УДК 004.7 DOI: 10.18372/2073-4751.74.17888

Столяр А.Л.,
orcid.org/0000-0002-7669-1202

АНАЛІЗ СУЧАСНИХ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

Національний авіаційний університет

stoliarannanau@gmail.com

Вступ

Інформаційні технології (ІТ) стали невід'ємною частиною повсякденного життя. Передбачається, що велика частина людства залучена до створення, зберігання, обробки та обміну даними. Інформаційні технології розвиваються швидкими темпами, а кількість пристроїв, підключених до Інтернету, тільки зростає. Як результат, також спостерігається експоненціальне зростання у кількості та складності кібератак [1-3].

З появою і розвитком нових технологій, з'являються і нові види загроз. Водночас збільшується кількість мережевого трафіку [4], що призводить до збільшення кількості кібератак, а також зменшення ефективності роботи систем виявлення загроз [5].

Одним із засобів виявлення вторгнень у комп'ютерну мережу є системи виявлення вторгнень (СІВ). Одним із видів таких систем є системи виявлення аномалій (САВ), які виявляють аномалії викликані внутрішніми та зовнішніми чинниками у комп'ютерних мережах.

Мета

Метою статті є аналіз аномалій, які зустрічаються та дослідження методів виявлення аномалій в комп'ютерних системах.

Основна частина

Загалом, термін «аномалія» розуміють як відхилення від очікуваного або нормального. Під аномаліями в мережах розуміють ненормальну поведінку або неочікуваний трафік, який суперечить очікуваній стандартній поведінці мережі [6, 8-10]. Аномалії можуть бути викликані різними факторами, включаючи пошкодження обладнання, збої програмного забезпечення та зловмисні дії [7]. Наявність аномалій може вказувати на потенційне порушення безпеки або загрозу.

В літературі [6, 9, 11-15] часто зустрічається класифікація аномалій на основі їх патерну:

1. Точкові аномалії – це аномалії, які виникають у певний момент часу та відрізняються від решти набору даних. Раптове збільшення мережевого трафіку, що не є характерним для нормальної поведінки мережі є прикладом такої аномалії.

2. Контекстуальні аномалії – це аномалії, які виникають у певному контексті. Наприклад, неочікувана мережева активність в неробочий час.

3. Колективні аномалії – це групи аномалій, які відрізняються від решти набору даних. Наприклад, велика кількість спроб входу з одного хоста.

Також ряд авторів виділяють аномалії за їх типами [7-9, 13, 16]:

1. Аномалії пов'язані з продуктивністю мережі. До цього типу відносяться всі аномалії, що впливають на продуктивність мережі. У багатьох випадках такі аномалії виникають без будь-якого свідомого планування або наміру.

Ці аномалії можуть бути викликані різними факторами, включаючи помилки проектування системи, несправності обладнання, людські помилки або зовнішні фактори, такі як умови навколишнього середовища та інші.

До таких аномалій належать: перезавантаження мережі, втрата пакетів під час передачі даних, затримки при передачі пакетів, вичерпання пропускної здатності та неправильна конфігурація мережі та ін.

2. Аномалії пов'язані з безпекою мережі. До цього типу зазвичай відносять всі

аномалії, які викликані злочинними діями третіх осіб. Вони спричинені з наміром заподіяти шкоди, порушити роботу мережі.

Ці аномалії можуть бути викликані різними типами кібератак, які використовують наявні вразливості в комп'ютерних мережах.

Виділимо наступні поширені причини виникнення аномалії, які можуть бути викликані діями зловмисників [7, 13]:

1. Аномалії викликані вторгненнями в мережу. Вторгнення відбувається коли зловмисник отримує несанкціонований доступ до мережі, використовуючи існуючі вразливості, грубу силу або соціальну інженерію.

Вторгнення в мережу може призвести до: появи незвичного трафіку (наприклад, раптового збільшення трафіку або зміни типу трафіку в мережі, що може свідчити про те, що зловмисник намагається вкрасти дані і т.д.); появи нових підозрілих підключень до мережі (наприклад, підключень до невідомих зовнішніх хостів та ін.); отримання доступу до мережевих ресурсів несанкціонованими користувачами (наприклад до серверів, баз даних, і т.д.); змін в конфігураціях мережі (наприклад появу нових користувачів, зміну параметрів мережі і т.д.); та ін.

