© Vyshnevska N., Kubiv S. Methodology for forming the input vector of observed network activity variables // Ukrainian Scientific Journal of Information Security, 2025, vol. 31, issue 2, pp. 80-86. ©

DOI: 10.18372/2225-5036.31.20699

МЕТОДИКА ФОРМУВАННЯ ВХІДНОГО ВЕКТОРА СПОСТЕРЕЖУВАНИХ ЗМІННИХ МЕРЕЖЕВОЇ АКТИВНОСТІ

Наталія Вишневська, Степан Кубів

Державний університет «Київський авіаційний університет»

ВИШНЕВСЬКА Наталія Сергіївна
Рік та місце народження: 1977 рік, м. Київ, Україна.
Освіта: Національний авіаційний університет, 2001 рік.
Посада: старший викладач кафедри кібербезпеки, Державного університету «Київський авіаційний університет».
Наукові інтереси: інформаційна безпека, кіберзахищеність інформаційних систем, виявлення кібератак.
Публікації: більше 20 наукових публікацій.
E-mail: nataliia.vyshnevska@npp.kai.edu.ua.
Orcid ID: 0000-0002-1358-467X

КУБІВ Степан Іванович, д.е.н. доцент
Рік та місце народження: 1962 р., с. Мшанець, Тернопільська область, Україна.
Освіта: ЛНУ імені Івана Франка і Національний університет «Львівська політехніка».
Посада: професор кафедри технічного захисту інформації Державного університету «Київський авіаційний університет».
Наукові інтереси: економічна безпека, безпека телекомунікаційних мереж , кібербезпека, штучний інтелект.
Публікації: більше 30 наукових публікацій, серед яких наукові статті, монографії, підручники та навчально-методичні посібники.
E-mail: vvkzeos@gmail.com.
Orcid ID: 0000-0002-2606-4550.

Анотація. *У статті запропоновано методику формування вхідного вектора спостережуваних змінних мережевої активності для систем виявлення та прогнозування кібератак. Методика передбачає поетапне формування вектора із залученням необроблених параметрів трафіку, їх нормалізацію, згладжування у ковзних вікнах та узгодження за часовою шкалою. Розглянуто включення параметрів з усіх функціональних блоків ознак, що забезпечує адаптацію моделі до різних типів атак. Результатом застосування методики виступає вхідний вектор, сумісний з ймовірнісними моделями з Марківськими властивостями, здатний відображати і миттєві зміни в мережі, і довгострокові тренди. Запропонований підхід забезпечує підвищення точності виявлення аномалій та зниження хибних спрацювань завдяки гнучкому налаштуванню структури вектора під характер профілю загроз.*

Ключові слова: *ймовірнісне моделювання, кібербезпека, вхідних вектор, мережевий трафік, виявлення кібератак, агрегація ознак.*

Аналіз останніх публікацій

Одним із ключових факторів побудови ефективної системи виявлення кібератак є якісне формування вхідного вектора параметрів, який забезпечує формалізоване представлення поточного стану мережевої активності. Цей вектор є основою для подальшого аналізу, прогнозування та прийняття рішень у рамках обраної моделі. У науковій літературі запропоновано низку підходів до автоматизованого або евристичного формування таких векторів, проте більшість із них орієнтовані переважно на задачі класифікації. Як наслідок, вони демонструють обмежену придатність у контексті побудови імовірнісних моделей з урахуванням часової залежності або Марківських властивостей.

Зокрема, у роботах [1, 2] представлено методики, засновані на застосуванні фільтраційних (filter-based), обгорткових (wrapper-based) процедур скорочення ознакового простору. Основна мета таких підходів полягає у зменшенні кількості параметрів при збереженні високої точності класифікації. Проте, такі методи мають низку суттєвих обмежень: відсутність механізмів адаптації до змінного профілю загроз (статичність набору ознак), неврахування часових залежностей між параметрами, обмежена здатність до багатовимірного опису динаміки мережевого середовища.

