УДК 004.056.5

Лахно Валерій Анатолійович
доктор технічних наук, професор, професор кафедри комп'ютерних систем, мереж та кібербезпеки,
Національний університет біоресурсів та природокористування України
ORCID: https://orcid.org/0000-0001-9695-4543
E-mail: lva964@nubip.edu.ua

Мамченко Сергій Миколайович
доктор педагогічних наук, професор кафедри комп'ютерних систем, мереж та кібербезпеки,
Національний університет біоресурсів та природокористування України
ORCID**:** https://orcid.org/0009-0006-8743-5606
E-mail: s.mamchenko@nubip.edu.ua

Матієвський Володимир Валерійович
старший викладач кафедри комп'ютерних систем, мереж та кібербезпеки,
Національний університет біоресурсів та природокористування України
ORCID: https://orcid.org/0000-0002-1954-8493
E-mail: m_vv@outlook.com

АСПЕКТИ ВИЯВЛЕННЯ КІБЕРЗАГРОЗ В МЕРЕЖЕВОМУ ТРАФІКУ УНІВЕРСИТЕТУ

Анотація. *Сучасні кібернетичні загрози для телекомунікаційних систем і мереж характеризуються високим ступенем прихованості, адаптивності та різноманітності. Це ускладнює їхнє оперативне виявлення в мережевому трафіку, зокрема, університету. В умовах мінливої структури кібератак традиційні методи, що базуються на сигнатурному аналізі та фіксованих правилах, виявляються недостатньо ефективними для ідентифікації нових або модифікованих загроз. У зв'язку з цим зростає значущість розробки інтелектуальних гібридних підходів. Такі методи здатні аналізувати поведінкові характеристики університетського трафіку та адаптуватися до його змін. У статті представлено метод виявлення кібернетичних загроз, заснований на поєднанні методів ансамблевої кластеризації та баєсівського імовірнісного моделювання. На першому етапі використовується машинне навчання для виділення прихованих поведінкових ознак мережевих з'єднань в університетській мережі на основі різних кластеризаційних алгоритмів. Отримані ембединги поведінки надалі слугують вхідними даними для побудови баєсівської мережі, що описує імовірнісні залежності між параметрами поведінки та ознаками аномальності. Запропонований підхід дозволяє не тільки фіксувати відхилення від нормальної поведінки в трафіку, але й забезпечує інтерпретованість рішень у сфері інформаційної безпеки. Практична цінність методу полягає в його потенціалі для застосування в системах моніторингу мережевого трафіку в корпоративних мережах.*

Ключові слова: *мережевий трафік, мережа, університет, поведінковий аналіз, баєсова мережа, кластеризація, машинне навчання, метод, кібербезпека.*

Вступ. Комп'ютерні мережі об'єктів інформатизації, перебуваючи в стані постійного розвитку, та в міру розвитку інформаційних технологій і посилення залежності бізнес-процесів від роботи мереж, стають дедалі вразливішими до широкого спектру кіберзагроз. Такі кіберзагрози проявляються у вигляді атак різної природи, починаючи від несанкціонованого доступу й закінчуючи складними багатоетапними вторгненнями [1, 2]. Зі збільшенням обсягів мережевого трафіку та ускладненням поведінкових шаблонів користувачів традиційні методи виявлення загроз та аномалій на базі сигнатур і статичних евристик, описані в роботах [2-6], як показано в [6-11], втрачають свою ефективність. Ця обставина обумовлена не тільки високою динамікою кібернетичних загроз, а й необхідністю оперативної адаптації до нових типів атак. А такі нові атаки, часто не мають заздалегідь визначених шаблонів [11, 12].

У даній статті розглядається новий гібридний підхід до аналізу мережевого трафіку, що поєднує методи машинного навчання (МН) та імовірнісного моделювання на базі баєсівських мереж (БМ). Запропонований метод ґрунтується на багатоступеневій обробці спостережуваних мережевих ознак. Спершу за допомогою ансамблевої кластеризації виявляються приховані поведінкові представлення, що відображають структуру взаємодій та поведінкову неоднорідність мережевої активності. А потім на їхній основі будується баєсівська модель, що дозволяє робити імовірнісні висновки про належність трафіку до нормальної або аномальної категорії. Запропонований у статті метод інтегрує переваги навчальних поведінкових моделей та пояснюваність імовірнісних структур, забезпечуючи високу адаптивність до розмаїття кібернетичних загроз при збереженні інтерпретованості отримуваних результатів.

