УДК 004.056

ТАКСОНОМІЯ АНОМАЛІЙ В КОНТЕКСТІ КІБЕРЗАХИСТУ ТА ЗАХИСТУ ІНФОРМАЦІЇ

М. І. Шовак[1,a], В. М. Ткач[1]

[1] Навчально-науковий Фізико-технічний інститут

Анотація

Стаття присвячена класифікації аномалій у контексті кіберзахисту та захисту інформації. У статті описано різні типи аномалій, які можуть бути ідентифіковані під час різних атак зловмисників. Також описано перелік методів виявлення, які краще використовувати для виявлення тих чи інших типів аномалій. Таксономія може бути корисною для побудови обширної системи захисту та швидкої ідентифікації аномалій в часі.

Ключові слова: таксономія, аномалія, методи виявлення аномалій, killchain

Вступ

Кібербезпека є актуальною темою у сучасному світі, оскільки кіберзлочинці неодноразово демонстрували свою здатність здійснювати атаки на різні типи організацій та навіть держав. Зі зростанням важливості кібербезпеки все більше уваги приділяється дослідженню проблем, які з нею пов'язані. Одним із життєво важливих аспектів підтримки кібербезпеки є виявлення аномалій.

Таксономія аномалій є важливим інструментом для виявлення потенційних загроз у системі. Метод класифікації дозволяє відображати різні види аномалій, організовуючи їх у окремі категорії та спрощуючи процес виявлення вразливостей у системі. У цій статті розглянуто велику кількість типів аномалій, таких як аномалії в поведінці користувачів, мережевому трафіку, системних журналах , даних, ресурсах і т. д.

Розуміння різних фаз атаки є важливим для виявлення аномалій. Різноманітні етапи, які може включати атака, це розвідка, напад, експлуатація та інші. Різні типи аномалій можуть бути видимі на різних фазах атак, тому важливо мати розуміння про те , які аномалії можуть бути спостережені на кожній фазі, і як їх можна виявити.

Дослідження проводилося на прикладі реальних сценаріїв атак та використання різних інструментів для виявлення аномалій. Результати цієї роботи допоможуть покращити підходи до кіберзахисту та забезпечити більш ефективний захист інформації в сучасному цифровому світі.

[a]sovakslavik@gmail.com

1. Класифікація аномалій

Основне поняття аномалії в контексті кіберзахисту

Аномалії – це відхилення від норми або очікуваної поведінки, які можуть виникати в різних контекстах, включаючи науку та технології. У контексті кібербезпеки аномалії – це відхилення від очікуваної поведінки системи, які можуть свідчити про порушення безпеки. Це може бути викликано різними факторами, як-от зловмисне програмне забезпечення, хакерство або внутрішні загрози. Аномалії відіграють вирішальну роль у кібербезпеці, де вони часто використовуються для виявлення зловмисних дій або інших загроз безпеці. Їхнє детектування передбачає ідентифікацію шаблонів, які відрізняються від нормальної поведінки системи, що може вказувати на наявність атаки або порушення безпеки. Виявлення аномалій має вирішальне значення для підтримки безпеки мереж, систем і даних [1].

Типи аномалій

У цій роботі була розглянута класифікація , яка грунтується на аномаліях , які можуть бути зустрінуті під час різних видів атак та даними у яких вони спостерігаються. Провівши огляд сучасних технік атак вдалося виділити такі категорії:

1. Аномалії в поведінці користувачів: ці аномалії виникають, коли користувачі поводяться незвично для свого звичайного способу діяльності, або коли вони проводять дії, які не мають логічного пояснення. Наприклад, користувач може намагатися звернутися до сервера, який не існує, або спробувати ввести неправильний пароль більше десяти разів.
2. Аномалії в мережевому трафіку: ці аномалії пов'язані з аномальною активністю в мережі,

такою як надмірний трафік або незвичайна адресація. Наприклад, зловмисники можуть використовувати мережу для розповсюдження шкідливого програмного забезпечення або отримання доступу до чутливої інформації.

