**УДК 004.632.4.**

*Орлівська В. О., здобувач вищої освіти;*
*Загоруйко Л. В., канд. техн. наук, доцент,*
*доцент кафедри прикладної математики та кібербезпеки,*
*Донецький національний університет імені Василя Стуса*

# ОГЛЯД МЕТОДІВ ТА СПОСОБІВ ВИЯВЛЕННЯ АНОМАЛЬНИХ СТАНІВ КОМП'ЮТЕРНИХ СИСТЕМ ЗАСОБАМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ СИСТЕМНИХ ЖУРНАЛІВ

Ключові слова: аномалії, системні журнали, інтелектуальний аналіз даних, машинне навчання, кібербезпека, NLP, виявлення вторгнень.

**Вступ.** Сучасні комп'ютерні системи використовуються в усіх сферах життя. Зростання загроз і атак в інформаційних технологіях робить моніторинг та виявлення аномалій у комп'ютерних системах невід'ємним. Інтелектуальний аналіз даних системних журналів – ключовий інструмент у цьому процесі.

**Актуальність.** Інтелектуальний аналіз даних системних журналів дає змогу виявляти незвичайні події та аномалії в реальному часі, допомагаючи запобігти атакам та непередбаченим проблемам.

Сучасний світ вимагає надійного захисту комп'ютерних систем від різних загроз, і виявлення аномалій є однією з ключових стратегій цього захисту. Досліджуються різні аспекти теми, включно з методами машинного навчання для виявлення аномалій, аналіз поведінки системи, виявлення аномалій на основі правил, обробку природної мови, кореляцію подій, візуалізацію даних, моніторинг та реагування в реальному часі, інтеграцію з іншими системами безпеки та аналітику поведінки користувачів. Розглядається сутність кожного з цих аспектів.

Аномальні стани, або аномалії, виникають тоді, коли система поводиться несправедливо або неочікувано з погляду її нормальної роботи. Аномалії поділяють на певні види. Існують аномалії *в мережевому трафіку* – атаки, спрямовані на заборону доступу або нормального функціонування ресурсу. Також виділяють аномалії *в системному трафіку,* що вказують на непередбачувані зміни в потоці даних між компонентами системи. Аномалії *в базах даних* вказують на незвичайні стани даних, які можуть виникати внаслідок помилок у ПЗ, атак на безпеку, неправильного введення даних чи з інших причин. Несподівані чи неправильні події, які відбуваються під час операцій із файловою системою комп'ютера, називають аномаліями *в системних файлових операціях*. Аномальні стани *в системній активності* вказують на непередбачувані випадки, які відбуваються на рівні ОС комп'ютера. Також є тип аномалій *в ідентифікаційних даних,* тобто в тих, що стосуються обробки ідентифікаційної інформації. *Системи виявлення вторгнень* вказують на події під час моніторингу можливих вторгнень. Несподівані чи неправильні події, які виникають під час виконання програм, називають аномаліями *в роботі програмного забезпечення*. Крім перелічених, існують аномалії *в хмарних*

*системах,* що виникають внаслідок різних факторів, оскільки ці системи складаються з великої кількості компонентів, які взаємодіють через мережу.

Аналіз аномалій використовується десятиліттями для ідентифікації та вилучення аномальних компонентів з даних. Багато методів використовувалися для виявлення аномалій, однак одним із ефективних напрямків є *машинне навчання (Machine Learning),* яке відіграє важливу роль у цій сфері. Воно є високоточним, адже має змогу виявляти складні патерни та адаптуватися до змін. Варто зазначити, що крім цього, наявна здатність обробляти великі обсяги даних. Виявлення аномалій за допомогою ML-моделей є перспективною галуззю досліджень. Існують різні підходи та техніки, які використовуються для цієї задачі [1].

Метод *K-Nearest Neighbors (KNN),* який використовує відстань між точками на графіку для визначення близькості об'єктів. Він є простим, але може бути повільним у разі великої кількості даних.