2. Аномалії викликані зловмисним програмним забезпеченням – це програмне забезпечення, призначене для проникнення в мережу або комп'ютерну систему з метою заподіяння шкоди. Шкідливе програмне забезпечення може приймати різні форми, наприклад віруси, черв'яки, трояни.

3. Аномалії викликані атаками на відмову в обслуговуванні (DoS – Denial-of-Service). DoS атаки спрямовані на порушення нормального функціонування мережі шляхом перезавантаження її трафіком задля спричинення збою в роботі або задля перешкоджання доступу звичайних користувачів до мережі.

DoS атаки можуть спричинити: надзвичайно великий обсяг трафіку, що може призвести до перезавантаження мережі та сповільнення часу відповіді мережі чи серверу; збої в доступі; збої в роботі обслуговуючих мережевих служб; незвичні патерни трафіку (наприклад часті повторювані запити до мережних служб та пристроїв та ін.); та ін.

4. Аномалії викликані застосуванням сканерів портів. Зловмисники можуть використовувати сканери портів для надсилання пакетів на порти мережевих вузлів з метою виявлення відкритих портів та служб у мережі. Зловмисники використовують сканування портів для виявлення вразливостей, які можуть бути використані для отримання несанкціонованого доступу до мережі.

Сканери портів можуть спричинити: незвичні патерни трафіку; підозрілі мережеві підключення; несанкціонований доступ до мережевих ресурсів, та ін.

5. інші.

Додатково аномалії можливо класифікувати і за іншими ознаками. Наприклад за джерелом походження (на зовнішні та внутрішні), за тривалістю (короткострокові та довгострокові), за впливом (низького ступеня впливу на продуктивність та безпеку мережі, середнього ступеня впливу, високого ступеня впливу), та ін.

Класифікацію згаданих аномалії приведено на рис.

Як уже згадувалося, наявність аномалій свідчить про те, що мережа може бути об'єктом зловмисних дій з боку третіх осіб. Мережеві аномалії, спричинені кібератаками або іншими факторами, можуть мати різні негативні наслідки, такі як низька продуктивність мережі, втрата або пошкодження даних, що, в свою чергу, може негативно вплинути на бізнес-операції компанії або організації. Більше того, чим довше аномалія залишається в мережі, тим більша загроза. Тому виявлення аномалій є критично важливим питанням для забезпечення працездатності мережі та запобігання майбутнім атакам.

