УДК 004.056.5:004.5
DOI: 10.31673/2409-7292.2023.030101

Савченко В. А., Смолєв Є. С., Гамза Д. Є.

МЕТОДИКА ВИЯВЛЕННЯ АНОМАЛІЙ ВЗАЄМОДІЇ КОРИСТУВАЧІВ З ІНФОРМАЦІЙНИМИ РЕСУРСАМИ ОРГАНІЗАЦІЇ

У статті розглянуто проблему виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації. Показано, що взаємодія користувачів з інформаційними ресурсами стає ключовим фактором ефективності та безпеки. Більшість з способів виявлення аномалій базується на аналізі різних технічних та інструментальних показників, таких як мережева активність, використання периферійних пристроїв, завантаженість системи, інтенсивність взаємодії з інформаційними системами та ін. Сучасні системи виявлення вторгнень (IDS – Intrusion Detection System) дозволяють здійснювати виявлення атак в реальному часі на основі бази даних шаблонів (сигнатур) атак, методів машинного навчання та множини даних, що характеризують взаємодію працівників з інформаційними активами організації. Разом з тим, більшість методів машинного навчання є достатньо складними для оперативної реалізації і не дозволяють приймати однозначні рішення щодо наявності аномалій. У роботі пропонується методика виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації, яка дозволяє використовувати результати застосування сучасних систем виявлення вторгнень і є достатньо простою для практичної реалізації адміністраторами інформаційної безпеки. Методика базується на використанні дводольного графа для відображення взаємодії користувачів (працівників організації) з активами (інформаційними системами) на підставі даних мережі, що збираються системою IDS. Результати моделювання за наведеними прикладами свідчать, що така методика є достатньо чутливою до різної активності користувачів. Методика дозволяє визначити, що взаємодія користувача з якимось інформаційним активом організації носить аномальний характер. Дані можуть бути передані адміністратору інформаційної безпеки для подальшого аналізу.

Ключові слова: інформаційний ресурс, аномалія взаємодії, виявлення вторгнень, інформаційна безпека, кібербезпека.

Вступ

В сучасному цифровому світі, де доступ до інформації стає все більш важливим для успішної діяльності організацій, взаємодія користувачів з інформаційними ресурсами стає ключовим фактором ефективності та безпеки. Організації вкладають значні зусилля в розробку та підтримку систем, що забезпечують доступ до даних та ресурсів для своїх співробітників та клієнтів. Однак виявлення аномалій в цій взаємодії може мати значний вплив на безпеку та функціональність цих систем. Аномалії взаємодії з інформаційними ресурсами організації можуть включати широкий спектр подій, від незвичайної активності користувачів до потенційних кібератак або порушень безпеки. Розуміння, виявлення та вчасна реакція на такі аномалії стають критичними для забезпечення надійності інформаційних систем.

У даному дослідженні ми розглянемо методику визначення аномалій взаємодії користувачів з інформаційними ресурсами організації. Метою є розкриття ключових аспектів цього процесу, включаючи типи аномалій, методи їх виявлення та практичне застосування в реальних сценаріях. Дослідження спрямоване на надання інсайтів, які допоможуть підвищити безпеку та ефективність інформаційних систем організацій через виявлення та вирішення аномалій в їх взаємодії з користувачами.

Постановка проблеми

Загальна проблема виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації полягає в тому, що це завдання є складним і вимагає комплексного підходу через декілька ключових причин:

Різноманітність аномалій: Аномальна поведінка користувачів може приймати різні форми, включаючи нестандартні паттерни доступу до інформації, незвичайну активність, спроби несанкціонованого доступу, або навіть інсайдерські загрози. Різноманітність цих аномалій ускладнює їх виявлення. Організації мають величезні обсяги даних, що генеруються внаслідок взаємодії користувачів з інформаційними системами. Аналіз цих великих обсягів даних для виявлення аномалій вимагає потужних інструментів обробки та аналізу. Поведінка

користувачів та структура інформаційних систем можуть змінюватись з часом. Що було нормальним вчора, може вже стати аномалією сьогодні. Потрібна система, яка може адаптуватися до змін в середовищі.

Неоднорідність даних: Дані про взаємодію користувачів можуть бути представлені у різних форматах та джерелах. Їх об'єднання та обробка для виявлення аномалій може бути складною через різницю в структурах та типах даних. Також, виявлення аномалій вимагає високої точності, оскільки неправильна інтерпретація може призвести до помилкової класифікації нормальної поведінки як аномальної або навпаки.