В дослідженнях [3], запропоновано використання автоенкодерів та методів реконструкції векторів, вхідний вектор формується на основі помилок реконструкції. Такий підхід дозволяє виявляти аномалії як відхилення від типових шаблонів, але має певні обмеження а саме процес формування вектора не керований та інтерпретований, структура результату залежить від гіперпараметрів моделі також не забезпечується гарантоване охоплення ключових функціональних блоків мережевої поведінки.

Окрему категорію становлять глибинні підходи автоматичного витягу ознак, зокрема із використанням convolutional нейронних мереж, attention-механізмів або wavelet-перетворень [4, 5]. Попри високу ефективність в задачах виявлення шаблонів, вони мають наступні обмеження: висока обчислювальна складність і ресурсоємність, відсутність прозорості (black-box character), відсутність можливості гнучкої адаптації структури вхідного вектора без модифікації архітектури моделі, низька сумісність з імовірнісними або марківськими структурами.

В дослідженнях [10] застосували Байєсівську машину опорних векторів (*Bayesian SVM)* для формування вхідного вектора, в якому ключові параметри оптимізувались з урахуванням їх внеску в апостеріорну ймовірність класифікації. Проте метод не враховував динаміку мережевих параметрів і мав обмежене масштабування. Авторами [6] запропоновано варіант використання байєсівського автоенкодеру (*Bayesian Autoencoder)*, в якому вхідний вектор будується як умовний розподіл параметрів. Незважаючи на інноваційність, підхід орієнтований переважно на реконструкцію, а не на структуроване представлення ознак. В оглядових роботах [7; 8] зроблено акцент на важливості інтеграції структурованих ознак у ймовірнісні графові моделі, зокрема у Байєсівські мережі. Водночас алгоритми побудови таких векторів залишаються неформалізованими або потребують ручного налаштування.

Значна частина існуючих підходів не враховує специфіку окремих типів атак. Наприклад, не передбачено адаптацію складу параметрів для виявлення DDoS, DNS тунелювання чи автоматизований підбір облікових даних, що суттєво знижує точність виявлення при зміні профілю загроз [2, 3, 5]. У таких підходах структура вектора не є контекстно-чутливою і не дозволяє врахувати відмінності в поведінкових ознаках різних сценаріїв атак [4, 9].

Ще одне обмеження є відсутність процедур попередньої обробки параметрів таких як нормалізація, згладжування або агрегація у часових вікнах [1, 4, 5]. Це ускладнює використання параметрів у ймовірнісному аналізі, оскільки відсутність уніфікованого масштабу значень або наявність шумів у даних призводить до зниження достовірності прогнозу [8].

Проаналізовані підходи мають наукову цінність і демонструють потенціал у межах задач класифікації. Але, у контексті побудови ймовірнісної моделі з Марківською властивістю, вони не враховують низку вимог а саме структурну репрезентативність параметрів, узгодженість у часі та підтримку багатомасштабності, можливість гнучкої адаптації до профілю загроз, сумісність з алгоритмами обчислення апостеріорної ймовірності станів системи. Зазначені недоліки обумовлюють необхідність розробки методики до структурованого представлення мережевих параметрів, яка дозволяє забезпечити репрезентативне, синхронізоване та масштабоване охоплення мережевої поведінки в кожен момент часу.

Мета та постановка завдання

Метою дослідження є розробка методики, яка враховує вищезазначені аспекти та забезпечує уніфіковану, адаптивну і стійку до змін структуру вхідного вектора для задач виявлення атак у реальному часі. Методика має забезпечувати репрезентативне охоплення ключових аспектів мережевої активності, можливість нормалізації та багатомасштабної агрегації даних, сумісність із ймовірнісними моделями, зокрема такими, що мають Марківську властивість станів.

Виклад основної частини

Введемо поняття вектора спостережуваних змінних мережевої активності X_t. Вектор X_t є сукупністю змінних, що відображають мережеву активність у момент часу t. Він складається з різних компонентів, кожен з яких відповідає за окремий аспект мережевої поведінки, що може сигналізувати про аномалії чи загрози.