3. Аномалії в системних журналах: ці аномалії пов'язані зі змінами в системних журналах, таких як незвичайна кількість записів або зміни в критичних записах. Наприклад, зловмисники можуть намагатися приховати свої дії, видаляючи або змінюючи журнал.
4. Аномалії в даних: цей тип аномалій виникає, коли дані, що зберігаються в системі, містять помилки або несподівані значення. Наприклад, це можуть бути дублікати даних, неправильні типи даних, неправильні формати даних тощо. Аномалії в даних можуть спричинити помилки в роботі програм, що використовують ці дані, а також порушити конфіденційність та цілісність даних.
5. Аномалії в ресурсах: цей тип аномалій виникає, коли система витрачає занадто багато ресурсів (таких як CPU, RAM, диск тощо) або коли деякі ресурси використовуються неправильно. Наприклад, це можуть бути процеси, що використовують занадто багато CPU часу, програми, що займають занадто багато місця на диску, або запити до баз даних, що використовують занадто багато пам'яті.
6. Аномалії в програмному забезпеченні: це незвичайні події або стан системи, які вказують на проблеми з програмним забезпеченням. Це можуть бути помилки в коді, вразливості, які можуть бути використані для злому системи або втрати даних. Наприклад, аномалія в програмному забезпеченні може включати в себе помилки в коді програми, які можуть призвести до її краху.
7. Аномалії в архітектурі системи: це відхилення від нормальної архітектурної конструкції системи, які можуть бути наслідком помилок в проектуванні, розробці та впровадженні системи, або ж в результаті її модифікацій. Ці аномалії можуть бути складними для виявлення, оскільки вони можуть проявлятися тільки під час певних умов або заходів користувача. Наприклад, несправна архітектура може призвести до витоку конфіденційної інформації через незахищені мережеві канали, або до ситуацій, коли відмовляє критична система при великих навантаженнях.
8. Аномалії в ідентифікації та автентифікації: ці аномалії пов'язані з безпекою і забезпеченням доступу до системи або ресурсів. Аномалії в ідентифікації виникають, коли користувач не може бути ідентифікований або ідентифікується неправильно, наприклад, коли користувач вводить невірний пароль або використовує незареєстрований акаунт. Також аномалії в автентифікації виникають, коли користувач ідентифікується правильно, але не може отримати доступ до ресурсу, наприклад, коли у користувача недостатньо прав доступу до файлу або каталогу.
9. Аномалії в захисті від вразливостей: цей тип аномалій виникає, коли система не може ефективно захиститись від потенційних атак з використанням відомих вразливостей. Наприклад, це можуть бути відхилення від найкращих практик захисту у різних системах , що породжують атаки, які використовують вразливості у програмному забезпеченні; схеми захисту, які можуть бути обхідні; людські помилки у конфігурації системи.
10. Хмарні аномалії: це аномалії, які виникають у хмарних обчислювальних середовищах через неавторизований доступ до хмарних ресурсів або незвичайну хмарну діяльність. Наприклад, якщо користувач, який зазвичай не має доступу до певних ресурсів хмарного постачальника, починає отримувати до них доступ, або ж використання хмарних сервісів , які раніше не використовувалися - все це може свідчити про проблему безпеки.
11. Аномалії, пов'язані з часом: це аномалії, які виникають у результаті незвичайної діяльності, що відбувається в певний час. Наприклад, якщо користувач входить у систему в незвичайний час або звертається до ресурсів системи в піковий час, це може свідчити про неавторизований доступ або напад на систему.

2. Засоби ідентифікації

Класифікація методів виявлення аномалій

Зазвичай методи поділяються на три класи: поведінкові методи, методи обчислювального інтелекту та методи машинного навчання. [2] Поведінкові методи базуються на аналізі поведінки системи та визначенні відхилень від звичної поведінки. Ці методи використовують велику кількість параметрів, що відображають стан системи, та дозволяють виявляти навіть невеликі зміни в поведінці. Системи , які побудовані на поведінкових методах використовують метод порівняння поточних показників з шаблоном нормальної поведінки для виявлення аномалій і сигналізування про можливу атаку.