Постановка проблеми. Задачу виявлення кіберзагроз у мережевому трафіку формулюють як проблему виявлення аномальної поведінки, що характеризує потенційно небезпечні або шкідливі дії в потоці мережевих з'єднань. Основна складність полягає в тому, що поведінка тих, хто атакує, може відрізнятися в кожному окремому випадку. Відповідно, такі дані не будуть заздалегідь представлені в навчальних вибірках (на чому базуються багато дослідників, наприклад, у роботах [2-10]). Крім того, аномалії часто мають поведінковий характер і проявляються не в значеннях окремих ознак, а у відхиленнях від типових шаблонів активності. Існуючі методи [8, 9, 11] або використовують статичні правила та сигнатури, не здатні впоратися з невідомими атаками, або застосовують методи машинного навчання (МН), які, хоча й здатні виявляти складні залежності, часто страждають від нестачі інтерпретованості та нездатності враховувати причинно-наслідкові зв'язки. В умовах, коли потрібно одночасно забезпечувати точність виявлення, стійкість до хибних спрацьовувань та можливість пояснення результатів, виникає необхідність у комплексних моделях, що поєднують емпіричну адаптацію з імовірнісною інтерпретацією..

Поставлене завдання полягає в розробці методу, здатного виявляти кіберзагрози на підставі поведінкового аналізу мережевого трафіку, спираючись на приховані закономірності, що виявляються машинним навчанням (МН), та баєсову мережу (БМ), яка вміє моделювати імовірнісні залежності між характеристиками поведінки та аномальністю.

Методи та моделі. Етап 1. Виявлення поведінкових ознак за допомогою ансамблевої кластеризації.

Нехай наявна вибірка мережевого трафіку:

$$\chi = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}, x^{(i)} \in \mathbb{R}^d, \tag{1}$$

де $x^{(i)} = \left(x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)}\right)$ – вектор ознак для і-го мережевого з'єднання (тривалість з'єднання, байти, протокол тощо).

Далі обираємо набір М кластеризаційних алгоритмів:

$$C = \{\mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_M\}.$$

Тут кожен C_j – це функція $\mathrm{C}_j: \mathbb{R}^d \longrightarrow \{1,2, \dots, K_j\}$, яка кожному вектору $x^{(i)}$ зіставляє мітку кластера $z_j^{(i)} = C_j\left(x^{(i)}\right)$.

Тоді отримаємо такий проміжний результат.

$$Z^{(i)} = \left(z_1^{(i)}, z_2^{(i)}, \dots, z_M^{(i)}\right) \in \prod_{j=1}^{M}\{1, \dots, K_j\}. \tag{2}$$

На даному етапі мета полягає в тому, щоб виявити приховані поведінкові закономірності в мережевому трафіку шляхом застосування до даних кількох алгоритмів кластеризації. Це дозволяє сформувати узагальнене представлення про можливі шаблони трафіку. Наприклад, припустимо, що кожен об'єкт у вибірці — це окреме мережеве з'єднання, описане вектором ознак. Як ми згадували раніше, наприклад, тривалість з'єднання, кількість переданих байт, тип протоколу тощо. Ці дані надходять на вхід кількох кластеризаційних алгоритмів, що

становлять ансамбль. В якості алгоритмів можуть бути обрані такі добре відомі методи, як K-means, DBSCAN, ієрархічна кластеризація, спектральна кластеризація та інші. Зауважимо, що вибір конкретних алгоритмів для ансамблю визначається з урахуванням кількох факторів. Серед цих факторів розглядаються відмінності в методах апроксимації щільності та відстаней (щільнісні методи або метричні), чутливість до форми та розміру кластерів, стійкість до шуму та викидів, масштабованість при роботі з великими обсягами трафіку. Отже, застосовуючи кожен з алгоритмів до вихідних даних, ми отримаємо, що кожне мережеве з'єднання виявляється віднесеним до якого-небудь кластера. Таким чином, кожному з'єднанню зіставляється мітка, тобто ідентифікатор кластера, до якого воно віднесене алгоритмом.