Представимо методику формування вектора для різних типів атак.

Крок 1. Мета формування вхідного вектора параметрів і його роль у моделі

Вектор виступає інформативним джерелом спостережуваних змінних для ймовірнісних моделей з Марківською властивістю, а також має забезпечити єдину уніфіковану форму подання різнорідних параметрів трафіку; збереження контексту мережевої поведінки в кожен момент часу; можливість подальшої нормалізації, згладжування та агрегації параметрів для зменшення впливу шумів і незначущих коливань; відображення як миттєвих змін, так і довготривалих трендів; адаптацію до різних типів атак завдяки охопленню ознак з усіх функціональних блоків; сумісність з алгоритмами апостеріорного оцінювання станів, що реалізують механізми прийняття рішень в імовірнісних моделях.

Крок 2. Формування набору параметрів на основі функціональних блоків

Формування вхідного вектора параметрів передбачає структуроване представлення багатовимірної інформації про мережеву активність у вигляді впорядкованого набору ознак. Для забезпечення репрезентативності, масштабованості та сумісності з ймовірнісними моделями з Марківською властивістю параметри групуються за функціональними блоками, які відображають певні аспекти поведінки системи. Такий підхід дозволяє охопити найважливіші сценарії атак і забезпечити репрезентативне, масштабоване та адаптивне представлення середовища.

До базових функціональних блоків належать :

1. Трафікові блоки (Traffic Features) включають обсяги, інтенсивність і розподіл трафіку (кількість пакетів на одиницю часу; обсяг переданих даних; середній розмір пакета; співвідношення вхідного/вихідного трафіку). Застосовані у всіх сценаріях, як базові ознаки. Особливо актуальні для виявлення DDoS-атак, флуктуацій у загальному навантаженні, зливів даних.

2. Поведінкові блоки (Behavioral Features) описують характер і сталість взаємодій (частота

повторних запитів; тривалість сесій; кількість унікальних IP; частка невдалих спроб з'єднання). Застосовані при виявленні атак типу brute force, credential stuffing, сканування. Актуальні при аналізі клієнтської поведінки.

3. Протокольні блоки (Protocol-Specific Features) фіксують особливості використання мережевих протоколів (частота SYN-пакетів; кількість нестандартних портів; частка DNS-запитів; відношення UDP/TCP). Вводяться якщо необхідна детекція аномалій на рівні протоколу - наприклад, DNS тунелювння, TCP-фрагментація, ICMP misuse.

4. Ентропійні та статистичні блоки (Entropy and Distribution Features) відображають ступінь невпорядкованості трафіку (ентропія IP-адрес; розподіл портів; частка унікальних поєднань "IP-порт"). Застосовуються при необхідності виявити stealth-атак, zero-day активності або змін у типових шаблонах мережевої взаємодії.

5. Інфраструктурні блоки (Host/Infrastructure Load Features) відображають навантаження на ресурси системи (CPU/Memory utilization; кількість одночасних з'єднань; активність на сервісних портах). Вводяться в сценаріях, де важливо оцінити стан цільової системи наприклад, DoS-атаки на сервери, атаки на IoT.

6. Часові блоки (Temporal Context Features) Містять контекст часу (час доби; день тижня; відхилення від середньої активності у цей період). Вводяться для зниження хибних спрацювань, дозволяють адаптувати модель до норм поведінки у відповідні часові інтервали.

7. Ознаки координованої активності (Correlated/Botnet Features) Ідентифікують скоординовану або ботнет-поведінку (пульсації трафіку; однотипна активність з різних джерел; періодичність звернень). Доцільні при виявленні botnet-мереж, атак за попередньо визначеним розкладом, взаємопов'язаних атак.

Крок 3. Умови вибору функціональних блоків

Склад функціональних блоків не фіксується, а визначається на основі типу атаки, до виявлення якої готується система (наприклад DDoS → трафікові та ентропійні, а credential stuffing → поведінкові та протокольні).

Рівня доступних спостережень – у разі обмеженого доступу до payload чи внутрішніх логів перевага надається трафіковим та протокольним блокам.