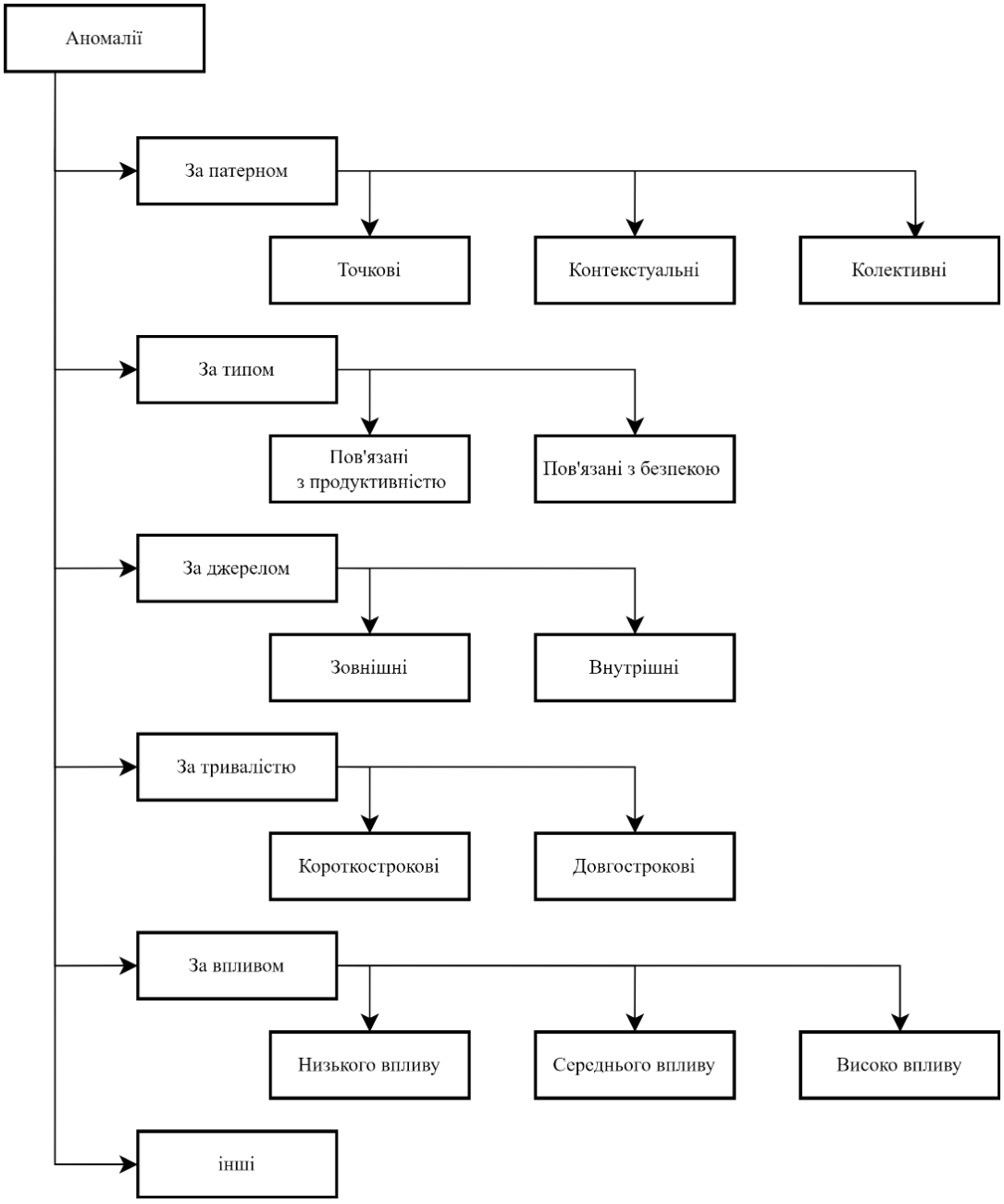

Рис. Класифікація аномалій в комп'ютерних мережах

Виявлення аномалій – це процес ідентифікації неочікуваних або нестандартних даних. Цей процес передбачає перевірку даних для виявлення патернів та відхилення від цих патернів, що можуть вказувати про потенційні проблеми в комп'ютерній мережі [9, 11, 16].

У роботах [2, 6, 9, 13-15, 17] виділяють наступні загальні категорії методів виявлення аномалій:

1. Контрольовані підходи до виявлення аномалій вимагають наявність помаркованих даних. Тобто точки даних позначаються як нормальні або аномальні. Ці помарковані дані використовуються для навчання моделі, яка визначає, чи є нова точка даних нормальною або аномальною. Контрольовані алгоритми корисні, коли є велика кількість помаркованих даних, але можуть бути неефективними, коли аномалій небагато або коли аномалії погано охарактеризовані.

2. Неконтрольовані підходи до виявлення аномалій не потребують помаркованих даних і базуються на виявленні неочікуваних патернів в самих даних. Цей

підхід припускає, що більшість точок даних є нормальними, і визначає точки, які значно відхиляються від цієї нормальної поведінки, як аномалії. Неконтрольовані підходи часто застосовують, коли помарковані дані недоступні або коли аномалії нечітко визначені.

3. Напівконтрольовані підходи до виявлення аномалій поєднують контрольовані та неконтрольовані аспекти. У цьому підході набір помаркованих даних використовується для навчання моделі, яка може виявляти аномалії, а потім використовується для класифікації додаткових точок даних як нормальних або аномальних. Модель може виявити нові аномалії, яких немає в помаркованих даних. Напівконтрольовані методи можуть бути більш успішними, ніж неконтрольовані методи, коли марковані дані невеликі, але можуть бути менш ефективними, ніж контрольовані методи, коли маркованих даних недостатньо або вони не відображають повний розподіл даних.

Порівняння цих трьох категорій наведено у табл. 1.