Оскільки використовується кілька алгоритмів, для кожного з'єднання формується вектор міток, де кожна компонента відповідає результату кластеризації за одним із методів. У такому випадку, проміжний результат цього етапу являє собою так зване "кластерне представлення" поведінки з'єднання. Це набір кластерних міток, отриманих за результатами роботи всіх методів ансамблю. Дана множина міток не слід інтерпретувати безпосередньо як ознаки в класичному розумінні. Замість цього вона слугує основою для подальшої побудови числових поведінкових ознак, що відображають узгодженість або розбіжність в оцінках різних кластеризаторів. Зазначені ознаки на наступних етапах методу використовуватимуться як вхідні змінні в імовірнісній моделі баєсової мережі, дозволяючи їй враховувати та інтерпретувати поведінкові закономірності, виявлені кластеризацією.

Для переходу до числового вектора поведінкових ознак використовується відображення:

$$\Phi: \prod_{j=1}^{M}\{1, \dots, K_j\} \longrightarrow \mathbb{R}^k, \tag{3}$$

де Ф – відображення (функція), що перетворює вихідні кластерні мітки у вектор поведінкових ознак. Тобто. Ф виконує роль ембедингу, беручи результати кластеризації (мітки кластерів від кожного алгоритму) та перетворюючи їх на числовий вектор фіксованої довжини.;

K_j – результат роботи j-го кластеризатора (із ансамбля з M методів) над об'єктом x_i. Або $K_j(x_j)$ – кластерна мітка, присвоєна j-м методом кластеризації для i-го мережевого з'єднання;

$\mathbb{R}^k$ – вектор із k дійсних чисел. Вектор кластерного ембедингу для i-го мережевого з'єднання є вектором ознак розмірності d, придатним для подачі на вхід БМ. Розмірність d залежить від кількості методів в ансамблі (M), та кількості кластерів, що видаються кожним методом.

Тоді, після того як для кожного мережевого з'єднання було сформовано вектор кластерних міток, отриманих від кількох алгоритмів кластеризації, виникає необхідність перетворити даний вектор у числове представлення. Таке представлення має бути придатним для подальшого аналізу в рамках баєсової моделі. Це перетворення називається кластерним ембедингом [13, 14].

Кластерні мітки, отримані на попередньому етапі, є категоріальними значеннями (наприклад, "кластер 1", "кластер 3" тощо), які не несуть кількісного змісту і не можуть бути безпосередньо використані в баєсовій мережі (БМ), де змінні вимагають або числового, або суворо імовірнісного представлення. Крім того, принципово отримати стійкі до шуму ознаки. Тоді ці ознаки відображатимуть структуру даних, виявлену на підставі кількох методів кластеризації, а не результатів окремого методу. Для цього виконується відображення вектора міток у новий простір ознак, або простір поведінкових ознак, де кожна компонента відображає участь об'єкта в тих чи інших кластерах.

Дане перетворення може бути реалізоване різними методами. Розглянемо конкретний приклад. Припустимо, для одного з'єднання отримано наступні мітки від трьох алгоритмів: K-means відніс з'єднання до кластера 2 з 4 можливих. DBSCAN визначив з'єднання як "шум" (не відніс до жодного кластера). Ієрархічна кластеризація помістила з'єднання в кластер 3 з 5 можливих. Отже, отримуємо вектор міток:

[Kmeans=2, DBSCAN=Noise, Hierarchical=3].

Щоб перетворити цей набір на числовий вектор (ембединг), можна застосувати, наприклад, one-hot кодування. Тоді K-means → [0, 1, 0, 0] (активний кластер 2 з 4). DBSCAN → [0, 0] (припустимо, два допустимих кластери, а "шум" кодується нулями). Та Hierarchical → [0, 0, 1, 0, 0] (активний кластер 3 з 5).