Забезпечення приватності: Під час виявлення аномалій необхідно зберігати конфіденційність та приватність даних користувачів, що може ускладнювати реалізацію деяких методів аналізу.

Оскільки ці проблеми є складними та різноманітними, виявлення аномалій взаємодії користувачів з інформаційними ресурсами вимагає використання різноманітних методів аналізу даних, машинного навчання, та розвитку спеціалізованих систем для ефективного вирішення цієї проблеми.

Аналіз наукових публікацій

Виявлення аномалій – напрямок, який з кожним роком стає все більш актуальним. Існують різні способи виявлення аномалій діяльності користувачів інформаційних систем. Більшість з них базується на аналізі різних технічних та інструментальних показників, таких як мережева активність, використання периферійних пристроїв, завантаженість системи, інтенсивність взаємодії з інформаційними системами та ін.

У статті [1] подано вичерпний огляд існуючої літератури, в якій розглядаються останні досягнення в методах виявлення аномалій для виявлення загроз безпеці в кіберфізичних системах. В статті [2] запропоновані підходи, щодо класифікації методів виявлення аномалій в сучасних системах виявлення атак. Показано, що методи виявлення аномалій в сучасних системах виявлення атак недостатньо опрацьовані в частині формальної моделі атаки, а, отже, для них досить складно суворо оцінити такі властивості як обчислювальну складність, коректність, завершеність. В роботі [3] методи виявлення аномалій було оцінено з аспекту їх можливості застосування до різних систем з мінімізацією вкладу користувача. Отримані результати показують, що найефективнішим методом виявлення аномалій, котрий може бути перенесений на різні системи та мінімізує роботу користувача є системи засновані на машинному навчанні.

Публікація [4] визначає три основні методологічні області для діагностування аномалій (машинне навчання, глибоке навчання, статистичні підходи) і підсумовує, як саме відповідні моделі застосовуються для виявлення аномалій. Крім того, автори пояснюють, які конкретні області застосування зазвичай розглядаються шляхом виявлення аномалій у контексті хмарних обчислювальних середовищ і які відповідні загальнодоступні набори даних часто використовуються для оцінювання. В роботі [5] запропоновано інтелектуальну систему виявлення аномалій та ідентифікації пристроїв розумних будинків із застосуванням колективної комунікації. Концепція роботи системи заснована на отриманні виграшу від об’єднання розумних будинків в соціальну мережу в частині підвищення безпеки як окремо взятого розумного будинку, так і всієї соціальної мережі поєднаних розумних будинків. Публікація [6] пропонує неконтрольований метод, який був розроблений для виявлення аномалій, коли інформація не позначена або класифікована. Були використані підходи з видобутку інформації на основі машинного навчання, розробленого для реалізації системи виявлення аномалій.

Сучасні системи виявлення вторгнень (IDS – Intrusion Detection System) впроваджуються для виявлення атак в реальному часі, використовують базу даних шаблонів (сигнатур) атак, методи машинного навчання, можуть реєструвати множину даних, що характеризують

© Савченко, В. А., Смолєв, Є. С., & Гамза, Д. Є. (2023). Методика виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації. Сучасний захист інформації, 4(56), 6–12. https://doi.org/10.31673/2409-7292.2023.030101.

взаємодію працівників з інформаційними активами організації, і добре зарекомендували себе при вирішенні задачі виявлення аномалій.

У статті [7] описується дослідження лог-майнінгу в області технологій мікросервісів з виявленням аномалій з журналів, тобто подій, які вимагають більш глибокої перевірки аналітиками. Автори пропонують новий підхід до пошуку числових представлень комп'ютерних журналів, не роблячи припущень щодо формату базових даних і не вимагаючи знання програм. У статті [8] представлено розподілений підхід для виявлення аномалій у реальному часі у великомасштабних середовищах. Метод має можливість виявлення послідовних та кількісних аномалій в межах а багатоджерельного потокового журналу.

Мета статті – розробити методику виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації, яка дозволяла б використовувати результати застосування сучасних систем виявлення вторгнень (IDS) і була б достатньо простою для практичної реалізації адміністраторами інформаційної безпеки.