Обчислювальних обмежень – у системах реального часу перевага надається менш ресурсоємним блокам (наприклад, трафіковим та часовим).

Рівня агрегації - при побудові багаторівневих моделей блоки можуть обиратись на різних рівнях узагальнення.

Потреб адаптації до профілю загроз - при зміні домену чи появі нових типів атак можливе автоматичне оновлення складу блоків.

Сформований набір параметрів репрезентує ключові ознаки мережевої активності, пов'язані з основними сценаріями загроз, зокрема DDoS-атаками, скануванням портів, тунелюванням DNS тощо. У перспективі методика може бути доповнена механізмами адаптивного уточнення структури вектора на основі накопичених даних, що дозволить забезпечити чутливість до нових або комбінованих типів атак без повного переосмислення архітектури моделі.

Крок 4. Побудова вектора χ_t на основі обраних параметрів. На цьому кроці здійснюється вибір параметрів із визначених функціональних блоків (трафікових, поведінкових, топологічних, тимчасових тощо) з урахуванням їх інформативності, незалежності та здатності відображати характерні особливості мережевої активності.

Для кожного дискретного моменту спостереження - такту часу t, який визначається заданою частотою вимірювання (наприклад кожна секунда або кожні 100 мс) формується вектор спостережень X_t. Кожен такий такт часу репрезентує фіксований інтервал в межах якого система реєструє та накопичує значення відібраних мережевих параметрів, які надходять з джерел моніторингу (NetFlow, IDS/SPS або систем збору логів).

На початковому етапі відібрані параметри попередньо не оброблені, то ж на кожному такті часу t формується вектор необроблених (сирих) значень X'_t

$$X'_t = \left[x'_{t,1}, x'_{t,2}, ..., x'_{t,n} \right]^{\mathrm{T}}$$

де $x'_{t,i}$ – зафіксоване значення i-го мережевого параметра в момент часу t, n - загальна кількість включених у вектор параметрів і відповідають функціональним блокам (трафікові, поведінкові, протокольні, ентропійні, інфраструктурні, часові, координованої активності). Сформований вектор включає наступні параметри, згруповані за функціональними блоками таблиця 1

Крок 5. Попередня обробка параметрів: нормалізація, згладжування та агрегація

Вектор X'_t, сформований на основі значень мережевих параметрів в момент часу t, наведених у табл. 1 проходить трансформацію для забезпечення інформативності даних та коректної інтеграції в аналітичну модель і складається з наступних операцій.

Таблиця 1

№	Параметр	Позначення	Тип ознаки	Опис
1	Обсяг мережевого трафіку	$x'_{t,1}$	Трафікова	Кількість байтів, переданих за фіксований інтервал часу
2	Частота переданих пакетів (PPS)	$x'_{t,2}$	Трафікова	Пакети за секунду - інтенсивність передавання
3	Частота SYN-запитів	$x'_{t,3}$	Протокольна	Індикатор можливих SYN Flood-атак
4	Аномальне використання протоколів	$x'_{t,4}$	Протокольна	Частка нестандартних/заблокованих протоколів у трафіку
5	Відсоток успішних з'єднань	$x'_{t,5}$	Поведінкова	Частка успішних TCP/HTTP-з'єднань
6	Географічне розташування трафіку	$x'_{t,6}$	Поведінкова	Кількість країн/регіонів серед джерел доступу
7	Частота HTTP/DNS-запитів	$x'_{t,7}$	Прикладна	Кількість звернень до сервісів у секунду
8	Довжина DNS-запитів	$x'_{t,8}$	Прикладна	Середня довжина доменних імен у запитах
9	Рівень завантаження CPU / RAM	$x'_{t,9}$	Інфраструктурна	Використання обчислювальних ресурсів сервера
10	Затримка мережі (Latency)	$x'_{t,10}$	Інфраструктурна	Час відповіді сервера на запит
11	Розподіл активності портів	$x'_{t,11}$	Структурна	Кількість одночасно активних TCP/UDP-портів

Z -нормалізація параметрів ґрунтується на приведенні до уніфікованого масштабу вимірювання відносно середнього значення та варіації, що дозволяє забезпечити порівнюваність різних ознак у межах єдиного вектору. і визначається за формулою

$$x_{t,i} = \frac{x'_{t,i} - \mu'_{x_i}}{\sigma'_{x_i}},$$

де μ'_{x_i} та σ'_{x_i} – середнє значення та стандартне відхилення параметра $x'_{t,i}$ за значеннями за попередні такти часу.