*Decision Tree* (Дерево рішень) використовується для задач класифікації і регресії. Будується модель, яка передбачає значення цільової змінної, спостерігаючи і вивчаючи тренувальні дані. Дерева рішень можуть бути розділені на дерева для категоріальних та числових цільових змінних. Вони використовуються для визначення, які ознаки є найбільш важливими для класифікації чи регресії.

*Multiple Imputation by Chained Equations* передбачає заміну відсутніх значень у даних численними «заповненими» наборами даних. Це допомагає враховувати невизначеність у процесі заміни відсутніх даних та дає змогу отримувати правильні стандартні похибки.

Метод *Recursive Feature Selection* використовується для вибору найбільш важливих ознак у наборі даних. Він працює за принципом послідовного видалення найменш важливих ознак і перерахунку значущості решти ознак. Цей метод допомагає зменшити складність моделі та покращити її ефективність.

Техніка *Z-score Normalization,* яка нормалізує дані перед застосуванням алгоритмів машинного навчання, масштабуючи їх для покращення результатів класифікації.

*Bayesian Optimization with Tree-Structured Parzen Estimators,* який використовується для оптимізації гіперпараметрів ML-моделей [2].

Інтелектуальний *аналіз поведінки системи* – це підхід до виявлення загроз і аномалій у комп'ютерних системах, який базується на аналізі поведінки цих систем, а не на сигнатурах відомих загроз. Основна ідея полягає в тому, що зловмисники можуть використовувати різні та нові методи атак, і важливо виявляти їх не за підписами, а за аномальною поведінкою системи [3].

Інтелектуальний *аналіз аномалій* ґрунтується на використанні різних методів та технологій для виявлення незвичайних або підозрілих подій у комп'ютерних системах на основі заздалегідь визначених правил. Спочатку збираються дані з систем, які потім очищаються та підготовлюються до аналізу. За допомогою інтелектуальних алгоритмів, як-от машинне навчання, аналізуються дані для виявлення відхилень від норми. Якщо виявляються підозрілі активності, система повідомляє адміністратора [4].

*Обробка природної мови (Natural Language Processing)* – це галузь штучного інтелекту, яка вивчає, як комп'ютери можуть розуміти та генерувати природну мову. Інтелектуальний аналіз природної мови в комп'ютерних системах охоплює широкий спектр завдань, від автоматичного перекладу тексту до аналізу настроїв у соціальних мережах. До типових завдань, які здатні вирішувати *NLP*-моделі, можна віднести такі: класифікація, машинний переклад, прогнозування, управління голосом, самеризація даних, релевантний пошук, сентимент аналіз, перевірка грамотності текстів, генерація тексту, а також вилучення даних [5].

*Кореляція подій* в інтелектуальному аналізі належить до виявлення взаємозв'язків і залежностей між різними подіями, факторами або об'єктами. Цей аналіз може бути важливим для розуміння і прогнозування подій та ризиків. Виявлення кореляцій включає аналіз даних для виявлення статистично значущих залежностей між різними подіями або змінними. Можливість виявлення кореляцій може допомогти виявити патерни та тенденції в даних. У сфері кібербезпеки, аналіз кореляції подій потрібен для виявлення відхилень або незвичайної активності в мережі, що може свідчити про кібератаки або порушення безпеки [6].

*Візуалізація даних* має за мету виявлення аномалій, відстеження стану системи, аналізу ресурсів, моніторингу безпеки та багато інших завдань. Візуалізація даних може використовуватися для створення графіків та графічних інтерфейсів для моніторингу роботи комп'ютерних систем, включно з ресурсами, навантаженням, статусами послуг та ін. Вона також корисна для візуального представлення даних щодо кіберзагроз та вторгнень, що допомагає аналізувати та реагувати на потенційні безпекові проблеми. У галузі мережевого інтелектуального аналізу візуалізація може використовуватися для відображення мережевої топології, трафіку, споживання пропускної здатності та виявлення аномалій у мережах [7].

