УДК 004.056.5:004.65

Ю. В. Стасєв, М. І. Карамушка, А. О. Сич, А. С. Карамушка

САМОНАВЧАЛЬНИЙ МЕТОД ЗАХИСТУ ТА РЕЗЕРВУВАННЯ БАЗ ДАНИХ У КОМПЛЕКСАХ ЗАСОБІВ АВТОМАТИЗАЦІЇ АВТОМАТИЗОВАНИХ СИСТЕМ УПРАВЛІННЯ

Анотація. *Розглянуто самонавчальний метод захисту та резервного копіювання баз даних комплексів засобів автоматизації автоматизованих систем управління. Метод ґрунтується на алгоритмі Isolation Forest, поведінковому профілі користувача та вазі таблиць за критичністю даних. Наведено математичні співвідношення для оцінки аномальності запиту, порогів сповіщення й резервування та умови перенавчання моделі. Показано, що позачергове резервування за сигналом методу не збільшує час між останньою копією та інцидентом.*

Ключові слова: *база даних, захист інформації, резервне копіювання, самонавчання, Isolation Forest.*

Annotation. *A self-learning method for protecting and backing up databases in automation equipment of automated control systems is considered. The method is based on the Isolation Forest algorithm, a user behavioural profile and table criticality weights. Mathematical relations are given for the query anomaly score, the alert and backup thresholds and the model retraining condition. It is shown that an extra backup triggered by the method never increases the time between the last copy and an incident.*

Keywords: *database, information protection, backup, self-learning, Isolation Forest.*

Кібератаки на державні інформаційні системи та сектор безпеки і оборони дедалі інтенсивніші: у 2025 р. CERT-UA опрацювала майже 6000 кіберінцидентів, а кількість ворожих атак зросла на 37 % [1]. У серпні 2026 р. повідомлено про перший повністю автономний кібернапад на державні системи Тайваню з використанням мультиагентної системи на основі штучного інтелекту [2]. Бази даних комплексів засобів автоматизації (КЗА) автоматизованих систем управління (АСУ), де зберігаються планові й розвідувальні відомості, є привабливою ціллю. Традиційні засоби захисту СУБД – автентифікація, розмежування доступу за ролями, шифрування, аудит – або блокують дію за заздалегідь заданим правилом, або лише фіксують її постфактум [3], а резервне копіювання виконується за розкладом незалежно від рівня ризику [4], [5]. Метою роботи є обґрунтування методу, який самонавчається, виявляє аномальні запити та ініціює позачергове резервне копіювання.

Постановка задачі. Кожен запит q до бази даних подається вектором ознак $x(q) = (x_1, \ldots, x_k)$: тип операції, час, обсяг оброблених даних, важливість таблиці та звичність звернення для користувача. Потрібно побудувати функцію f: $x \to \{0, 1\}$, де 0 – нормальна, а 1 – аномальна активність, яка періодично уточнюється на нових даних без ручного перенавчання, а її результат слугує підставою для рішення про резервування. Штатні засоби PostgreSQL такої функції не виконують [3]. Раніше запропонований прототип детектора на основі наївного байєсівського класифікатора порівнює запит лише з очікуваною поведінкою ролі й не адаптується до нових даних [6].

Метод складається з п'яти етапів: збір даних із журналу запитів, формування ознак, оцінювання, рішення та самонавчання. Кожній таблиці j адміністратор ставить у відповідність вагу w_j (наприклад, від 1 до 3), а чутливість запиту визначається як

$$x_s(q) = x_{op}(q) \cdot w_{jj} \tag{1}$$

де $x_{op}(q)$ – коефіцієнт типу операції (наприклад, SELECT – 1, DELETE – 3). Поведінковий профіль користувача u – частка його запитів до таблиці j:

$$p_{u,jj} = \frac{n_{u,jj}}{n_u} \tag{2}$$

де $n_{u,j}$ – кількість запитів користувача u до таблиці j, n_u – загальна кількість його запитів. Якщо $p_{u,j} = 0$, запит одразу виглядає незвичним [7]. Оскільки Isolation Forest обирає порогові значення випадково між мінімумом і максимумом ознаки, множення ознаки на сталу не змінює результат; тому вага вводиться окремими ознаками w_j і x_s.