Об'єднуючи все, отримуємо ембединг довжиною 11 [0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0]. Цей вектор вже можна використовувати в наступних імовірнісних моделях, оскільки він чисельне інтерпретований та відображає характеристики об'єкта. Крім того, цей вектор агрегує інформацію одразу від кількох кластеризаторів і не залежить від вихідних чисельних ознак.

Головна мета – надати БМ «чисті» та високорівневі ознаки, що відображають структуру поведінки, а не лише технічні параметри трафіку. Відповідно, БМ працює не з необробленими даними, а зі стійкими та інтерпретованими характеристиками поведінки. Це суттєво для задачі виявлення аномалій у мережі, де сама «аномальність» найчастіше проявляється як поведінкове відхилення від норми.

У підсумку ми отримаємо поведінковий профіль об'єкта $x^{(i)}$, отриманий на основі кластерної належності.

Тоді на наступному етапі у нас йде побудова баєсової мережі поверх витягнутих ознак. Баєсова мережа – це імовірнісна модель, що описує залежність між випадковими величинами [15, 16].

Етап 2. Формалізуємо склад вузлів БМ.

Нехай:

$\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_k)$ – змінні, що відповідають компонентам $h^{(i)}$;

$A \in \{0,1\}$ – бінарна випадкова величина, що відображає аномальність (*A=1* – аномалія, *A=0* – норма).

Тут $h^{(i)}$ – прихована (латентна) змінна, що характеризує аномальність *i*-го об'єкта. Наприклад, мережевого з'єднання. Це не спостережувана напряму характеристика, а прихована гіпотеза про те, чи належить об'єкт до класу «нормальний трафік» або «аномалія». Фактично це цільовий вузол у БМ, що моделює гіпотезу про те, чи є об'єкт потенційною кіберзагрозою.

Об'єднуємо всі змінні у множину $\boldsymbol{v} = (Z_1, \dots, Z_k, \boldsymbol{A})$. Множина $\boldsymbol{v}$ представляє собою повний набір змінних, що використовуються в БМ. Ми об'єднуємо їх, щоб визначити структуру БМ, або, іншими словами, щоб встановити, між якими змінними можуть існувати імовірнісні залежності. Також $\boldsymbol{v}$ необхідна для того, щоб задати область факторизації, оскільки БМ будується як факторизація сумісного розподілу всіх змінних зі $\boldsymbol{v}$, з використанням спрямованого ациклічного графа (DAG) [8]. Або у формалізованому вигляді

$$G = (V, E), V = \boldsymbol{v}, E \subseteq V \times V.$$

Кожне ребро $(Z_i, A) \in E$ інтерпретується як «ознака поведінки Z_i яка впливає на ймовірність того, що з'єднання аномалія».

Структура мережі, представлена у вигляді орієнтованого ациклічного графа, слугує «каркасом» для моделювання імовірнісних відношень між змінними. Визначивши, які вузли (змінні, такі як прихована змінна, що характеризує аномальність, та витягнуті поведінкові ознаки) безпосередньо пов'язані, ми сформуємо базу для факторизації сумісного розподілу ймовірностей. Іншими словами, це означає, що кожна змінна в мережі розглядається разом з набором своїх безпосередніх попередників, тим самим дозволяючи адекватно описати її імовірнісну поведінку. Перехід до сумісного розподілу здійснюється за допомогою розбиття повної імовірнісної міри на добуток умовних розподілів, де кожна компонента відповідає вузлу графа і залежить тільки від змінних, з якими він безпосередньо пов'язаний у структурі (іншими словами, від його батьків у графі).

Сумісний розподіл у БМ задається так:

$$P(Z_1, \dots, Z_k, \boldsymbol{A}) = \prod_{v \in V} P(v | pa(v)), \quad (4)$$

де $pa(v) \subseteq V$ – множина батьків вузла v.

Основна мета – обчислити апостеріорний розподіл:

$$P(\boldsymbol{A} = 1 | Z_1 = z_1, \dots, Z_k = z_k), \tag{5}$$

що і є передбаченням імовірності аномалії для об'єкта $x^{(i)}$.

