтральний результат, 1 – успіх).

Механізм темпоральної еволюції. Еволюція когнітивного профілю описується системою різницевих рівнянь (оскільки APT-операції відбуваються дискретними етапами):

$$\Phi(t+1) = \Phi(t) + \Delta t \cdot G(\Phi(t), E(t), H(t)) + \varepsilon(t), \tag{7}$$

де $G: R^4 \times R^n \times R^m \to R^4$ – функція когнітивного оновлення;

$E(t) \in R^n$ – вектор зовнішніх умов (рівень захищеності мережі, час до виявлення, доступність цілей);

$H(t) \in R^m$ – вектор накопиченої історії операцій (успіхи, невдачі, виявлення);

$\varepsilon(t) \sim N(0, \Sigma)$ – стохастичний шум з коваріаційною матрицею Σ;

Δt – дискретний крок часу.

Функція G моделюється рівністю (8):

$$G(\Phi(t), E(t), H(t)) = A \cdot \tanh(W_1\Phi(t) + W_2E(t) + W_3H(t) + b), \quad (8)$$

де A, W_1, W_2, W_3 – матриці параметрів, b – вектор зміщення, tanh (гіперболічний тангенс) забезпечує обмеженість змін.

Адаптивний механізм прийняття рішень. Для кожної потенційної цілі $\omega \in \Omega$ функція корисності в момент t визначається як зважена комбінація індивідуальних компонентів упереджень:

$$U(\omega, t) = w_1(t)\cdot U_c(\omega, t) + w_2(t)\cdot U_l(\omega, t) + w_3(t)\cdot U_s(\omega, t) + w_4(t)\cdot U_a(\omega, t), \quad (9)$$

де $U_c(\omega, t)$ – корисність з точки зору упередження підтвердження;

$U_l(\omega, t)$ – корисність з урахуванням аверсії до втрат;

$U_s(\omega, t)$ – корисність з врахуванням невідшкодованих витрат;

$U_a(\omega, t)$ – адаптивна компонента на основі накопиченого досвіду.

Ваги $w(t) = [w_1(t), w_2(t), w_3(t), w_4(t)]$ визначаються через контекстно-залежний механізм уваги

$$w(t) = \text{softmax}(\Phi(t) \odot C(t)), \quad (10)$$

де softmax – математична функція, яка перетворює вектор дійсних чисел у вектор ймовірностей, сума яких дорівнює 1; $C(t)$ контекстний вектор поточної ситуації, $\odot$ – операція Адамара (поелементний добуток).

Кожна компонента корисності адаптується до конкретного когнітивного упередження:

$$U_c(\omega, t) = \theta_c(t) \cdot \text{conf}(\omega, t) + (1 - \theta_c(t)) \cdot \text{obj}(\omega, t), \quad (11)$$

де conf(ω, t) – ступінь підтвердження попередніх гіпотез цілі ω, obj(ω,t) – об'єктивна цінність цілі.

$$U_l\upsilon(\omega, t) = \begin{cases} \rho(\omega) & \text{якщо } \rho(\omega) \geq 0 \\ -\lambda_l(t)|\rho(\omega)| & \text{якщо } \rho(\omega) < 0 \end{cases}, \quad (12)$$

де $\rho(\omega)$ – очікувана винагорода/втрата від цілі ω.

$$U_s(\omega, t) = \rho(\omega) + \mu_s(t) \cdot k(\omega, t), \quad (13)$$

де $k(\omega, t)$ – кумулятивні невідшкодовані витрати на ціль ω до моменту t.

Алгоритм оновлення параметрів. Модель включає механізм навчання через підкріплення для оновлення когнітивних параметрів:

$$\Phi(t + 1) = \Phi(t) + \eta \cdot \nabla\Phi\, J(\Phi(t)) \cdot \eta(R(t)), \quad (14)$$

де $\eta \in (0,1)$ – швидкість навчання;

J(Φ) – функція якості;

R(t) – вектор винагород за період [t, t+1];

$\eta(R) = \dfrac{1 - e^{-\beta|R|}}{1 + e^{-\beta|R|}}$ – функція чутливості до результатів з параметром $\beta > 0$.