До поведінкових методів відносяться:

- Статистичний аналіз
- Вейвлет аналіз
- Фрактальний аналіз
- Моделі на основі кінцевих автоматів

Методи обчислювального інтелекту використовують комп'ютерні алгоритми для виявлення аномалій. Вони базуються на аналізі великої кількості даних та використовують різноманітні алгоритми для знаходження відхилень від звичної поведінки. Ці методи зазвичай використовуються для розв'язання задач, для яких традиційні математичні методи не є ефективними або не підходять. Серед методів обчислювального інтелекту можна виділити:

- Моделі на основі нейронних мереж
- Генетичні алгоритми

- Метод опорних векторів
- Моделі на основі нелінійної динаміки
- Рольові алгоритми

Методи машинного навчання – це підхід до розв'язання задач штучного інтелекту, що полягає у використанні алгоритмів та статистичних моделей, щоб зробити прогнози, виявляти закономірності, класифікувати об'єкти і здійснювати інші дії на основі даних. Такі методи можуть використовувати як наглядні дані, так і статистичні характеристики, для пошуку незвичайних відхилень від очікуваного зразка поведінки. Результатом такого аналізу може бути виявлення невідомих аномальних подій, які можуть бути потенційно шкідливими для системи. Деякі вищеперераховані методи можуть входити до цього класу методів , але розглянуті нижче , не можна віднести до інших класів.

- Дерева рішень
- Байесовські мережі
- Кластерний аналіз
- Алгоритми регресії

Методи виявлення

Статистичний аналіз: ці моделі використовують статистичні методи для виявлення аномалій. Вони базуються на розрахунку статистичних показників, таких як середнє значення, дисперсія, кореляція тощо. Ці показники порівнюються зі значеннями, отриманими у реальних даних, і якщо відхилення виявляються достатньо значними, то можна стверджувати, що спостерігається аномалія [1]. Статистичний аналіз використовувався у напрацюванні Ma X., Chen Y., Wang S., Li Z. [3]. В ньому описується метод виявлення аномалій у часових рядах, який базується на визначенні аномалій шляхом порівняння поточних даних з нормальним розподілом, що визначається на основі історичних даних. Автори також розглядають використання різних статистичних методів, таких як метод вибіркових середніх та метод експоненціально зваженого ковзаючого середнього.

Вейвлет аналіз: це метод аналізу сигналів, який дозволяє розкласти сигнал на декілька складових частин з різною частотою. В основі вейвлет-аналізу лежить застосування математичних функцій – вейвлетів, які дозволяють виявити навіть дуже малі зміни в сигналі, що робить цей метод дуже ефективним для аналізу даних.У роботі [4] запропоновано підхід на основі вейвлет-аналізу для виявлення аномалій в мережевому трафіку, який є стійким до шуму та здатним виявляти аномалії в реальному часі.

Фрактальний аналіз: це метод виявлення аномалій у кібербезпеці, який використовується для аналізу розмірності і складності даних. Фрактал є геометричною формою, яка має самоподібність на всіх масштабах. У фрактальному аналізі, дані розглядаються як самоподібна фрактальна структура, і аномалії можуть бути виявлені шляхом порівняння розмірності і складності цих структур.

Моделі на основі кінцевих автоматів: ці моделі використовуються для виявлення аномалій у послідовностях дій. Кінцевий автомат зображає собою математичну модель, яка складається з кінцевої кількості станів і правил переходу між ними. Яскравим прикладом побудований на кінцевих автоматах є алгоритм REP (Regular Expression Pattern matching), який використовує регулярні вирази для опису нормальних шаблонів поведінки в даних. Автомат будується на основі цих регулярних виразів, і даний автомат визначає, чи належить послідовність даних до нормальної поведінки. Якщо послідовність даних не відповідає регулярному виразу, то вона вважається аномальною. використання кінцевих автоматів, було розглянуто у роботі P. Garcia-Teodoro [5]. У дослідженні пропонується система виявлення вторгнень в мережі на основі кінцевих автоматів, яка використовується для виявлення аномалій в мережевому трафіку.

