КОМП'ЮТЕРНА ІНЖЕНЕРІЯ

УДК 004.75 DOI: https://doi.org/10.32515/2664-262X.2024.10(41).2.11-22

Є.В. Мелешко, проф., д-р техн. наук, **М.С. Якименко,** доц., канд. ф.-м. наук,
В.В. Міхав, д-р філос., **Я.П. Шуліка,** асп.
Центральноукраїнський національний технічний університет, м. Кропивницький, Україна
e-mail: elismeleshko@gmail.com, m.yakymenko@gmail.com, mihaw.wolodymyr@gmail.com, yar.shulika@gmail.com

Математична модель виявлення аномальних зв'язків між компонентами складної комп'ютерної системи

Об'єктом дослідження є процес виявлення аномалій у високонавантажених складних комп'ютерних системах (ВНСКС). Сфера практичного використання включає ВНСКС, такі як сервери банківських транзакцій, хмарні платформи, де необхідно забезпечити стабільну роботу в умовах великої кількості запитів. Проблема, що вирішується в дослідженні, полягає у відсутності моделей виявлення аномалій у ВНСКС у реальному часі з заданою точністю в умовах обмежених ресурсів. Створено та досліджено математичну модель виявлення аномальних зв'язків між компонентами складної комп'ютерної системи (ВАЗККС). Результати тестування моделі показали наступні показники ефективності: точність – 84%, точність позитивних прогнозів – 87%, повнота – 74%, зважена середня точність (ЗСТ) – 78%. Позитивні результати дослідження пояснюються наступними передумовами, модель ВАЗККС використовує ортогональні векторні функції та проекційні матриці для аналізу аномалій у складній комп'ютерній системі. Що дозволяє будувати просторові розклади, за допомогою яких можна виявляти складні взаємозв'язки між компонентами складної комп'ютерної системи, використовуючи тільки власні вектори та значення. Таким чином, модель може застосовуватися для оперативного аналізу даних та виявлення аномалій в умовах обмежених ресурсів.

високонавантажені складні комп'ютерні системи, виявлення аномалій, математичні моделі, динамічний хаос, складні мережі

Постановка проблеми. Високонавантажені веб-сервіси та складні комп'ютерні системи та мережі становлять ключову частину сучасної інфраструктури, оскільки забезпечують стабільну роботу банківських послуг, комерційних платформ, онлайн-освітніх ресурсів, соціальних мереж та багатьох інших критичних галузей. З огляду на значне навантаження, яке ці системи щоденно витримують, для їхньої стабільної роботи необхідно дотримуватися високих вимог до показників безпеки та надійності.

Одна з головних проблем для таких систем – це виявлення аномалій у реальному часі. Аномалії можуть сигналізувати про несправності у функціонуванні системи, невідповідності в процесах або потенційні кібератаки. Високонавантажені системи є особливо чутливими до навіть незначних збоїв, оскільки вони можуть спричиняти значні затримки або повну недоступність сервісу для багатьох користувачів одночасно, що призводить до фінансових втрат та втрати довіри клієнтів.

Сучасні веб-сервіси стикаються з такими проблемами як спроби DDoS-атак, значні перепади в запитах користувачів, проблеми з підключенням до баз даних, витоки пам'яті, а також вплив несподіваних змін у конфігурації мережі чи обладнання. Тому забезпечення своєчасного виявлення аномалій є критично важливим аспектом безпеки високонавантажених систем. Автоматизація процесу аналізу та своєчасна ідентифікація потенційних проблем в режимі реального часу дозволяють зменшити ризики та мінімізувати час простою.

Високонавантажені системи відрізняються складною динамікою процесів та великою кількістю компонентів, що потребує застосування інструментів для глибокого аналізу взаємодії між ними. Одним з перспективних підходів є використання моделей виявлення аномальних зв'язків між компонентами системи, що дозволяє здійснювати оцінку стабільності поведінки, враховуючи складні залежності між навантаженням на процесор, пам'ять та мережеві компоненти.