Запропонована у цій роботі АКТМ представляє кілька ключових переваг порівняно з існуючими підходами до моделювання поведінки APT-операторів:

1. На відміну від статичних моделей, АКТМ враховує еволюцію когнітивних упереджень протягом тривалого періоду. Це критично важливо для APT, оскільки оператори можуть змінювати свою поведінку на основі попередніх успіхів, невдач або зміни умов захищеності цільової мережі.

2. Модель не розглядає упередження ізольовано, а моделює їх взаємодію через механізм контекстно-залежних ваг. Це відображає реальність прийняття рішень, де кілька психологічних факторів можуть одночасно впливати на вибір тактики.

3. АКТМ забезпечує можливість прогнозувати майбутню поведінку кіберзлочинців на основі їх когнітивного профілю та історії дій. Це може значно покращити ефективність проактивних захисних стратегій.

4. Параметри моделі можуть бути налаштовані для конкретних APT-груп на основі спостережуваних поведінкових патернів, що дозволяє створювати більш точні моделі загроз.

5. Розуміння когнітивних упереджень дає можливість розробляти захисні механізми, що експлуатують ці упередження. Наприклад, можна створювати

«приманки», що використовують упередження підтвердження для відволікання уваги зловмисників від справжніх цілей.

6. Модель може ідентифікувати моменти, коли когнітивний профіль зловмисника кардинально змінюється, що може сигналізувати про зміну фази атаки або появу нових учасників кібератаки.

Практичне застосування АКТМ потребує калібрування параметрів на основі емпіричних даних про поведінку конкретних APT-груп. Це може бути досягнуто через аналіз журналів безпеки, систем-пасток, які імітують вразливий комп'ютер або мережу для приваблення кіберзлочинців, та інших джерел інформації про тактики зловмисників. Валідація моделі може здійснюватися через порівняння прогнозованих поведінкових траєкторій з реальними спостереженнями використовуючи метрики когнітивної узгодженості.

Запропонована модель відкриває нові перспективи для розвитку адаптивних систем кібербезпеки, що здатні передбачати та протидіяти еволюції тактик APT-операторів через глибоке розуміння їх психологічних мотивацій та когнітивних обмежень.

Висновки. У роботі досліджено феномен когнітивних упереджень (логічних патернів) як критичного фактору формування поведінки кіберзлочинців у контексті APT. Встановлено, що нехтування базовою частотою призводить до переоцінення значущості окремих індикаторів цілей; упередження підтвердження спонукає до селективного аналізу інформації; аверсія до втрат визначає вибір між агресивними та прихованими методами проведення атак; помилка невідшкодованих витрат зумовлює прив'язаність до субоптимальних векторів атаки навіть за умов низької ефективності.

Розроблено математичний апарат для формалізації когнітивних упереджень через параметричні моделі, що дозволяє кількісно оцінювати їх вплив на процеси прийняття рішень зловмисниками. Запропоновано параметри θ_c для упередження підтвердження, λ_l для аверсії до втрат та μ_s для невідшкодованих витрат, які забезпечують можливість прогнозування вибору цілей і тактик кіберзлочинців на основі їх когнітивного профілю. Створено адаптивну когнітивно-темпоральну модель, що представляє принципово новий підхід до моделювання динамічної еволюції поведінки операторів APT. Модель інтегрує множинні когнітивні упередження через механізм softmax-нормалізації та забезпечує адаптивний перерозподіл уваги між різними психологічними факторами залежно від контексту операції. Доведено, що темпоральна компонента моделі дозволяє ідентифікувати критичні точки зміни поведінкових патернів на основі накопиченого досвіду успіхів і невдач зловмисників.