Методика виявлення аномалій при взаємодії користувачів з інформаційними активами організації. Існують різні підходи до аналізу дій користувачів і виявлення відхилень від нормальної поведінки. У даній роботі пропонується методика, яка базується на використанні дводольного графа [9] для відображення взаємодії користувачів (працівників організації) з активами (інформаційними системами) на підставі даних мережі, що збираються системою IDS.

Множину користувачів позначимо як $U = \{u_1,..., u_n\}$, інформаційні активи визначимо множиною $A = \{a_1,..., a_m\}$, а набір користувачів, які зверталися до активів a_i за якийсь певний період часу, множиною U_{A_i}. Позначимо як G_{A_i} – повний граф U_{A_i}, де значенням ваги між парами вершин виступає величина подібності.

Дводольний граф, що відображає факт звернення користувачів до активів, позначимо двійковою матрицею A_U. При цьому $A_U(i, j) = 1$, якщо користувач u_i здійснює доступ до активу a_i, і $A_U(i, j) = 0$, якщо ні. Для оцінки зв'язності користувачів з активами пропонується використовувати статистичну міру IDF (Inverse Document Frequency). У якості міри I_{DF} пропонується взяти сигмоїдальну функцію у вигляді:

$$I_{DF}(U_i) = \frac{1}{1 + e^{\frac{\gamma}{2} - \frac{\gamma \cdot E \times U_i}{|A|}}} \tag{1}$$

де E = (1,1,...,1) – одиничний вектор розмірності m;
$|A|$ – потужність множини A;
U_i – стовпець i-користувача матриці As (вектор доступу);
γ – коефіцієнт чутливості функції.

Отриману після трансформації матрицю позначимо I_{DF}^{UA}. Подібність між парами користувачів може бути отримано на основі їх векторів доступу. Для вимірювання схожості двох векторних активів пропонується використовувати косинусну подібність [10]:

$$C(u_i, u_j) = \frac{I_{DF}(U_i) \times I_{DF}(U_j)}{\|I_{DF}(U_i)\| \times \|I_{DF}(U_j)\|} = \frac{\sum_{k=1}^{m} I_{DF}(U_{i,k}) \times I_{DF}(U_{j,k})}{\sqrt{\sum_{i=1}^{m} (I_{DF}(U_{i,k}))^2} \times \sqrt{\sum_{i=1}^{m} (I_{DF}(U_{j,k}))^2}}. \tag{2}$$

Якщо дано два вектори ознак X і Y, то косинусна схожість може бути представлена, використовуючи скалярний добуток і норму. У разі взаємодії користувача з інформаційними

© Савченко, В. А., Смолєв, Є. С., & Гамза, Д. Є. (2023). Методика виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації. Сучасний захист інформації, 4(56), 6–12. https://doi.org/10.31673/2409-7292.2023.030101.

активами організації косинусна схожість двох користувачів змінюється в діапазоні від 0 до 1, оскільки кут між двома векторами частоти не може бути більше, ніж 90°. Косинусна схожість є ефективною у якості оціночної міри, особливо для розріджених векторів, так як при цьому враховуються лише ненульові значення [10].

В результаті обчислень буде отримано матрицю подібності взаємодії користувачів з інформаційними активами. Передбачається, що в разі, якщо один з користувачів є зловмисником, то його дії знайдуть своє відображення на матриці подібності. Навколо кожного активу утворюється індивідуальна група користувачів, які з ним працюють і до нього звертаються. Для обчислення подібності між групами користувачів необхідно обчислити середню схожість між усіма парами користувачів (загальна подібність користувачів):

$$C\left(G_{A_k}\right)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}C\left(U_i,U_j\right)}{\left|U_{A_k}\right|\times\frac{U_{A_k}-1}{2}},\forall U_i\neq U_j\in U_{A_k}\forall U_j, \tag{3}$$

де $\left|U_{A_k}\right|$ – кількість користувачів в групі.

Якщо $C\left(G_{A_k}\right)$ має високе значення, то це означає, що користувачі мають сильну взаємодію щодо активу a_k. Для виявлення аномальних дій користувача необхідно визначити середню подібність для підгрупи $C\left(G_{A_j}\right)\forall i\vee j=k$, у якій окремий користувач k порівнюється з іншими користувачами, та визначити рейтинг цього користувача відносно середнього значення по організації:

$$R\left(u_k,A\right)=\frac{C\left(G_{U_k}\right)-C\left(G_{A_k}\right)}{C\left(G_{A_k}\right)}\cdot 100\%,k=1,\ldots,m. \tag{4}$$

де $C\left(G_{U_k}\right)$ – підгрупа користувачів, які порівнюються з користувачем u_j. Чим більше значення $R\left(u_k,A\right)$, тим імовірніше, що доступ користувача u_j до активів a_i є аномальним.