Враховуючи ці вимоги, розробка математичної моделі для виявлення аномалій у високонавантажених веб-сервісах має на меті не лише підвищення точності виявлення, але й забезпечення своєчасної реакції на потенційні загрози в умовах обмежених ресурсів. Це дозволяє уникнути значних фінансових втрат та підвищити рівень довіри користувачів до таких сервісів.

Отже, актуальною є задача розробки математичної моделі виявлення аномалій у високонавантажених складних комп'ютерних системах в умовах обмежених ресурсів.

Аналіз останніх досліджень і публікацій. Було проведено аналіз останніх наукових досліджень і публікацій у сфері виявлення аномалій у комп'ютерних системах.

В дослідженні [1] проведено аналіз використання технологій обчислення в пам'яті для прискорення роботи високонавантажених систем глибинного навчання. Особливий акцент зроблено на перспективи покращення енергоефективності і зниження затримок під час обробки великих обсягів даних, що особливо актуально для масштабованих комп'ютерних систем, які обслуговують численні запити одночасно. Автори надають огляд сучасних трендів і прогнозів розвитку технологій. У висновках автори підкреслюють актуальність досліджень високонавантажених складних комп'ютерних систем у різних напрямках, в тому числі у напрямках безпеки та надійності. Однак, у зв'язку з узагальненим характером праці, автори не надають результатів практичних експериментів та не конкретизують досягнуті переваги дослідження.

У статті [2] представлені сучасні тенденції в розробці мережевих систем виявлення вторгнень, що є критично важливими для захисту високонавантажених складних комп'ютерних систем від кіберзагроз. Автори детально описують поточний стан методів виявлення аномалій у мережевих потоках та обговорюють необхідність створення адаптивних і масштабованих рішень для реального часу, що допоможе забезпечити надійну роботу таких систем навіть у пікові навантаження. Проте, залишаються невирішені питання щодо узгодження різних методів виявлення аномалій та адаптації цих методів до нових загроз. Основною причиною цього є відсутність єдиного стандарту для оцінки та порівняння ефективності різних моделей, а також високі вимоги до обчислювальних ресурсів, необхідних для реалізації цих підходів у реальному часі.

Стаття [3] описує основні мережі з'єднань, що використовуються в суперкомп'ютерних системах, такі як NVIDIA InfiniBand, Intel Omni-Path та Cray Slingshot, і представляє їхню еволюцію та тенденції розвитку в умовах зменшення дії закону Мура. Особлива увага приділяється продуктивності мереж а також швидкості реакції на можливі впливи. Автори підкреслюють що ці характеристики є ключовими для підвищення ефективності паралельних обчислень у високопродуктивних комп'ютерах. Це важливо для розуміння майбутніх викликів у розвитку мережевих рішень для високонавантажених систем. Проте, залишається кілька невирішених питань, таких як забезпечення масштабованості цих мереж для подальшого підвищення продуктивності. Пов'язано це з питаннями технічного обмеження, архітектурними особливостями сучасних мережевих рішень і вимогами до

обчислювальних ресурсів.

Аналіз літератури [4–8] показав, що ряд наукових праць присвячено математичному моделюванню процесів у високонавантажених системах з акцентом на можливість захисту даних. Так, наприклад, в статті [4] автори пропонують математичну модель для виявлення аномалій у складних комп'ютерних системах. Автори наголошують на необхідності універсального та науково обґрунтованого підходу до моніторингу поведінки високонавантаженої системи. Дослідження спрямоване на встановлення загального критерію для ідентифікації аномальної активності на основі однорідності вибірок вхідних даних. Покращуючи критерії однорідності вибірки та ізолюючи спостереження, що вказують на аномальну поведінку, пропонована модель спрямована на підвищення надійності та безпеки комп'ютерних систем. Однак модель розрахована тільки на обробку серій з невеликих вибірок. Крім того, модель має високу обчислювальну складність.