Оцінювання виконується алгоритмом Isolation Forest [8], [9], який будує ансамбль випадкових дерев: аномальні запити ізолюються за меншу кількість розділень, тобто мають меншу довжину шляху $h(x)$. Оцінка аномальності

$$s(x, \psi\psi) = 2^{-\frac{E[h(x)]}{c(\psi\psi)}} \quad (3)$$

де $E[h(x)]$ – середня довжина шляху за всіма деревами, ψ – розмір підвибірки, а

$$c(\psi\psi) = 2H(\psi\psi - 1) - \frac{2(\psi\psi - 1)}{\psi\psi} \quad (4)$$

– середня довжина шляху в бінарному дереві пошуку, $H(i) \approx \ln i + 0{,}5772$ – гармонійне число [9]. Значення $s \to 1$ свідчать про аномалію, $s < 0{,}5$ – про нормальну поведінку. Оскільки кожне з t дерев будується на підвибірці розміру ψ, складність навчання становить $O(t\psi \log \psi)$ і не залежить від розміру журналу [9], що важливо за обмежених ресурсів КЗА АСУ.

Пороги визначаються за стійкою статистикою: за оцінками s_i навчального вікна обчислюють медіану M та медіанне абсолютне відхилення $MAD = \text{median}|s_i - M|$:

$$\theta\theta = M + k \cdot 1{,}4826 \cdot MAD \quad (5)$$

$$\theta\theta_b = M + k_b \cdot 1{,}4826 \cdot MAD \quad (6)$$

де $k_b > k$, а множник 1,4826 узгоджує MAD зі стандартним відхиленням нормального розподілу. На відміну від квантиля, такі пороги слабо залежать від аномалій, що потрапили до вікна навчання. Рішення ухвалюється за таблицею 1.

Таблиця 1 – Рівні рішення методу

Умова	**Дія методу**
$s \le \theta$	Запис у журнал
$\theta < s \le \theta_b$	Сповіщення адміністратору
$s > \theta_b$	Сповіщення та створення точки відновлення

Самонавчання реалізується контролем дрейфу: на кожному вікні з N запитів обчислюється середня оцінка $\bar{s}$. Якщо

$$|ss - ss_0| > \delta\delta \quad (7)$$

де $\bar{s}_0$ – базове значення, а δ – задане порогове значення, модель і профіль перебудовуються на даних поточного вікна, а M, MAD, θ та θ_b обчислюються заново [10]. Розмітка даних і втручання адміністратора не потрібні; кожне перенавчання фіксується в журналі, щоб повільна атака не стала для моделі «звичайною» поведінкою.

Ефективність резервування оцінюють показником RPO – максимально допустимим інтервалом між останньою копією та інцидентом [11]. Нехай $t_{\text{інц}}$ – момент початку інциденту, $t_{\text{план}}$ – момент останньої планової копії, $t_{\text{дод}}$ – останньої додаткової копії, створеної за сигналом методу. Тоді

$$\Delta\Delta = tt_{\text{інц}} - mmmmx(tt_{\text{план}}, tt_{\text{дод}}) \quad (8)$$

За умови лише планового копіювання з періодом T маємо $\Delta \le T$, а за рівномірного розподілу моменту інциденту $E[\Delta] = T/2$. Оскільки $\max(t_{\text{план}}, t_{\text{дод}}) \ge t_{\text{план}}$, додаткові копії не збільшують Δ для жодного інциденту. Ціною є навантаження: між двома додатковими копіями встановлюється мінімальний інтервал τ, тому їх кількість за добу не перевищує 1440/τ (τ у хвилинах), а підвищення k_b зменшує цю кількість.

Висновки. Запропонований метод поєднує виявлення аномальних запитів і резервування, самонавчається під зміну поведінки користувачів та не потребує розмітки даних. Він доповнює, а не замінює штатний захист PostgreSQL, обмежено виявляє дії внутрішнього порушника, схожі на звичайну роботу, і потребує перевірки на реальних журналах запитів у «тіньовому» режимі.