Запропонована методика виявлення аномальних дій користувачів на основі аналізу мережевих даних може бути представлена у вигляді послідовності кроків.

1. Побудова множин користувачів і активів.
2. Побудова дводольного графа взаємодії.
3. Обчислення статистичної міри IDF.
4. Обчислення матриці схожості дій користувачів.
5. Обчислення загальної схожості дій користувача.
6. Виявлення аномальних дій.

Алгоритм виявлення аномалій при взаємодії користувачів з інформаційними активами організації

1. $A_U\leftarrow\begin{Vmatrix} a_{u_{1,1}} & \ldots & a_{u_{1,n}} \\ \ldots & \ldots & \ldots \\ a_{u_{m,1}} & \ldots & a_{u_{m,n}} \end{Vmatrix}$ – формування матриці доступу користувачів до активів;
2. $\gamma\leftarrow$ – введення значення чутливості алгоритму;

3. $m = Length\left[A_U\right]$ – визначення кількості активів;
4. $n = Length\left[A_{U_1}\right]$ – визначення кількості користувачів;
5. $E \leftarrow (1,\ldots,1_m)$ – формування одиничного вектора;
6. $I_{DF}(U) = Table\left[\frac{1}{1+e^{\frac{\gamma}{2}-\frac{\gamma \cdot E \times Table\left[\left\{a_{u_{i,j}}\right\},\{m\}\right]}{m}}},\{j,n\}\right]$ – визначення параметра IDF;
7. $I_{DF}\left(A_U\right) = Table\left[If[a_{u_{i,j}} = 1, I_{DF}(U)_i, 0],\{i,n\},\{j,m\}\right]$ – підстановка значень $I_{DF}(U)$ в матрицю сигмоїдальних функцій;
8. $C\left(u_i,u_j\right) = Table\left[\frac{\sum_{k=1}^{m} I_{DF}\left(A_U\right)_{k,i} \cdot I_{DF}\left(A_U\right)_{k,j}}{\sqrt{\sum_{k=1}^{m} I_{DF}\left(A_U\right)_{k,i}^2} \cdot \sqrt{\sum_{k=1}^{m} I_{DF}\left(A_U\right)_{k,j}^2}},\{i,n\},\{j,n\}\right]$ – обчислення матриці схожості дій користувачів;
9. $C\left(G_{U_k}\right) = Table\left[\frac{1}{n-1} \cdot \sum_{j=1}^{n} If\left[i = j, 0, C\left(u_i,u_j\right)\right],\{i,n\}\right]$ – визначення подібності дій окремих користувачів;
10. $\bar{C}\left(G_{U_k}\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} C\left(G_{U_k}\right)_i$ – визначення середнього значення подібності дій користувачів;
11. $R\left(U_k,A\right) = Table\left[\frac{C\left(G_{U_k}\right)_i - \bar{C}\left(G_{U_k}\right)}{\bar{C}\left(G_{U_k}\right)} \cdot 100\%,\{i,n\}\right]$ – виявлення аномальності користувачів.

Приклад застосування методики. Розглянемо дію методики на абстрактному прикладі. Припустимо, що в організації є 10 користувачів та 15 інформаційних активів. Тоді дводольний граф взаємодій користувачів з інформаційними активами можна описати бінарною матрицею A_S розмірністю 15×10. Розглянемо основні сценарії застосування методики.

Сценарій 1. Дві групи користувачів працюють з виділеними інформаційними активами і їх дії щодо доступу до активів не перетинаються. Такий сценарій описується матрицею (рис. 1.а), а результатом застосування алгоритму буде графік (рис. 1.б). Як бачимо, така ситуація дає «чистий» графік, на якому аномалії не відслідковуються.

Сценарій 2. На відміну від попереднього сценарію один з користувачів (№5) робить спроби проникнути до активу №8 (матриця на рис. 1.в). У такому випадку застосування алгоритму одразу ж дає результат при якому аномальність поведінки користувача №5 становить 7.2% на фоні решти користувачів, у яких ступінь аномальності коливається від –3.4% до +1.3% (рис. 1.г).