Моделі на основі нейронних мереж: ці моделі використовують штучні нейронні мережі для виявлення аномалій. Цей тип є більш ефективними при роботі з великими обсягами даних, оскільки вони можуть розпізнавати складні зв'язки та залежності між даними [6]. Вони використовуються для виявлення різних типів аномалій, включаючи контекстуальні та колективні аномалії. Одним з прикладів моделей виявлення аномалій на основі нейронних мереж є використання автокодерів. Автокодер – це нейронна мережа, яка навчається відтворювати вхідні дані на виході, після проходження через прихований шар. У разі виявлення аномалій, автокодер не зможе точно відтворити вхідні дані, що свідчить про аномальність [7].

Генетичні алгоритми – це методи пошуку та оптимізації, які базуються на механізмах природного добору та еволюції. Генетичні алгоритми шукають оптимальне рішення для задачі шляхом знаходження та комбінування кращих кандидатів, які відповідають вимогам. За допомогою генетичних алгоритмів можна ідентифікувати аномальну поведінку користувача. Одним з прикладів методу, що використовує генетичні алгоритми, є генетичний програмний аналіз (Genetic Programming, GP). GP використовує генетичні алгоритми для створення програм, які можуть виявляти аномальні патерни в поведінці користувачів або систем. Алгоритм може бути навчений розпізнавати зразки поведінки користувачів, такі як інтервали між діями, кількість дій тощо, та використовувати ці знання для виявлення аномалій. У статті [8] автори пропонують метод виявлення аномалій в поведінці користувачів на основі аналізу логів серверів та застосування генетичних алгоритмів.

Метод опорних векторів – це метод машинного навчання, який використовується для класифікації даних та пошуку незвичайних зразків. Один з методів SVM(Support Vector Machines), який використовується для виявлення аномалій, називається One-Class SVM. Цей метод використовується для знаходження границі, яка охоплює нормальні точки даних. Точки, які лежать за межами цієї границі,

вважаються аномальними. Цей метод може бути використано для виявлення аномалій у задачах ідентифікації та автентифікації. В роботі H. Khan та ін. [9] запропоновано використання SVM для виявлення аномальних вхідних сигналів у системі контролю доступу до приміщень.

Моделі на основі нелінійної динаміки: ці моделі використовують теорію нелінійної динаміки, яка дозволяє аналізувати складні системи зі змінними параметрами. Основна ідея полягає в тому, що аномальні дані можуть мати іншу динаміку, ніж нормальні дані, і це може бути використано для їх виявлення. Для виявлення аномалій в програмному забезпеченні A. Mosleh та співавтори [10] було запропоновано метод на основі нелінійної динаміки та рекурентної нейронної мережі LSTM.

Рольові алгоритми: це методи виявлення аномалій у кібербезпеці, які базуються на аналізі ролей користувачів і розподілі доступів до ресурсів. Ідея полягає у тому, що якщо зловмисник намагається здійснити дії, які відрізняються від його звичайної поведінки та ролі, це може свідчити про наявність атаки. Рольові алгоритми можуть бути використані для виявлення аномалій у процесах ідентифікації та автентифікації. В роботі W. Li та ін. [11] запропоновано використання рольових алгоритмів для виявлення аномальних активностей користувачів у системі контролю доступу до мережі.

Дерева рішень: основна ідея цього підходу полягає у побудові дерева з різних правил і умов, щоб зробити прогнози на основі вхідних даних. Важливе місце в контексті виявлення аномалій займає дерево вирішення випадків інцидентів (Incident Decision Tree), яке дозволяє інформаційним безпековим фахівцям приймати рішення про класифікацію інцидентів як аномальних або нормальних. Для виявлення аномалій у архітектурі системи, Y. Zhang, Y. Wen та Q. Li [12] було запропоновано метод на основі дерев рішень. Метод був успішно використаний для виявлення аномалій у поведінці веб-сайтів та баз даних.