Визначено практичні застосування запропонованої моделі для розробки персоналізованих систем виявлення загроз, оптимізації стратегій активної оборони та створення адаптивних систем-пасток для імітації вразливого комп'ютера чи мережі для введення в оману кіберзлочинців. Обґрунтовано можливість експлуатації когнітивних упереджень для підвищення ефективності захисних механізмів шляхом створення умов, що провокують прогнозовані помилки в рішеннях кіберзлочинців. Міждисциплінарний підхід, який поєднує методи когнітивної психології, теорії ігор та машинного навчання, відкриває нові перспективи для вирішення актуальних проблем кібербезпеки в умовах зростаючої складності та тривалості сучасних кібератак. Результати дослідження створюють теоретичну основу для подальшого розвитку поведінково-орієнтованих систем кібербезпеки.

Список літератури

1. FireEye Mandiant Services. M-Trends. Advanced Persistent Threat Groups Special Report. 2020. URL: https://www.fireeye.com/current-threats/annual-threat-report/m-trends.html.
2. Farwell J.P., Rohozinski R. Stuxnet and the Future of Cyber War. *Survival*. 2011. Vol. 53(1). P. 23–40. DOI: https://doi.org/10.1080/00396338.2011.555586.

3. Nasir F., Sohail R. State, Surveillance, and Cyberspace: An Intelligence Analysis of Operation Aurora. *Social Science Review Archives*. 2025. Vol. 3(2). Article 1668–1675. DOI: https://doi.org/10.70670/sra.v3i2.789.
4. Lyngaas S. FireEye unmasks a new North Korean threat group. Cyberscoop. 2018. https://cyberscoop.com/apt38-north-korea-fire-eye/.
5. Sfetcu N. APT: Definition, History and Features. In book: Advanced Persistent Threats in Cybersecurity – Cyber Warfare. 2024. DOI: 10.58679/mm28378. https://www.telework.ro/en/e-books/advanced-persistent-threats-in-cybersecurity-cyber-warfare/.
6. Aggarwal P., Thakoor O., Jabbari S., Cranford E. A., Lebiere C., Tambe M., Gonzalez C. Designing effective masking strategies for cyberdefense through human experimentation and cognitive models. *Computers & Security*. 2022. Vol. 117. Article 102671. DOI: https://doi.org/10.1016/j.cose.2022.102671.
7. Kovalchuk O., Shevchuk R., Banakh S., Cryptocurrency Crime Risks Modeling: Environment, E-Commerce, and Cybersecurity Issue. *IEEE Accpess*. 2024. Vol. 12. P. 50673–50688. DOI: 10.1109/ACCESS.2024.3386428.
8. Kovalchuk O., Shynkaryk, Masonkova M. Econometric Models for Estimating the Financial Effect of Cybercrimes. *11th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany. 2021, P. 381–384. DOI: 10.1109/ACIT52158.2021.9548490.
9. Kovalchuk O., Shynkaryk M., Masonkova M., Banakh S. Cybercecurity: Technology vs Safety. *10th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany. 2020. P. 765–768. DOI: 10.1109/ACIT49673.2020.9208951.
10. Midtgård K, Selart M. Cognitive Biases in Strategic Decision-Making. *Administrative Sciences*. 2025. Vol. 15(6). Article 227. DOI: https://doi.org/10.3390/admsci15060227.
11. Tversky A., Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*. 1973. Vol. 5 (2). P. 207–232. DOI: https://doi.org/10.1016/0010-0285(73)90033-9.
12. Johnson C.K., Gutzwiller R.S., Gervais J. Ferguson-Walter K. Decision-Making Biases and Cyber Attackers. November. *2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, Melbourne, Australia. 2021. P. 140–144. DOI: 10.1109/ASEW52652.2021.00038.
13. Cyberproof. AI Data Security: Key Threats and Protection. 2024. URL: https://www.cyberproof.com/blog/ai-data-security-key-threats-and-protection/?utm_campaign=Data%20Security&utm_source=ppc&utm_medium=Google%20Search%20Ads&campaign_id=220.
14. Mutalib N.H.A., Sabri A.Q.M., Wahab, A.W.A. et al. Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review. *Artif Intell Rev*. 2024. Vol. 57. Article 297. DOI: https://doi.org/10.1007/s10462-024-10890-4.
15. Mat N.I.C., Jamil N., Yusoff Y., Kiah M.L.M. A systematic literature review on advanced persistent threat behaviors and its detection strategy. *Journal of Cybersecurity*. 2024. Vol. 10(1). Article 23. DOI: https://doi.org/10.1093/cybsec/tyad023.
16. Imperva. Advanced persistent threat (APT). 2025. URL: https://www.imperva.com/learn/application-security/apt-advanced-persistent-threat/.
17. Huang S. *et al*. PsybORG+: Modeling and Simulation for Detecting Cognitive Biases in Advanced Persistent Threats. MILCOM 2024. *IEEE Military Communications Conference (MILCOM), Washington, DC, USA,* 2024, P. 1–6. DOI: 10.1109/MILCOM61039.2024.10773763.
18. Schmidt, U., Zank, H. What is Loss Aversion? *J Risk Uncertainty*. 2005. Vol. 30. P. 157–167. DOI:10.1007/s11166-005-6564-6.