Сумісний розподіл виражає повне множинне залежностей, виділених у ході побудови графа, і дозволяє потім проводити розрахунки умовних імовірностей, необхідні для виявлення кіберзагроз. Отже, об'єднання концепцій кроків методу, що включають отримання структури БМ та сумісного розподілу, відображає фундаментальну ідею баєсівського підходу, при якій попередньо визначена структура залежностей задає правила, за якими можна розкласти та описати складний розподіл імовірностей, що лежить в основі моделі виявлення аномалій у мережевому трафіку.

Після формалізації сумісного розподілу всіх змінних, включених у структуру БМ, наступним логічним етапом є процедура навчання моделі. Навчання БМ — це процес визначення числових параметрів, що відповідають умовним імовірностям у вузлах графа. Для кожної змінної, включеної в мережу, потрібно оцінити її умовний розподіл, що задається її батьками в графі залежностей.

Якщо структура мережі заздалегідь фіксована (наприклад, спираючись на апріорні знання експертів), завдання навчання зводиться до оцінювання параметрів розподілів. У випадку дискретних змінних, до яких належать як латентна змінна, що відображає аномальність, так і ознаки, отримані з кластерного ембедингу, навчання здійснюється шляхом підрахунку відносних частот за навчальною вибіркою. Тоді забезпечується максимізація правдоподібності, тобто параметрична настройка моделі таким чином, щоб вона найкращим чином пояснювала спостережувані дані про мережевий трафік.

Власне, навчання БМ є центральним етапом нашого методу. На цьому етапі імовірнісна модель набуває конкретного числового змісту, що відображає статистичний взаємозв'язок між поведінковими ознаками трафіку та гіпотезою про аномальність. А отримана навчена модель згодом послужить основою для імовірнісного висновку, дозволяючи оцінювати ступінь належності нових мережевих з'єднань до потенційно небезпечних на підставі їхнього поведінкового профілю.

Результати дослідження. Запропонований метод аналізу мережевого трафіку для виявлення кіберзагроз можна концептуалізувати у вигляді виразу (6).
Іншими словами, метод можна представити як композицію послідовно застосовуваних відображень, див. вираз (6). Причому кожне з яких реалізує певну функціональну трансформацію над даними, наближаючи нас до формування імовірнісної моделі поведінки.

$$x^{(i)} \xrightarrow{C} Z^{(i)} \xrightarrow{\Phi} h^{(i)} \xrightarrow{\text{BN inference}} P\left(A = 1 \middle| h^{(i)}\right). \tag{6}$$

На вхід подаються спостережувані характеристики мережевих з'єднань, що представляють собою низько рівневі мережеві ознаки $x^{(i)}$. Ці дані проходять через блок машинного навчання, зокрема – ансамблеву кластеризацію, результатом якої стає перетворення вихідних спостережень у поведінкове представлення. Цей етап можна інтерпретувати як перше відображення. Це перше відображення виявляє приховані поведінкові закономірності, характерні для різних типів активності в мережі, та кодує їх у вигляді ембедингів.

Наступним етапом є друге відображення – побудова баєсової мережі на отриманих ембедингах, тобто

$$h^{(i)} \xrightarrow{\text{BN inference}} P\left(A = 1 \middle| h^{(i)}\right).$$

Це відображення встановлює імовірнісні залежності між поведінковими ознаками та гіпотезою про аномальність з'єднання, дозволяючи об'єднати поведінку, зафіксовану

кластеризаторами, з імовірнісною моделлю, пристосованою враховувати невизначеність, причинно-наслідкові зв'язки та частково спостережувані дані.

Підсумкова модель, таким чином, являє собою композицію перетворень, що послідовно переходить від вихідних мережевих ознак до імовірнісної оцінки ризику $a^{(i)}$ – це бінарна (або категоріальна зміна), що вказує, до якого класу насправді належить мережеве з'єднання. А параметр $\hat{A}^{(i)}$ – прогноз моделі, тобто результат імовірнісного висновку в БМ.

Зазвичай $\hat{A}^{(i)}$ означає

$$\hat{A}^{(i)} = \begin{cases} 1, \text{якщо } P\left(h^{(i)} = 1 \middle| z^{(i)}\right) > \tau, \\ 0, \text{ в інших випадках,} \end{cases} \quad (8)$$

де $\tau -$ заздалегідь вибраний поріг.