Сценарій 3. У більш довільних ситуаціях, як наприклад на рис. 1.д, користувачі, які працюють в окремих групах, і залучають декілька груп активів, алгоритм також дає можливість визначити аномальну поведінку окремих користувачів. У даному випадку це №4 та №7.

Результати моделювання за наведеними прикладами свідчать, що така методика є достатньо чутливою до різної активності користувачів. Від'ємні значення вказують на те, що активність користувача є нижчою за середню по організації (групі користувачів), а додатні відхилення – на підозрілу активність користувача.

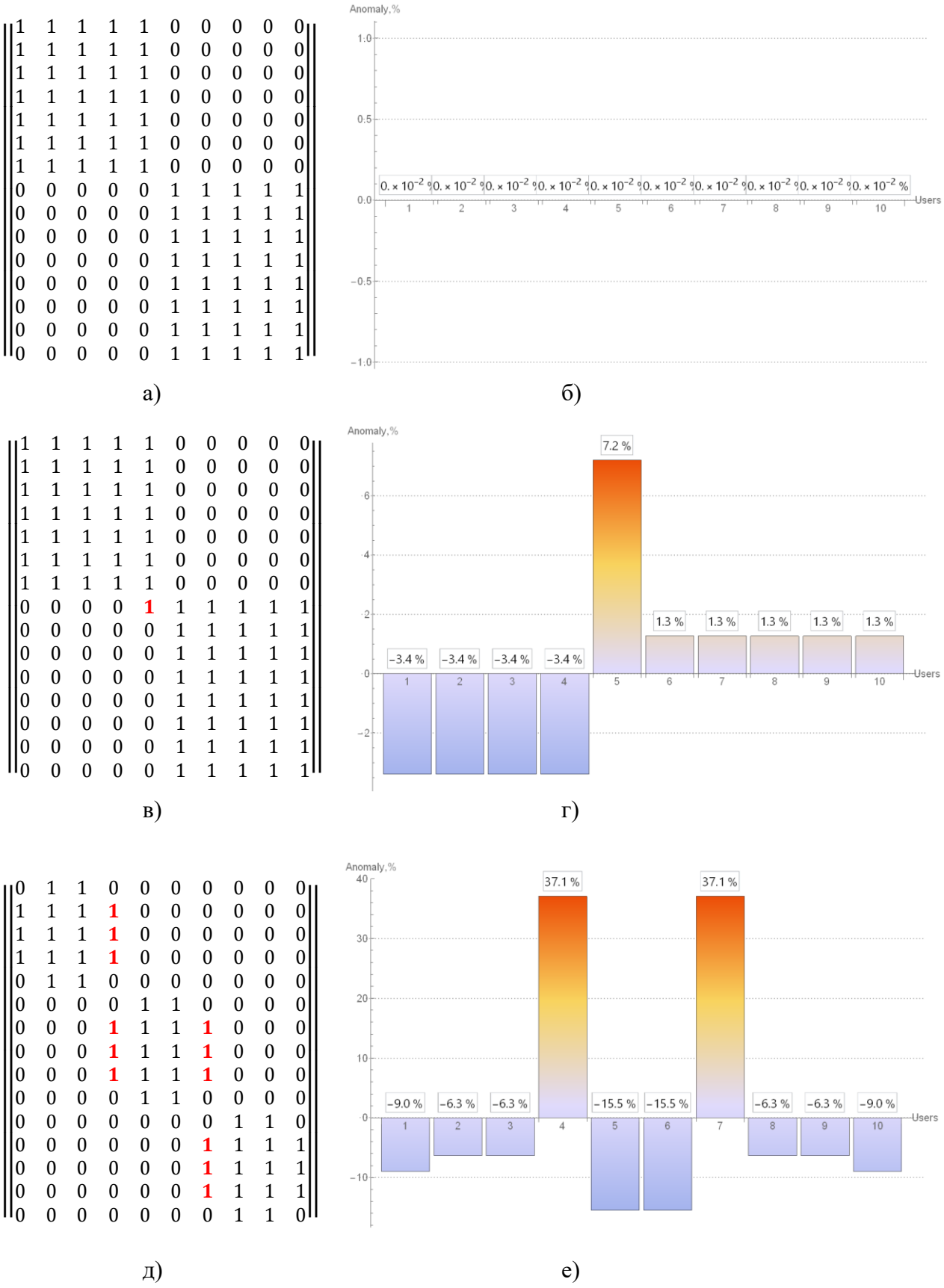

Рис. 1. Приклади застосування алгоритму виявлення аномалій при взаємодії користувачів з інформаційними активами

© Савченко, В. А., Смолєв, Є. С., & Гамза, Д. Є. (2023). Методика виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації. Сучасний захист інформації, 4(56), 6–12. https://doi.org/10.31673/2409-7292.2023.030101.