Таблиця 1. Порівняння контрольованих, неконтрольованих та напівконтрольованіх методів виявлення аномалій

	Переваги	Недоліки	Приклад
Контрольовані	Ефективний для відомих моделей загроз; Низький рівень хибно-позитивних результатів; Може забезпечити раннє сповіщення про відомі загрози.	Обмежується виявленням відомих аномалій; Неефективний проти нових загроз або загроз нулевого дня; Залежить від своєчасного оновлення сигнатур.	Виявлення на основі сигнатур; Виявлення на основі правил; Support Vector Machines (SVM); Decision Trees.
Неконтрольовані	Можуть виявляти нові та невідомі аномалії; Адаптуються до нових загроз; Менша залежнить від попередньо визначених правил.	Можуть генерувати більше помилкових спрацьовувань; Відсутність позначених даних для контролю; Може вимагати великих обчислювальних ресурсів.	Isolation Forest; One-Class SVM; DBSCAN (Density-Based Clustering); Gaussian Mixture Models (GMM).
Напівконтрольовані	Збалансоване виявлення відомих і невідомих загроз; Адаптивність до нових і нових загроз; Зменшує помилкові спрацьовування порівняно з неконтрольованими методами.	Потребують комбінацію помаркованих і немаркованих даних; Продуктивність моделі залежить від якості даних.	Напівконтрольоване навчання з кластеризацією;; Самонавчання з Support Vector Machines (SVM).

В літературі зустрічаються наступні групи методів виявлення аномалій:

Автори [2] виділяють статистичні методи (уніваріативні, мультиваріативні, моделі часових рядів), методи на основі знань (скінченні автомати, мови опису, експертні системи), методи на основі машинного навчання (контрольовані, неконтрольовані, напівконтрольовані).

Автори [6] виділяють методи основані на класифікації (методи машинного навчання, на основі байєсівських мереж, на базі опорних векторних машин, на базі правил), методи типу найближчий сусід (використання відстані до k-го найближчого сусіда, використання відносної щільності), методи основані на кластеризації, статистичні методи (параметричні, непараметричні), інформаційно-теоретичні методи, спектральні методи.

Автори [9, 13] виділяють статистичні методи (параметричні, непараметричні), методи на основі класифікації, кластерні методи, методи «м'якого» обчислення (генетичні алгоритми, штучні нейронні мережі, нечіткі набори, грубі набори, штучні імунні системи), методи на основі знань (підходи на основі правил та експертної системи, онтологічний і логічний підходи), комбінаційні методи.

Автори [11] виділяють статистичні методи (параметричні, непараметричні), спектральні методи, інформаційно-теоретичні методи, методи на основі машинного навчання (методи класифікації, методи типу найближчий сусід, кластерні методи) та ін.

Автори [15] виділяють статистичні методи, методи на основі класифікації, кластерні методи, методи скінченних автоматів, інформаційно-теоретичні методи, гібридні та інші.

Розглянемо більш детально різні методи:

1. Статистичні методи.

Статистичні методи використовують математичні моделі на основі статистичних характеристик даних для виявлення даних, які відхиляються від прогнозованих значень.

При цьому підході будується статистична модель нормальної поведінки мережі на основі статистичного аналізу. Після побудови моделі, її тренують на помаркованих даних, щоб навчити визначати нормальну поведінку мережі. Після навчання модель використовується для виявлення аномалій у нових немаркованих даних. За допомогою моделі виявляються будь-які точки даних, які відхиляються від прогнозованої типової поведінки. Ці точки даних позначають як потенційно аномальні.

Статистичні методи поділяють на два класи: параметричні та непараметричні [6, 9, 11, 13]. Параметричні статистичні методи базуються на припущенні, що дані відповідають заданому розподілу ймовірностей. Ці методи передбачають обчислення параметрів передбачуваного розподілу, а потім використання цих параметрів для виявлення аномалій. Непараметричні статистичні методи не базуються на заданому розподілу ймовірності, а натомість оцінюють його на основі самих даних.

Прикладом параметричних статистичних методів є методи на основі моделі Гауса [6]. Типовим підходом до пошуку аномалій або викидів у наборі даних є виявлення аномалій на основі моделі Гауса. Основний підхід полягає в тому, щоб змоделювати дані за допомогою Гаусового або нормального розподілу, а потім визначити будь-які точки даних, які суттєво відхиляються від цього розподілу, як потенційні аномалії.

Щоб використовувати цей підхід необхідно спочатку оцінити параметри розподілу Гауса для набору даних. Оцінка виконаються за допомогою стандартних статистичних методів. Після оцінки параметрів розподілу Гауса, модель може бути використана для розрахунку ймовірності виявлення для кожної точки даних у наборі даних. Тобто, аномалії – це точки даних із низькою ймовірністю, визначені за допомогою розподілу Гауса.