Список використаних джерел:

1. Державна служба спеціального зв'язку та захисту інформації України. (б. д.). CERT-UA у 2025 році опрацювала майже 6000 кіберінцидентів, кількість ворожих атак зросла на 37%. https://cip.gov.ua/ua/news/cert-ua-u-2025-roci-opracyuvala-maizhe-6000-kiberincidentiv-kilkist-vorozhikh-atak-zrosla-na-37

2. Hackers used autonomous AI agents to attack Taiwan. Is this the future of cyberwarfare? (2026,

August 13). CNN Business. https://www.cnn.com/2026/08/13/tech/china-taiwan-ai-agent-cyberattack-intl-hnk

3. Percona. (n.d.). PostgreSQL security: A comprehensive guide to hardening your database. https://www.percona.com/blog/postgresql-database-security-what-you-need-to-know/ PostgreSQL Global Development Group. (n.d.). pg_basebackup. In PostgreSQL documentation. https://www.postgresql.org/docs/current/app-pgbasebackup.html

4. Severalnines. (n.d.). PostgreSQL replication for disaster recovery. https://severalnines.com/blog/postgresql-replication-disaster-recovery/

5. Shebaro, B., Sallam, A., Kamra, A., & Bertino, E. (2013). PostgreSQL anomalous query detector. In Proceedings of the 16th International Conference on Extending Database Technology (EDBT '13) (pp. 741–744). ACM. https://doi.org/10.1145/2452376.2452469

6. Будзинський, О. В. (2025). Метод виявлення вразливостей та автоматизованого реагування в системах захисту корпоративних баз даних. Сучасний захист інформації, 2(62), 180–186.

7. Щавінський, Ю., & Будзинський, О. (2025). Аналіз актуальних проблем безпеки корпоративних баз даних в умовах сучасної інфраструктури та шляхи їх вирішення. Кібербезпека: освіта, наука, техніка, 3(27), 390–405. https://doi.org/10.28925/2663-4023.2025.27.726

8. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17

9. Heigl, M., Anand, K. A., Urmann, A., Fiala, D., Schramm, M., & Hable, R. (2021). On the improvement of the isolation forest algorithm for outlier detection with streaming data. Electronics, 10(13), Article 1534. https://doi.org/10.3390/electronics10131534

10. Splunk. (n.d.). Recovery point objective (RPO) vs. recovery time objective (RTO): What's the difference? https://www.splunk.com/en_us/blog/learn/rpo-vs-rto.html

Стасєв Юрій Володимирович – доктор технічних наук, професор, професор кафедри бойового застосування та експлуатації АСУ Харківського національного університету Повітряних Сил імені Івана Кожедуба, м. Харків, e-mail: Stasev543@gmail.com. ORCID https://orcid.org/0000-0002-3893-0313

Карамушка Микита Ігорович – курсант Харківського національного університету Повітряних Сил імені Івана Кожедуба, м. Харків, e-mail: nikita1karamushka@gmail.com. ORCID https://orcid.org/0009-0004-1098-9366

Сич Анастасія Олегівна – слухач Харківського національного університету Повітряних Сил імені Івана Кожедуба, м. Харків, e-mail: anastasiagukal96@gmail.com. ORCID https://orcid.org/0009-0008-5403-0615

Карамушка Ангеліна Сергіївна – курсант Харківського національного університету Повітряних Сил імені Івана Кожедуба, м. Харків, e-mail: storchaevaa1999@gmail.com.

Yurii Volodymyrovych Stasiev – Doctor of Technical Sciences, Professor, Professor of the Department of Combat Use and Operation of Automated Control Systems, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, e-mail: Stasev543@gmail.com. ORCID https://orcid.org/0000-0002-3893-0313

Mykyta Ihorovych Karamushka – Cadet of Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, e-mail: nikita1karamushka@gmail.com. ORCID https://orcid.org/0009-0004-1098-9366 ***Sych Anastasiia Olegivna*** – Undergraduate of Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, e-mail: anastasiagukal96@gmail.com. ORCID https://orcid.org/0009-0008-5403-0615

Anhelina Serhiivna Karamushka – Cadet of Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, e-mail: storchaevaa1999@gmail.com.