Наприклад, $a^{(i)}$ істинна мітка з датасету, яка використовується для оцінки якості. Тоді $\hat{A}^{(i)} -$ передбачення моделі, тобто результат баєсівського висновку про аномальність трафіку.

Запропонований підхід є розвитком існуючих методів виявлення аномалій у трафіку, він об'єднує ідеї поведінкового аналізу та імовірнісного висновку. Ключова новизна запропонованого рішення полягає у використанні ансамблевої кластеризації як механізму вилучення прихованих ознак поведінки, які потім використовуються не просто для класифікації, а для побудови динамічно адаптованої Баєсової мережі (БМ). Така мережа не лише виявляє загрози, але й здатна враховувати мінливість мережевої активності. Наприклад, це є актуальним в умовах кіберзагроз, що постійно еволюціонують. Вважаємо, що викладений у статті метод розширює класичну схему «кластеризація → маркування» і пропонує гнучку гібридну модель. У підсумку модель поєднує емпіричну поведінку та імовірнісну інтерпретацію, що надає викладеному методу добру.

Висновки. Запропонований у статті гібридний метод виявлення мережевих кіберзагроз поєднує в собі здібності поведінкового моделювання та імовірнісного висновку. Метод формує підхід до аналізу мережевого трафіку в умовах невизначеності вихідних параметрів для аналізу трафіку. Використання в методі на першому етапі ансамблевої кластеризації дозволить витягувати стійкі та інформативні представлення поведінкових ознак. А побудова на другому етапі Баєсової мережі на їх основі, відповідно, забезпечить здійсненність інтерпретованого висновку та врахування причинних залежностей між параметрами трафіку та загрозами для безпеки мережі. Представлений у статті підхід дозволяє гнучко адаптувати структуру моделі під зміну поведінки користувачів та динаміку атакуючих стратегій. Зауважимо, що викладена концептуальна схема формалізованого алгоритму для гібридного методу виявлення кіберзагроз у мережевому трафіку відкриває ряд напрямків для подальших досліджень. На наш погляд, перспективним є розвиток методів автоматичного навчання структури БМ на підставі аналізу динаміки трафіку. А також впровадження в модель навчання додаткових джерел інформації, наприклад, таких як часові та/або контекстні залежності.

Список використаних джерел

1. Lakhno, V., Yerbolat, K., Bagdat, Y., Kryvoruchko, O., Desiatko, A., Tsiutsiura, S., & Tsiutsiura, M. (2022). Модель захисту локальної мережі навчального закладу серверної системи віртуалізації. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2(18), 6-23. https://doi.org/10.28925/2663-4023.2022.18.623.
2. Korpan, Y. V. (2015). Класифікація загроз інформаційній безпеці в комп’ютерних системах при віддаленій обробці даних. Реєстрація, зберігання і обробка даних, 17(2), 39-46.
3. Ilyenko, A., Ilyenko, S., Diana, K., & Mazur, Y. (2023). Практичні підходи щодо виявлення вразливостей в інформаційно-телекомунікаційних мережах. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 3(19), 96-108. https://doi.org/10.28925/2663-4023.2023.19.96108.
4. Makarenko, O., & Yanko, A. (2022). Концепція системи виявлення та запобігання вторгнень до мережі. Системи управління, навігації та зв’язку. Збірник наукових праць, 2(68), 59-67.