Ще одним прикладом є метод на основі регресійної моделі [18]. Цей метод передбачає використання регресійного

аналізу для моделювання нормальної поведінки набору даних і виявлення аномалій на основі відхилень від цієї моделі.

Регресійна модель спочатку навчається на заданому наборі даних. Потім навчена модель застосовується до кожної точки даних в наборі для прогнозування вихідної змінної. Для кожного спостереження розраховується різниця між очікуваним і фактичним значенням, а ті що з великими відхиленнями позначаються як аномалії.

Прикладом непараметричних статистичних методів є методи на основі гістограм [19]. Основна ідея полягає у створенні гістограми даних, яка буде використана для визначення точок даних, які відрізняються від інших.

Для побудови гістограми діапазон можливих значень ознак ділиться на набір відрізків. Отримана гістограма забезпечує компактне представлення частотного розподілу даних. Кількість відрізків і їх ширину вбирають, виходячи з характеристик даних і бажаної чутливості методу. Якщо при порівнянні, якісь точки даних не відносяться до якогось відрізку, їх помічають як аномальні.

Ще один приклад – це методи на основі функції ядра [6].

Основна ідея виявлення аномалій цим методом полягає у використанні функції ядра для побудування моделі нормальної поведінки набору даних. Функція ядра використовується для відображення даних у високовимірному просторі ознак. Аномалії ідентифікуються як точки даних, які мають низьку ймовірність приналежності до нормального розподілу або мають особливості, які значно відхиляються від очікуваних значень на основі функції ядра.

Також до непараметричних статистичних методів відносять методи на основі Вейвлет-аналізу [7, 20-21]. Основним принципом вейвлет-аналізу є розділення сигналу на діапазони частот за допомогою вейвлетів та оцінка кожного діапазону окремо.

Обчисливши статистичні властивості вейвлет-коефіцієнтів у кожному діапазоні, аномалії можна знайти шляхом порівняння вейвлет-коефіцієнтів кожної точки даних із очікуваними значеннями для цього діапазону та визначити будь-які коефіцієнти, які значно відрізняються.

Параметричні методи можуть бути більш ефективними, якщо дані відповідають заданому розподілу, але вони можуть бути чутливими до викидів і можуть працювати погано, якщо порушуються припущення щодо розподілу. З іншого боку, непараметричні методи можуть бути застосовані до більш широкого діапазону розподілів; однак вони можуть бути не такими надійними, як параметричні методи, і потребують більше даних для ефективного прогнозування очікуваної поведінки даних.

2. Методи на основі кластеризації.

Методи на основі кластеризації групують схожі точки даних у кластери та ідентифікують ті, які не належать жодному кластеру або належать до невеликого кластера як аномалії.

Одним із популярних методів на основі кластеризації є метод k-середніх [22-23]. Метод k-середніх ділить набір даних на k груп на основі подібності точок даних. Кожна точка даних належить кластеру з найближчим до неї центроїдом.

Щоб використовувати k-середні для ідентифікації аномалій, спочатку алгоритм навчають на нормальних даних, тобто даних, які не містять аномалій. Це створить k кластерів, які представляють звичайну поведінку набору даних. Потім навчена модель k-середніх може бути використана для прогнозування кластерного призначення кожної нової точки даних. Точка даних, яка знаходиться далеко від центроїдів усіх k кластерів, швидше за все, є аномалією.

Серед методів на основі кластеризації виділимо EM-алгоритм [6, 9]. Цей підхід передбачає, що основна маса точок даних у наборі даних належить до одного розподілу, тоді як аномалії чи викиди належать до іншого.

Спочатку EM-алгоритм оцінює параметри розподілу, до якого належать точки

даних. Потім алгоритм проходить два кроки: 1. Крок очікування: метод оцінює ймовірність того, що кожна точка даних належить кожному кластеру на основі поточних оцінок параметрів. 2. Крок максимізації: метод змінює оцінки параметрів залежно від ймовірностей, обчислених на фазі очікування. EM-алгоритм чергує дві фази, поки оцінки параметрів не співпадають.