Байесовські мережі є графічними моделями ймовірності, які використовуються для моделювання та прогнозування подій. Вони засновані на теоремі Байеса, яка дозволяє оцінювати ймовірність події на основі знань про інші пов'язані події. Байесовські мережі можуть використовуватись для моделювання поведінки користувачів та виявлення аномальних дій. У роботі J. Han і J. Huang [13] було запропоновано метод виявлення аномалій в системних журналах на основі байєсовських мереж, який може виявляти аномалії шляхом порівняння з нормальними шаблонами та попереднім навчанням на зразках нормальної поведінки.

Кластерний аналіз: використовується для виявлення структури в наборі даних, шляхом розділення елементів на групи (кластери) з подібними характеристиками. Один з методів кластерного аналізу, який використовується для виявлення аномалій у кібербезпеці, - це метод кластеризації на основі густоти (DBSCAN). В цьому методі, кластери формуються на основі густоти точок в наближенні до певної точки. Аномалії можуть бути виявлені як точки, що не попадають в жоден кластер, або як точки, що утворюють самостійний кластер. В статті M. H. Nguyen та ін. [14] розглянуто використання кластерного аналізу для виявлення аномальних активностей користувачів у системі автентифікації за допомогою аналізу відхилень від типових поведінкових шаблонів.

Алгоритми регресії: це методи машинного навчання, які використовуються для прогнозування числових значень на основі існуючих даних. Алгоритми регресії можуть бути використані для визначення незвичайно високих або низьких значень відповідних параметрів мережі. Один з прикладів алгоритму регресії, який може бути використаний для виявлення аномалій в кібербезпеці, лінійна регресія. Вона дозволяє визначити високу або низьку кількість запитів до сервера в певний час, що може свідчити про DDoS-атаку.

Відображення типів аномалій до методів їхнього виявлення

Після аналізу літератури стосовно методів виявлення аномалій, огляду кожного методу та побудови класифікації типів аномалій було побудовано їхнє відображення.

На рис. 1 можна побачити перелік типів аномалій та методів виявлення, які б доцільно було використати для ідентифікації кожного з типів аномалій. Вона допоможе вибрати найбільш ефективні методи для виявлення різних типів аномалій залежно від їх характеристик та особливостей.

Висновки

У роботі було розглянуто основні аномалії , які можуть зустрічатися в цифрових інформаційних системах. На основі характерних особливостей аномалій, їх вдалося класифікувати на 11 різних типів. Для кращого забезпечення безпеки інформаційних систем були проаналізовано класифікацію різних методів виявлення аномалій та самі методи. На основі отриманої інформації та розглянутої літератури в результаті отримано таблицю розподілу типів аномалій до різних методів виявлення. В подальшому для швидкої ідентифікації аномалій в часі може бути використана структура Killchain. За допомогою неї, на основі переліку аномалій, буде можливість ідентифікувати фазу атаки, яка відбувається в конкретний момент часу над системою жертви.

Перелік використаних джерел

1. *Varun Chandola Arindam Banerjee V. K.* Anomaly detection: A survey. — 07/30/2009.
2. *Лихошерст В. Р.* Алгоритм класифікації та кластерного аналізу DenStream для вирішення задач з забезпечення інформаційної безпеки. — 2021.