В статті [5] пропонуються методи для ймовірнісного аналізу і моделювання динаміки складних систем, що можуть бути застосовані для моделювання високонавантажених складних комп'ютерних систем і аналізу їх надійності та поведінки в умовах аномальних змін. Це відповідає цілям побудови моделей для виявлення аномалій в поведінці систем, особливо з акцентом на моделювання і розуміння динаміки переходів між станами. Автори запропонували математичну модель для обчислення умовних ймовірностей того, що система опиниться в певному стані у конкретний час, за умови, що на початковий момент система була в будь-якому з можливих станів. Але обмеження наданого графового підходу моделювання, що передбачає наявність тільки трьох станів, зменшує практичну цінність запропонованої моделі в умовах ускладнення безпекової ситуації в реальному часі.

В статті [6] проведено математичне моделювання процесу тестування системи на безпеку з використанням перевіреного графового методу моделювання з додаванням елементів теорії невизначеної логіки. Це дало можливість математично формалізувати складний процес і отримати математичні вирази для розрахунку ймовірнісно-часових характеристик. Аналогічний підхід представлено у статті [7]. Але автори моделей не проводили їх верифікацію на високонавантажених складних комп'ютерних системах. Тому можна зробити висновки про доцільність інших підходів математичної формалізації складних технічних систем.

Один з таких підходів описаний в статті [8]. В ній автори рекомендують системи виявлення аномалій математично формалізувати за допомогою основних положень теорії динамічного хаосу. Зокрема, в якості основного показника автори пропонують використовувати показник BDS-статистики (статистика Брока-Дечерта-Шейнкмана). Це сучасний математичний апарат для виявлення аномалій в часових рядах, який може бути використаний при описі показників інформаційних процесів. В цілому математичний апарат теорії динамічного хаосу має перспективи в формалізації саме складних високонавантажених складних комп'ютерних систем. Однак автори цієї статті не провели аналізу складності та швидкості процесу виявлення аномалій.

На основі проведеного аналізу літературних джерел було виявлено кілька критичних невирішених питань, що потребують подальшого дослідження. Перш за все, одним із ключових викликів є підвищення швидкості виявлення аномалій з урахуванням вимог точності для ВНСКС. Це особливо актуально в умовах виконання завдань такими системами в реальному часі. Також залишаються нерозв'язаними питання масштабованості методів виявлення аномалій, які повинні ефективно працювати з великою кількістю даних і змінюватися відповідно до динамічних умов середовища. Відсутність єдиного стандарту для оцінки ефективності різних підходів

ускладнює об'єктивне порівняння результатів, що створює бар'єри для впровадження нових методів у промислових застосуваннях. Але можна відмітити що процеси масштабованості нових методів у великій мірі залежать так само і від швидкості виконання завдань.

Постановка завдання. Метою цієї роботи є розробка математичної моделі виявлення аномальних зв'язків між компонентами складної комп'ютерної системи в умовах обмежених ресурсів на основі положень теорії динамічного хаосу. Це дозволить підвищити швидкість виявлення аномалій у поведінці високонавантажених складних комп'ютерних систем що, в свою чергу, повинно підвищити її безпеку.

Виклад основного матеріалу. Розглянуто динаміку поведінки складної комп'ютерної системи, представлену в тривимірному просторі. Оцінимо три основні характеристики – завантаження пам'яті, процесора та мережевого пристрою. Для цієї оцінки застосуємо основні положення теорії динамічного хаосу [9, 10].