19. Kahneman D., Tversky A. Prospect theory – Analysis of decision under risk. *Econometrica*. 1979. Vol. 47. P. 263–291. DOI: 10.2307/1914185.
20. Standen M., Lucas M., Bowman D., Richer T.J., Kim J., Marriott, D. CybORG: A Gym for the Development of Autonomous Cyber Agents. *IJCAI-21 1st International Workshop on Adaptive Cyber Defense*, 2021. DOI: https://doi.org/10.48550/arXiv.2108.09118.
21. Jeon S.-E., Oh Y.-S., Lee Y.-J., Lee I.-G. Suboptimal Feature Selection Techniques for Effective Malicious Traffic Detection on Lightweight Devices. *CMES - Computer Modeling in Engineering and Sciences*. 2024. Vol. 140(2). P. 1669–1687. DOI: https://doi.org/10.32604/cmes.2024.047239.
22. Grinde B. The Evolution of Human Social Behavior. *Encyclopedia*. 2024. Vol. 4(1). P. 430–443. DOI: https://doi.org/10.3390/encyclopedia4010029.
23. Ковальчук О.Я, Банах С.В. Логіка права: стратегія кримінального профілювання. *Актуальні проблеми правознавства*. №3 (31). 2022. С. 70–76. DOI:10.35774/app2022.03.169

MODELING DYNAMICS OF CYBERCRIMINALS' LOGICAL PATTERNS

O. Kovalchuk[1], R. Ivanytskyy[2], L. Babala[1]

[1]West Ukrainian National University
11, Lvivska Str., Ternopil, 46009, Ukraine
[2]Ternopil Volodymyr Hnatiuk National Pedagogical University
2, M. Kryvonosa. Str., Ternopil, 46009, Ukraine
Email: olhakov@gmail.com, romikiv@ukr.net, ludaduma7@gmail.com

The study investigates the impact of cognitive biases on the behavior of cybercriminals in the context of advanced persistent threats (APTs). Four key cognitive biases are analyzed: base-rate neglect, confirmation bias, loss aversion, and the sunk cost fallacy. It is established that base-rate neglect leads to overestimating the significance of specific target indicators; confirmation bias drives selective information analysis; loss aversion determines the choice between aggressive and stealth attack methods; and the sunk cost fallacy results in attachment to suboptimal criminal strategies. A mathematical framework is developed to formalize cognitive biases through parametric models with corresponding coefficients for confirmation bias, loss aversion, and sunk cost fallacy. An adaptive cognitive-temporal model (ACTM) is proposed, integrating multiple biases into a single dynamic system with mechanisms of learning and adaptation. The model accounts for the temporal evolution of the cognitive profile of cybercriminals and enables adaptive redistribution of attention among biases depending on specific cyberattack conditions. Practical applications of the model are identified for developing personalized threat detection systems, optimizing active defense strategies, and designing adaptive honeypot systems for deceptive targeting of adversaries. The study substantiates the possibility of exploiting cognitive biases to enhance the effectiveness of defensive mechanisms by provoking predictable decision-making errors among cybercriminals.

Keywords: mathematical modeling, logical patterns, cognitive profile, cybersecurity, Advanced Persistent Threats (APT), decision-making, cybercrime.