Заключним етапом обробки параметрів агрегація(згладжування) в ковзних часових відрізках, дозволяє підвищити стійкість до випадкових коливань і забезпечити репрезентативне подання поточних тенденцій у мережевому середовищі. Для кожного параметра $x_{t,i}$, зафіксованого в момент часу t, виконується згладжування за рахунок аналізу значень, які спостерігалися протягом певної кількості попередніх тактів часу, включно з поточним.

$$x_{t,i} = \frac{1}{m}\sum_{j=0}^{m-1} x'_{i,(t-1)},$$

де m – ширина вікна згладжування. Зокрема, формується підмножина значень параметра $\left\{x_{t+(m-1),i}, x_{t+(m-2),i}, ..., x_{t,i}\right\}$.

Отримані агреговані значення замінюють або доповнюють поточне значення параметра у сформованому векторі спостережень. Такий підхід дозволяє зберегти не лише поточний стан системи, а й врахувати короткочасну динаміку її зміни, така властивість є важливою для побудови ймовірнісних моделей із часовою залежністю.

У результаті вектор вхідних параметрів на такті часу t включає або оброблені агреговані значення або комбінацію агрегованих компонент, забезпечуючи багатовимірне й узгоджене в часі представлення стану мережі для подальшої обробки в межах системи виявлення кібератак.

Крок 6. Фінальна інтеграція. Після обробки формується остаточний вхідний вектор X_t, який подається до ймовірнісної моделі:

$$X_t = \left[x_{t,1}, x_{t,2}, ..., x_{t,n}\right]^{\top}, \quad (2)$$

де кожен елемент $x_{t,i} \in \mathbb{R}^n$ є нормалізованим, згладженим значенням конкретного мережевого показника в момент часу t, n — загальна кількість компонентів вектора X_t. Кількість компонентів у векторі X_t залежить від складності мережі та глибини аналізу, яка необхідна для ефективного моніторингу, та може змінюватись залежно від вимог до аналізу мережевої активності. У базових сценаріях може бути використано лише кілька основних параметрів, тоді як у складних системах вектор може включати десятки змінних для досягнення максимальної точності прогнозування.

Для аналізу змін у часі також визначаються вектори попередніх станів, $X_t = \left(x_{t-1}, x_{t-2}, ..., x_{t-n}\right)$ представляє значення цих показників у попередній момент часу $t-1$, вони використовуються для оцінки відхилень, трендів, обчислення ковзних середніх та побудови показників аномальності.

Для забезпечення репрезентативного охоплення ключових аспектів мережевої поведінки у кожному векторі включається щонайменше один параметр з кожного функціонального блоку.

7. Вимоги до якості вектора

7.1. Вектор має відповідати наступним функціональним вимогам

Пріоритезація параметрів здійснюється шляхом або вибору найбільш інформативних параметрів для включення у вектор X_t у разі зміни типу очікуваної атаки, або динамічного перерозподілу ваг цих параметрів у процесі агрегації індикаторів аномалій.

Залежно від контексту, модель дозволяє або відключати малозначущі параметри, або знижувати їхній вплив при прийнятті рішення. Такий підхід забезпечує адаптивність моделі без необхідності перебудови її структури при зміні профілю загроз.

Обов'язкова репрезентативність. У процесі формування вектора передбачається включення принаймні по одному параметру з кожного ключового функціонального блоку ознак. Такий підхід дозволяє гарантувати багатовимірне охоплення мережевої поведінки, навіть у випадках атаки, що проявляється лише в одному класі ознак.