Висновки

Наведена методика дозволяє визначити, що взаємодія користувача з якимось інформаційним активом організації носить аномальний характер. Дані можуть бути передані адміністратору інформаційної безпеки для подальшого аналізу. Тим не менш, цей підхід не дозволяє достовірно визначити, чи є дана активність зловмисною діяльністю, так як такий аналіз не враховує контекст взаємодії і причину його виникнення, крім того не враховуються інші, особисті особливості конкретного користувача. Наприклад, цілком можливо, що подібна взаємодія відбувається в рамках виконання термінового доручення керівництва. Застосування даної методики доцільно в сукупності з аналізом інших показників, що дозволяють визначити наявність у користувача схильності до зловмисної діяльності, наприклад, з урахуванням лояльності персоналу.

Перелік посилань

1. Jeffrey, N.; Tan, Q.; Villar, J.R. A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems. Electronics 2023, 12, 3283. https://doi.org/10.3390/electronics12153283
2. Рубан, І.В., Мартовицький, В.О., Партика, С.О. Класифікація методів виявлення аномалій в інформаційних системах. Системи озброєння і військова техніка, 2016, № 3(47). – С. 100–105.
3. Горбенко, В.О. Методи виявлення аномалій поведінки користувача в інформаційних системах / В.О. Горбенко, В.М. Ткач // Теоретичні і прикладні проблеми фізики, математики та інформатики: матеріали XVI Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, 26-27 квітня 2018 року, м. Київ / КПІ ім. Ігоря Сікорського, ФТІ. – Київ: ВПІ ВПК «ПОЛІТЕХНІКА», 2018. – С. 51-52. – Бібліогр.: 4 назви. https://ela.kpi.ua/handle/123456789/25237
4. Hagemann, T., and Katsarou, K. 2020. A Systematic Review on Anomaly Detection for Cloud Computing Environments. In 2020 3rd Artificial Intelligence and Cloud Computing Conference (AICCC 2020), December 18–20, 2020, Kyoto, Japan. ACM, New York, NY, USA 14. https://doi.org/10.1145/3442536.3442550
5. Нічепорук, А.О., Нічепорук, А.А., Савенко, О.С., Казанцев, А.Д. Інтелектуальна система виявлення аномалій та ідентифікації пристроїв розумних будинків із застосуванням колективної комунікації. Електротехнічні та комп'ютерні системи. 2021. № 34 (110). 50-61.
6. Mezones Santana, H.L., Cobeña Macias, T.E., Quimiz Moreira, M.A. (2022). Anomaly Detection Method in Computer Systems by Means of Machine Learning. In: Zambrano Vizuete, M., Botto-Tobar, M., Diaz Cadena, A., Durakovic, B. (eds) Innovation and Research - A Driving Force for Socio-Econo-Technological Development. CI3 2021. Lecture Notes in Networks and Systems, vol 511. Springer, Cham. https://doi.org/10.1007/978-3-031-11438-0_32
7. Cinque, M., Della Corte R., Pecchia A., Micro2vec: Anomaly detection in microservices systems by mining numeric representations of computer logs, Journal of Network and Computer Applications, Volume 208, 2022, 103515, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2022.103515.
8. Vervaet, A. MoniLog: An Automated Log-Based Anomaly Detection System for Cloud Computing Infrastructures. IEEE 37th International Conference on Data Engineering (ICDE), 2021. ffhal04078128
9. Поляничко, М.А. Методика виявлення аномальної взаємодії користувачів з інформаційними активами виявлення інсайдерської діяльності // Праці навчальних закладів зв'язку. 2020. Т. 6. № 1. С. 94–98. https://doi.org/10.31854/1813-324X-2020-6-1-94-98
10. Singhal, Amit (2001). Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4): 35–43. http://singhal.info/ieee2001.pdf

Надійшла 08.09.2023 Рецензент: д.т.н., професор Вишнівський В.В.