Після ідентифікації кластерів цей підхід шукає аномалії, оцінюючи відстань між кожною точкою даних і центроїдом кластера, до якого вона належить. Точка даних вважається аномалією, якщо вона розташована далеко від центроїда кластера.

Також до методів на основі кластеризації відносять алгоритм DBSCAN [24-25]. Метою алгоритму є виявлення щільних областей точок даних, розділених областями меншої щільності.

Метод працює шляхом побудови області навколо кожної точки даних залежно від міри відстані, яку вказує користувач. Вважається, що точки в цій області доступні безпосередньо з початкової точки. Основна точка – це точка, яка має невелику кількість точок у своїй безпосередній близькості. Неосновні точки – це ті, які знаходяться поблизу основної точки, але не мають достатньо сусідів, щоб вважатися основними.

Потім метод ітеративно відвідує місця, які є негайно доступними з кожної основної точки, щоб створити кластери. Створені кластери складаються з основних точок і будь-яких неосновних точок поблизу них.

Аномальні точки визначаються як ті, що не входять до жодного кластера або є частиною кластера з невеликою кількістю точок.

3. Методи на основі класифікації.

Підходи до виявлення аномалій на основі класифікації виявляють аномалії або викиди в наборі даних шляхом навчання класифікатора розрізняти нормальні та аномальні точки даних. Цей підхід передбачає, що аномалії якимось чином відрізняються від нормальних точок даних, і їх можна ідентифікувати, навчивши класифікатор розпізнавати ці відмінності.

Одним із популярних методів на основі класифікації є використання нейронних мереж [13, 17]. При виявленні аномалій на основі нейронних мереж нейронна мережа навчається виявляти закономірності, що містяться в нормальних або типових наборах даних.

Нейронна мережа навчається на наборі даних, що містить лише нормальні точки даних, і вчиться виявляти основні закономірності та кореляції між характеристиками вхідних даних, які зазвичай зустрічаються в нормальних точках даних.

Після навчання нейронну мережу можна використовувати для виявлення аномалій і викидів у нових наборах даних. Коли на вхід нейромережі подаються нові точки даних, вихід мережі порівнюється з початковими вхідними даними. Якщо результат суттєво відрізняється від входу, точка даних вважається аномалією.

Виявлення аномалій можна виконати за допомогою різних нейронних мереж, включаючи рекурентні, згорткові нейронні мережі.

Також до методів на основі класифікації відносять метод k-найближчих сусідів [26]. Виявлення аномалій за методом k-найближчих сусідів порівнює нові точки даних з k найближчими сусідами в навчальному наборі даних.

Метод починається зі створення метрики відстані, яка порівнює схожість двох точок даних. Метод вибирає k навчальних прикладів, які є найближчими до нової точки даних на основі наданої метрики відстані. Потім обчислюється середня відстань між новою точкою даних та її k найближчими сусідами. Якщо відстань перевищує заданий поріг, нова точка даних класифікується як аномалія.

Серед методів на основі класифікації виділимо метод опорних векторів [27]. Методи виявлення аномалій на основі опорних векторів навчаються на наборах даних, що містять лише нормальні дані. Створюючи гіперплощину в просторі

ознак, яка оптимізує відстань між регулярними точками даних і межею рішення, алгоритм опорних векторів навчається розрізняти регулярні точки даних від інших точок даних.

Алгоритм обчислює відстань між кожною новою точкою даних і межею прийняття рішення. Якщо відстань перевищує заздалегідь визначений поріг, точка даних класифікується як аномалія.

Виявлення аномалій є важливою технікою в комп'ютерних мережах для визначення ненормальної поведінки, яка може свідчити про порушення безпеки чи інші проблеми.

Підходи до виявлення аномалій (табл. 2), такі як статистичні методи, методи кластеризації та класифікації, мають свої переваги та недоліки. Статистичні методи можуть успішно виявляти аномалії, які статистично відрізняються від типової поведінки, і можуть працювати з великими обсягами даних, але можуть мати труднощі з виявленням невеликих аномалій і схильні до помилкових спрацьовувань і помилкових негативних результатів. Алгоритми кластеризації чудово виявляють аномалії, які демонструють нетипову поведінку, відмінну від нормальної мережевої активності, але вони є обчислювально інтенсивними і вимагають великих обсягів пам'яті для підтримки кластерів. Методи класифікації можуть виявити аномальну поведінку, яку неможливо виявити статистичними методами або методами кластеризації, але вони вимагають великої кількості маркованих даних для навчання і схильні до помилкових спрацьовувань і помилкових відмов, якщо навчальні дані не є репрезентативними або якщо модель навчена неправильно.