Узгодження параметрів у часі. Кожен параметр має бути агрегованим у межах фіксованого інтервалу часу Δt , що забезпечує синхронізацію різних показників у межах одного такту спостереження, зменшення впливу короткочасних флуктуацій, можливість багатомасштабного аналізу на рівні подальших кроків моделі. Типові значення Δt - 1, 5 або 10 секунд - обираються залежно від цільового середовища.

Пріоритезація параметрів відповідно до типу атаки. Вибір параметрів здійснюється з урахуванням пріоритету інформативності для конкретного класу загроз. Наприклад, при очікуванні DDoS SYN Flood - включаються параметри SYN-rate, backlog, CPU; для Credential Stuffing - спроби авторизації, error-коди, IP churn, при DNS Tunneling - довжина DNS-запитів, частота subdomains, періодичність beaconing. Такий підхід дозволяє адаптувати склад вектора до поточних ризиків без втрати цілісності структури.

Обмеження розмірності та контроль надмірності. Рекомендована довжина вектора X_t від 8 до 15 параметрів, залежно від складності системи. Такий підхід дозволяє зберегти обчислювальну ефективність, уникнути ефекту надлишкової розмірності, зосередитися на найбільш релевантних індикаторах.

Занадто велика кількість параметрів призводить до надмірного навантаження на обчислювальні ресурси, збільшує розмірність ознакового простору, підвищує ризик мультиколінеарності та ускладнює побудову стабільної ймовірнісної моделі. Крім того, розширення вектора без контролю інформативності параметрів може спричинити зростання хибних спрацювань або перенавчання моделі, знижуючи стійкість до нових або непередбачуваних типів трафіку. Тому оптимізація кількості ознак дозволяє зберігати баланс між чутливістю системи та її обчислювальною стійкістю.

Адаптивна гнучкість та оновлення. Методика допускає динамічне оновлення вмісту вектора без зміни загальної структури моделі. Такий підхід дає можливість зміни набору параметрів при зміні типу атак або середовища, автоматизовану переоцінку інформативності ознак, вивантаження неактуальних ознак та додавання нових при появі нових загроз. І досягається завдяки використанню фіксованої структури обробки вектора (нормалізація, згладжування, оцінка аномалій), яка залишається сталою незалежно від кількості чи типу включених параметрів. Динамічне оновлення вмісту вектора реалізується шляхом активації/деактивації окремих ознак або оновлення пріоритетів у вагових коефіцієнтах без потреби зміни формату, логіки обробки або структури вихідної моделі. Такий підхід забезпечує адаптацію до нових загроз без порушення цілісності обчислювального процесу.

У подальшому формуванні методу ця гнучкість реалізується через формування локального вектора активних ознак із подальшою агрегацією та нормалізацією (Етап 2 методу).

7.2. Вектор має відповідати наступним аналітичним вимогам:

Здатність до нормалізації. Кожне значення параметра має бути перетворені до одного масштабу (через Z-score або інші способи) з середнім значенням 0 та стандартним відхиленням 1, для можливості їх порівняння. Такий підхід забезпечує коректне порівняння параметрів різної природи, покращує ефективність ймовірнісного аналізу, особливо в рамках Байєсівської моделі з Марківською властивістю.

Стабільність. Вектор має бути нечутливим до неінформативного шуму, а параметри очищені від сплесків, що не несуть ознак атаки.

Статистична оброблюваність. Значення мають допускати статистичну агрегацію, обчислення ковзних середніх, відхилень, ентропії, кореляцій тощо. Оскільки виявлення аномалій здійснюється у рамках багатомасштабного аналізу, кожен параметр у векторі X_t додатково обробляється у кількох часових масштабах з відповідними ширинами ковзного вікна. Конкретні значення ширини обираються залежно від типу мережевої активності, характеру очікуваних атак і необхідного рівня чутливості. Такий підхід дозволяє виявляти як короткочасні сплески активності, так і тривалі аномалії. Для кожного масштабу формується нормалізована послідовність значень, яка використовується для аналізу трендів, побудови векторів відхилення та бінарної матриці аномалій.