5. Трокоз, Є.М., Покотило, О.А., & Щур, Н.О. (2024). Моделювання загроз канального рівня в OWASP Threat Dragon з розробкою стратегії захисту. Технічна інженерія, (1 (93)), 246-254. https://doi.org/10.26642/ten-2024-1(93)-246-254.
6. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016.
7. Jeffrey, N., Tan, Q., & Villar, J. R. (2023). A review of anomaly detection strategies to detect threats to cyber-physical systems. Electronics, 12(15), Article 3283. https://doi.org/10.3390/electronics12153283.
8. Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), Article 1333. https://doi.org/10.3390/electronics12061333.
9. Samrin, R., & Vasumathi, D. (2017, December). Review on anomaly based network intrusion detection system. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 141–147). IEEE. https://doi.org/10.1109/ICEECCOT.2017.8284615.
10. Yang, Z., Liu, X., Li, T., Wu, D., Wang, J., Zhao, Y., & Han, H. (2022). A systematic literature review of methods and datasets for anomaly-based network intrusion detection. Computers & Security, 116, 102675. https://doi.org/10.1016/j.cose.2022.102675.
11. Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2019). A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. IEEE Communications Surveys & Tutorials, 21(2), 1851–1877. https://doi.org/10.1109/COMST.2019.2891891.
12. Bereziński, P., Jasiul, B., & Szpyrka, M. (2015). An entropy-based network anomaly detection method. Entropy, 17(4), 2367–2408. https://doi.org/10.3390/e17042367.
13. Xie, J., Girshick, R., & Farhadi, A. (2016, June). Unsupervised deep embedding for clustering analysis. In Proceedings of the 33rd International Conference on Machine Learning (pp. 478–487). PMLR. https://proceedings.mlr.press/v48/xieb16.html. https://doi.org/10.48550/arXiv.1511.06335.
14. Jiang, Z., Zheng, Y., Tan, H., Tang, B., & Zhou, H. (2016). Variational deep embedding: An unsupervised and generative approach to clustering. arXiv. https://arxiv.org/abs/1611.05148. https://doi.org/10.48550/arXiv.1611.05148.
15. Лєнков, С.В., Джулій, В.М., Берназ, Н.М., & Божук, С.О. (2017). Аналіз існуючих методів та алгоритмів виявлення атак в бездротових мережах передачі даних. Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка, (56), 124-132.
16. Голубенко, О.І., Лемешко, А.В., Цвик, О.С., & Мішкур, Ю.В. (2023). Забезпечення інформаційної безпеки в локальних мережах за допомогою контролю трафіку. ITSynergy, (2), 44-51. https://doi.org/10.53920/ITS-2023-2-3.

Lakhno Valeriy
Doctor of Technical Sciences, Professor, Professor of the Department of Computer systems, networks and cybersecurity,
National University of Life and Environmental Sciences of Ukraine,
ORCID: https://orcid.org/0000-0001-9695-4543
E-mail: lva964@nubip.edu.ua

Mamchenko Sergii
Doctor of Educational Sciences, Professor, Professor of the Department of Computer Systems,

Networks and Cybersecurity,
National University of Life and Environmental Sciences of Ukraine
ORCID: https://orcid.org/0009-0006-8743-5606
E-mail: s.mamchenko@nubip.edu.ua

Matiievskyi Volodymyr
Senior lecturer, Department of Computer Systems, Networks and Cybersecurity,
National University of Life and Environmental Sciences of Ukraine
ORCID: https://orcid.org/0000-0002-1954-8493
E-mail: m_vv@outlook.com

ASPECTS OF DETECTING CYBER THREATS IN UNIVERSITY NETWORK TRAFFIC

Abstract. *Modern cyber threats to telecommunications systems and networks are characterized by a high degree of concealment, adaptability, and diversity. This complicates their rapid detection in network traffic, particularly at universities. Given the changing nature of cyberattacks, traditional methods based on signature analysis and fixed rules are proving insufficiently effective for identifying new or modified threats. In this regard, the development of intelligent hybrid approaches is becoming increasingly important. Such methods are capable of analyzing the behavioral characteristics of university traffic and adapting to its changes. The article presents a method for detecting cyber threats based on a combination of ensemble clustering and Bayesian probabilistic modeling methods. At the first stage, machine learning is used to identify hidden behavioral features of network connections in the university network based on various clustering algorithms. The resulting behavior embeddings are then used as input data for constructing a Bayesian network that describes the probabilistic dependencies between behavior parameters and anomaly features. The proposed approach not only allows detecting deviations from normal traffic behavior, but also ensures the interpretability of decisions in the field of information security. The practical value of the method lies in its potential for use in network traffic monitoring systems in corporate networks.*

Keywords*: network traffic, network, university, behavioral analysis, Bayesian network, clustering, machine learning, method, cybersecurity.*