Таблиця 2. Порівняння методів виявлення аномалій

	Підхід	Переваги	Недоліки
Статистичні	аналізують розподіл даних і відхилення від очікуваних моделей для виявлення аномалій.	Ефективний для виявлення точкових аномалій; Добре підходить для простих і чітко визначених статистичних заходів; Обчислювально ефективний і інтерпретований.	Обмежується певним розподілом даних і простими шаблонами; Менш ефективний для виявлення складних або контекстних аномалій; Чутливий до викидів і шуму в даних.
Класифікації	передбачають навчання моделі на позначених даних для розрізнення нормальних і аномальних випадків.	Ефективний для виявлення відомих аномалій; Підходить для широкого діапазону типів аномалій; Може забезпечити високу точність і запам'ятовування при навчанні з якісними даними.	Потрібні позначені дані для навчання, які можуть бути дефіцитними та дорогими; Неефективно бориться з новими загрозами або загрозами нульового дня; Продуктивність моделі залежить від якості та репрезентативності навчальних даних.
Кластеризації	групують схожі точки даних і ідентифікують аномалії як екземпляри, які не належать жодному кластеру.	Ефективний для виявлення колективних і контекстуальних аномалій; Менша залежність від позначених даних; Може виявити нові або несподівані моделі.	Може залежати від вибору метрики відстані та кількості кластерів; Обмежена інтерпретація складних багатовимірних даних; Може боротися з точковими аномаліями в щільних скупченнях.

На практиці для підвищення точності та надійності систем виявлення аномалій у комп'ютерних мережах часто використовують поєднання цих стратегій.

Висновки

Виявлення аномалій є важливим інструментом для забезпечення безпеки та продуктивності мережі. Різні методи виявлення аномалій можуть автоматизувати цей процес, вивчаючи шаблони мережевого трафіку і розпізнаючи аномалії, які відрізняються від нормальної поведінки. Ці аномалії можуть вказувати на цілий ряд ризиків, включаючи мережеві вторгнення, атаки на відмову в обслуговуванні та зараження шкідливим програмним забезпеченням.

Кожен підхід має свої переваги та недоліки. Метод, що використовується, слід обирати відповідно до конкретної системи та типу даних. Поєднуючи переваги декількох підходів і усуваючи їх обмеження, гібридні методи виявлення аномалій можуть стати ефективним засобом підвищення точності і надійності систем виявлення аномалій.

Література

1. Julian Jang-Jaccard, Surya Nepal. A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*. 2014. V. 80, Iss. 5. P. 973–993.

2. Khraisat, A., Gondal, I., Vamplew, P. et al. Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*. 2019. V. 2. 22 p.

3. Корченко А. Методи ідентифікації аномальних станів для систем виявлення вторгнень. Монографія. Київ: ЦП «Компринт», 2019. 361 с.

4. Packet Clearing House, Internet exchange point directory reports. URL: http://www.pch.net/ixpdir/summary.

5. Fu, Zeyuan. Computer Network Intrusion Anomaly Detection with Recurrent Neural Network. *Mobile Information Systems*. 2022. P. 1–11.

6. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. *ACM Comput. Surv*. 2009. V. 41. 72 p.

7. Чемерис К.М., Дейнега Л.Ю. Застосування методу вейвлет-аналізу для виявлення атак в мережах. *Наука і техніка Повітряних Сил Збройних Сил України*. 2022. № 1(46). С. 99–107.

8. Ali A., Khan M., Azam S., Bukhari H., Mahmood W. ADAM: A Practical Approach for Detecting Network Anomalies Using PCA. *National Conference on Emerging Technologies* / 2004. P. 44–47.

9. An Trung Tran. Network Anomaly Detection. *Seminar Innovative Internet Technologies and Mobile Communications SS2017* / 2017. P. 55–61.

10. Foorthuis R. On the nature and types of anomalies: a review of deviations in data. *Int J Data Sci Anal*. 2021. № 12. P. 297–331.