7.3. Вектор має відповідати наступним модельованим вимогам.

Сумісність із ймовірнісними моделями. Параметри мають бути незалежними або слабо корельованими, якщо використовується наївна байєсівська модель.

Підтримка Марківських властивостей. Вектор X_t має містити достатньо інформації для переходу до

стану з Y_t урахуванням Y_{t-1}, тобто відповідати умовам Марківської або Квазімарківської системи.

Можливість побудови функцій ризику. Параметри повинні дозволяти обчислювати функцію $f(X_t) \to P(Y_t = 1)$.

7.4. Вектор має відповідати наступним обчислювальним вимогам

Розумна розмірність. Обсяг вектора має бути достатнім, для інформативності, але не надмірним.

Масштабованість. методика формування структури вектора X_t має дозволяти розширювати/видаляти параметри без порушення структури роботи моделі.

Обробка в реальному часі: всі параметри мають бути обчислювані на основі локальних або останніх спостережень без затримок.

7.5. Вектор має відповідати наступним прикладним вимогам

Контекстна адаптація. Вміст складу вектора X_t має змінюватися відповідно до типу системи, загроз, середовища або сценарію використання.

Пріоритетність. Певні параметри повинні мати більшу вагу залежно від очікуваного типу атаки.

Підтримка різних часових шкал. Параметри повинні дозволяти формування узагальнених ознак (індикаторів) для різних часових масштабів (миттєві, коротко- та довготермінові тренди).

Методика формування вхідного вектора забезпечує цілісне, репрезентативне та адаптивне представлення стану мережевої активності в кожний момент часу. Завдяки структурованому підходу - від збирання необроблених параметрів до їх нормалізації та агрегації - вектор охоплює ключові аспекти мережевої поведінки, зберігаючи стабільність у змінному середовищі. Гарантоване багатовимірне охоплення, підтримка часової синхронізації, здатність до масштабування та оновлення вмісту вектора без порушення архітектури моделі створюють умови для подальшого ефективного застосування у ймовірнісному аналізі, виявленні аномалій та прогнозуванні кібератак із урахуванням Марківської структури станів.

Дотримання рекомендації запропонованої методики при формуванні вхідного вектора спостережуваних змінних мережевої активності дає змогу сформувати вектор з компонентів, що відображають актуальну мережеву активність у момент часу t. Це дає змогу створити гнучку систему, яка адаптується до потреб конкретної мережі, і дозволяє на основі аналізу показників виявляти потенційні загрози та забезпечувати своєчасне реагування на загрози, що знижує ризики безпеки та підвищує загальну стійкість мережевої інфраструктури.

Висновки

В роботі розглянута методика формування вхідного вектора спостережуваних параметрів мережевої активності, задачею якого є виявлення кібератак на основі ймовірнісного аналізу з Марківською властивістю. Аналіз існуючих підходів засвідчив, що більшість сучасних рішень зосереджені на класифікаційних задачах та не забезпечують необхідний рівня адаптивності, структурної репрезентативності, часової узгодженості та сумісності з ймовірнісними моделями.

Запропонований підхід в порівнянні з іншими методиками, стійкість до багатьох класів атак забезпечується за рахунок репрезентативного охоплення всіх функціональних блоків мережевої поведінки. Виявлення короткочасних, і тривалих аномалій забезпечується за рахунок Багатомасштабної обробки даних із застосуванням ковзних середніх та Z-нормалізації. В залежності від типу загроз **та** перебудови моделі гнучку адаптацію структури вектора, без втручання в логіку моделі; Можливість побудови апостеріорної оцінки стану системи та динамічного порогового реагування **за** рахунок Інтеграції з ймовірнісними та Марківськими властивостями**.**

Розроблена методика створює концептуальну основу для формування вхідних даних для систем виявлення атак, орієнтовані на ймовірнісне моделювання мережевої поведінки. Також може бути використана як базовий компонент для побудови адаптивних систем кіберзахисту, що функціонують у режимі реального часу та враховують зміни профілю загроз.