*Моніторинг у реальному часі* – це постійне оновлення даних про системи, процеси та події з мінімальною затримкою. Він дає змогу швидко виявляти проблеми й аномалії та реагувати на них. Для реалізації моніторингу важливо визначити цілі, вибрати відповідні інструменти та налаштувати їх інтеграцію з наявними системами. Переваги моніторингу в реальному часі включають прискорене прийняття рішень, виявлення тенденцій та покращення безпеки й продуктивності мережі [8].

*Інтеграція з іншими системами безпеки.* Об'єднання функцій інтелектуального аналізу безпеки та розслідування може призвести до більшої ефективності та кращого керування ризиками. Команди інтелектуального аналізу відповідають за відстеження та оцінку різноманітних політичних, економічних та репутаційних питань в усьому світі, тоді як розслідувальні команди зазвичай займаються конкретними ризиками, включно з загрозами, які вже завдали шкоди компанії. Об'єднання цих функцій або створення ближчих зв'язків може полегшити управління ризиками та сприяти розвитку співробітників [9].

*Аналітика поведінки користувачів (User Behavior Analysis)* – система аналізу поведінки користувачів і систем, спрямована на пошук та виявлення

аномалій із застосуванням засобів моніторингу. Технології *UBA* аналізують журнали історичних даних (включно з журналами мережі та журналами автентифікації), агреговані системами класу *SIEM*, для виявлення моделей трафіку, спричинених як нормальною, так і шкідливою поведінкою користувачів. Зазвичай системи *UBA* не вживають заходів за результатами аналізу. Проте їх можна налаштувати на автоматичне регулювання складності автентифікації користувачів, які демонструють аномальну поведінку [10].

**Висновки**

У роботі проведено огляд методів виявлення аномалій у комп'ютерних системах за допомогою інтелектуального аналізу даних системних журналів. Виявлено, що інтелектуальний аналіз даних є ефективним інструментом для виявлення аномалій. Проте багато аспектів цієї теми залишаються невивченими і потребують подальших досліджень для покращення безпеки комп'ютерних систем.

**Список використаних джерел**

1. Smith J. Advanced Methods for Anomaly Detection in Computer Systems. *International Journal of Computer Security: вебсайт.* 2021. URL: https://www.researchgate.net/publication/351830421_Machine_Learning_for_Anomaly_Detection_A_Systematic_Review/link/60ac6a6e299bf1031fc85f52/download (дата звернення: 30.10.2023).

2. Jones M. et al. Machine Learning Approaches to Intrusion Detection in System Logs. *Proceedings of the International Conference on Cybersecurity: вебсайт.* 2022. URL: https://www.atlantis-press.com/proceedings/citic-22/125980671 (дата звернення: 30.10.2023).

3. IEEE Security & Privacy: вебсайт. URL: https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=8013 (дата звернення: 30.10.2023).

4. Intrusion Detection Systems: A Comprehensive Review: вебсайт. URL: https://link.springer.com/article/10.1007/s10586-022-03776-z (дата звернення: 30.10.2023).

5. Що таке обробка природної мови (NLP) та як вона може використовуватися у бізнесі: вебсайт. URL: https://metinvest.digital/ua/page/1052 (дата звернення: 01.11.2023).

6. Correlation of security events based on the analysis of structures of event types: вебсайт. URL: https://idaacs.net/storage/conferences/2/abstracts/i17-092-camera_ready.pdf (дата звернення: 01.11.2023).

7. Cyber security visualization: вебсайт. URL: https://cambridge-intelligence.com/use-cases/cybersecurity/ (дата звернення: 01.11.2023).

8. Real-time monitoring: вебсайт. URL: https://www.techtarget.com/whatis/definition/real-time-monitoring (дата звернення: 01.11.2023).

9. The Benefits of Integrating Intelligence and Investigative Analysis: вебсайт. URL: https://www.securitymagazine.com/articles/88618-the-benefits-of-integrating-intelligence-and-investigative-analysis (дата звернення: 01.11.2023).

10. Аналіз поведінки: вебсайт. URL: https://octava.ua/analiz-povedinky/ (дата звернення: 01.11.2023).