11. Baddar S., Merlo A., Migliardi M. Anomaly Detection in Computer Networks: A State-of-the-Art Review. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*. 2014. № 5. P. 29–64.

12. Mohiuddin A., Abdun N.M., Jiankun H. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*. 2016. V. 60. P. 19–31.

13. Bhuyan M.H., Bhattacharyya D.K., Kalita J.K. Network Anomaly Detection: Methods, Systems and Tools. *IEEE Communications Surveys & Tutorials*. 2014. V. 16, №. 1. P. 303–336.

14. Molina-Coronado B., Mori U., Mendiburu A., Miguel-Alonso J. Survey of Network Intrusion Detection Methods from the Perspective of the Knowledge Discovery in Databases Process. – 36 p.

15. Fernandes G., Rodrigues J.J.P.C., Carvalho L.F. et al. A comprehensive survey on network anomaly detection. *Telecommun Syst*. 2019. № 70. P. 447–489.

16. Thottan M., Ji C. Anomaly Detection in IP Networks. *IEEE Transactions On Signal Processing*. 2003. V. 51, № 8. P. 2191–2204.

17. Naseer S., et al. Enhanced Network Anomaly Detection Based on Deep Neural Networks. *IEEE Access*. 2018. V. 6. P. 48231–48246.

18. Rafferty M., Brogan P., Hastings J., Laverty D.M., Liu X., Khan R. Local Anomaly Detection by Application of Regression Analysis on PMU Data. 2018. P. 1-5.

19. Kind A., Stoecklin M.P., Dimitropoulos X. Histogram-based traffic anomaly detection. *IEEE Transactions on Network and Service Management*. 2009. V. 6, № 2. P. 110–121.

20. Cohen A., Atoui. M.A.A. On Wavelet-based Statistical Process Monitoring. *Transactions of the Institute of Measurement and Control*. 2022. № 44 (3). P. 525–538.

21. Lu W., Ghorbani A. Network Anomaly Detection Based on Wavelet Analysis. *EURASIP J. Adv. Sig. Proc* / 2009. 16 p.

22. Munz G., Li S., Carle G. Traffic Anomaly Detection Using K-Means Clustering. 2007. 8 p.

23. Syarif I., Prugel-Bennett A., Wills G. Unsupervised clustering approach for network anomaly detection. 2022. 11 p.

24. Çelik M., Dadaşer-Çelik F., Dokuz A.Ş. Anomaly detection in temperature data using DBSCAN algorithm. *International Symposium on Innovations in Intelligent Systems and Applications* / 2011. P. 91–95.

25. Chen Z., Li Y.F. Anomaly Detection Based on Enhanced DBScan Algorithm. *Procedia Engineering*. 2011. V. 15. P. 178–182.

26. Dang T.T., Ngan H.Y.T., Liu W. Distance-based k-nearest neighbors outlier detection method in large-scale traffic data. *2015 IEEE International Conference on Digital Signal Processing (DSP)* / Singapore, 2015. P. 507–510.

27. Catania C.A., Bromberg F., Garino C.G. An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection. *Expert Systems with Applications*.2012. V. 39, Iss. 2. P. 1822–1829.

Столяр А.Л.

АНАЛІЗ СУЧАСНИХ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

Проаналізовано визначення поняття аномалії, коротко описано причини їх виникнення та можливий вплив на комп'ютерні мережі. Розгляну три типи аномалій: поодинокі (точкові), контекстуальні та групові аномалії.. Також описано на основі яких характеристик відбувається виявлення аномальної поведінки. Наведено класифікації методів виявлення аномалій, які описано в науковій літературі. Розглянуто стандартні статистичні методи, методи на основі кластеризації та методи на основі класифікації.

Ключові слова: *аномалія, методи виявлення аномалій, комп'ютерна мережа.*

Stoliar A.L.

ANALYSIS OF CONTEMPORARY METHODS FOR DETECTING ANOMALIES IN COMPUTER NETWORKS

The definition of the concept of anomaly is analyzed, the reasons for their occurrence and possible impact on computer networks are briefly described. Ehree types of anomalies are considered: individual (point), contextual and group anomalies. It is also described on the basis of which characteristics abnormal behavior is detected. Classifications of anomaly detection methods described in the scientific literature are given. Standard statistical methods, methods based on clustering and methods based on classification are considered.

Keywords: *anomaly, anomaly detection methods, computer network.*