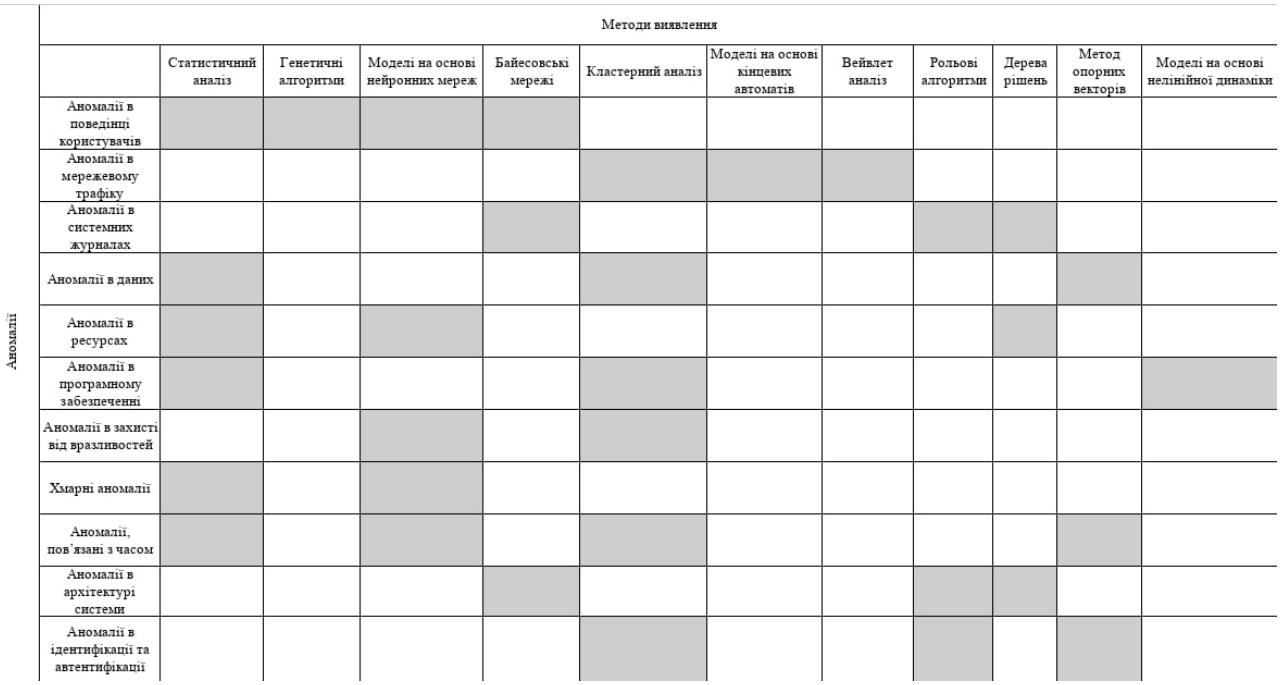

Рис. 1. Візуалізація відношення методів виявлення до типів аномалій

3. *Farooq M. U.*, *Ahamed S. I.* Anomaly detection in network traffic using machine learning techniques: A comprehensive review. // Journal of Network and Computer Applications. — 2018. — Vol. 107. — P. 24–37.
4. Anomaly detection in network traffic based on wavelet analysis / H. Lu, J. Zheng, X. Liu, Y. Zhang // Journal of Network and Computer Applications. — 2011. — Vol. 34. — P. 243–251.
5. Anomaly-based network intrusion detection: Techniques, systems and challenges / P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, E. Vázquez // Computers and Security. — 2009. — Vol. 28. — P. 18–28.
6. *Chandola V.*, *Banerjee A.*, *Kumar V.* A review of unsupervised anomaly detection methods for time series // ACM computing surveys. — 2013. — P. 1–41.
7. *Badr W.* Uncovering Anomalies with Variational Autoencoders (VAE): A Deep Dive into the World of Unsupervised Learning. — 2022.
8. *Huang Y. T.*, *Liao C. H.* Anomaly intrusion detection based on genetic algorithm and statistical analysis for web servers. — 2012.
9. *Khan H.*, *Hussain M.*, *Ullah Z.* Support vector machine based anomaly detection for access control system. — 2017. — C. 161—165.
10. *Mosleh A.*, *Pham V. T.*, *Liu Z.* Anomaly detection in software systems using nonlinear dynamics and LSTM recurrent neural networks. // Journal of Systems and Software. — 2018. — Vol. 135. — P. 65–77.
11. *Li W.*, *Zhang Y.*, *Jiang X.* Anomaly Detection for User Access Control with Role Mining Algorithms. — 2019.
12. *Zhang Y.*, *Wen Y.*, *Li Q.* Anomaly Detection of System Architecture Based on Decision Tree. — 2011.
13. *Han J.*, *Huang. J.* Anomaly detection in system logs using Bayesian networks // Network Operations and Management Symposium (NOMS). — 2012. — P. 1283–1286.
14. *Nguyen M. H.*, *Kang B.*, *Kim D.* Anomaly Detection in Authentication System Using Clustering Analysis. — 2018.