Література:

[1] Iglesias, F., Zseby, T. Analysis of network traffic features for anomaly detection // *Machine Learning*. 2014. Vol. 101, № 1–3. P. 59–84. DOI: 10.1007/s10994-014-5473-9.

[2] Ma, Q., Sun, C., Cui, B. A novel model for anomaly detection in network traffic based on support vector machine and clustering // *Security and Communication Networks*. 2021. Article ID 2170788. DOI: 10.1155/2021/2170788.

[3] Hasan-Torabi, H., Farahani, A., Ghasemi-Gol, M. An autoencoder-based anomaly detection framework for cybersecurity // *Journal of Information Security and Applications*. 2022. Vol. 66. 103179. DOI: 10.1016/j.jisa.2022.103179.

[4] Song, H., Takabi, H., Jang, J. Deep learning for network intrusion detection: CNN and attention-based architectures // *IEEE Transactions on Network and Service Management*. 2023. Vol. 20, № 1. P. 420–435. DOI: 10.1109/TNSM.2022.3203041.

[5] Duan, H., Yang, K., Wang, Y. Hybrid wavelet and CNN feature extraction for anomaly-based intrusion detection // *Applied Soft Computing*. 2022. Vol. 122. 108316. DOI: 10.1016/j.asoc.2022.108316.

[6] Casajus-Setién, C., Zurita, J. M., Carvajal, I., González-Pardo, A. A Variational Autoencoder-based Bayesian model for cyberattack detection // *Information Sciences*. 2022. Vol. 590. P. 542–562. DOI: 10.1016/j.ins.2022.02.025.

[7] Jensen, F. V., Nielsen, T. D. Bayesian Networks and Decision Graphs. 2nd ed. Berlin: Springer, 2007. 447 p. ISBN: 978-1-84628-828-4.

[8] Kozhukhovskyi, A. D., Korovaichenko Yu. Yu. Методологія прогнозування інсайдерських загроз на основі

байєсівських мереж//*Інформаційна безпека*. 2023. № 3. С. 40–50. DOI: 10.31673/2409-7292.2023.030404.

[9] Корнієнко, Ю. В., Кузнєцов, С. А., Мельник, О. П. Ідентифікація та прогнозування самоподібного трафіку в мережах//*Проблеми телекомунікацій*. 2022. № 2. С. 5–19. DOI: 10.28925/2518-6739.2022.2.1.

[10] Sotiris, S., Kokolakis, S., Gritzalis, D. Bayesian support vector machine for intrusion detection//Proceedings of the 5th International Conference on Information Systems Security and Privacy. 2010. P. 25–36.

Vyshnevska N., Kubiv S. Methodology for forming the input vector of observed network activity variables

This paper presents a methodology for constructing the input vector of observed network activity variables for cyberattack detection and prediction systems. The proposed approach involves a step-by-step formation of the vector, beginning with the collection of raw traffic parameters, followed by their normalization, smoothing within sliding time windows, and temporal alignment. The methodology includes the integration of parameters from all functional feature blocks, enabling model adaptation to various types of attacks. As a result, the input vector is fully compatible with probabilistic models exhibiting Markov properties and is capable of capturing both instantaneous fluctuations and long-term behavioral trends in network traffic. The proposed approach enhances anomaly detection accuracy and reduces false positives by enabling flexible adjustment of the vector's structure in accordance with the threat profile dynamics.

Keywords: *probabilistic modeling, cybersecurity, input vector, network traffic, feature aggregation*

Вишневська Наталія Сергіївна, старший викладач кафедри Кібербезпеки, Державний університет «Київський авіаційний університет».
Vyshnevska Nataliya, Senior Lecturer, Department of Cybersecurity, State University "Kyiv Aviation University".

Кубів Степан Іванович д.е.н., професор кафедри Технічного захисту інформації, Державного університету «Київський авіаційний університет».
Kubiv Stepan, Doctor of Economic Sciences, Professor, Department of Technical Information Security, State University "Kyiv Aviation University".