Розглянуто такий випадок, у якому інваріантним атрактором системи є граничний цикл. Цей цикл може задаватися деяким τ-періодичним розв'язком $x=\xi(t)$, де $x_0=\xi(0)$ є фіксованою точкою циклу. Розв'язок на інтервалі $[0, \tau)$ задає природну параметризацію точок циклу: $M=\{\xi(t) \mid 0 \le t \le \tau\}$. Зроблено припущення, що цикл атрактору M є E-стійким. Тоді навколо цього циклу формується стаціонарно розподілений набір випадкових траєкторій складної системи. В такому разі визначення стохастичної чутливості атрактору M відповідно до теорії динамічного хаосу для складної комп'ютерної системи зводиться до побудови та аналізу τ-періодичного розв'язку $W(t)$ матричного виразу:

$$V = F(t)V + VF^T(t) + P(t)S(t)P(t), \tag{1}$$

з T-періодичними коефіцієнтами:

$$\mathrm{F}(\mathrm{t}) = \partial f / \partial x\left(\xi(t)\right),\ S(t) = G(t)G^T(t),$$

$$G(t) = \sigma\left(\xi(t)\right),\ P(t) = P_{\xi(t)}.$$

Матриця $W(t)$, що є функцією стохастичної чутливості циклу M, являється єдиним розв'язком рівняння (1) у просторі Σ симетричних матриць n×n, які визначені і є достатньо гладкими на R^1 з наступними умовами періодичності:

$$\forall t \in R^1 : V(t+\tau) = V(t), \tag{2}$$

та вродженості:

$$\forall t \in R^1 : V(t)r(t) = 0, r(t) = f\left(\xi(t)\right). \tag{3}$$

Для ймовірнісної інтерпретації матриці $W(t)$ розглянемо наступну стохастичну систему:

$$dy = F(t)ydt + P(t)G(t)dw(t). \tag{4}$$

Ця система має періодичний режим, який пов'язаний із розв'язком $\overline{y}(t)$. Отже, шуканим розв'язком $W(t)$ системи (1)–(4) і є коваріаційна матриця випадкового періодичного процесу $\overline{y}(t)$, що відповідає стабільному стану складної системи. Вона є результатом асимптотичної поведінки процесу і забезпечує характеристику зв'язків між

компонентами системи у довгостроковій перспективі.

Випадкові траєкторії лінійної складної системи створюють навколо циклу набір, що лежить у деякому інваріантному околі системи в області U. Нехай PL_t – це гіперплощина, яка ортогональна циклу в точці $\xi(t)$ ($0 \le t \le Q$). Тоді позначимо через U_t окіл точки $\xi(t)$, що знаходиться в PL_T: $U_t = U \cap PL_t$. В такому разі можна зробити припущення, що $U_t \cap U_s = \emptyset$, якщо $t \neq s$.

Ймовірнісний опис набору цих випадкових траєкторій зручно пов'язати з векторною функцією X_t. Значення X_t є точками перетину випадкових траєкторій лінійної складної системи в області U_t. Ймовірнісний розподіл набору траєкторій з часом стабілізується, тому випадкова змінна X_t в околі оюласті U_t має певний стаціонарний розподіл з деякою густиною $\rho_t(x, \varepsilon)$.

Було використано геометричне представлення для дослідження розкиду випадкових траєкторій навколо циклу. На рис. 1 у вигляді зірочок наведено точки перетину випадкових траєкторій складної системи із січною площиною PL_t, ортогональною циклу у точці $\xi(t)$. Коваріація розподілу цих точок задана матрицею $W(t)$, яка є єдиним розв'язком вище наведеної системи рівнянь (1)–(4).

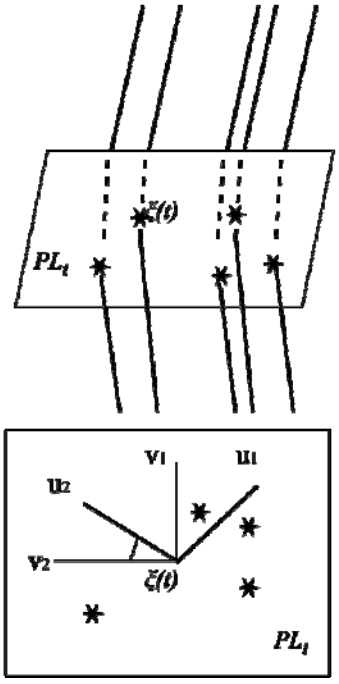


Рисунок 1 – Точки перетину випадкових траєкторій складної системи із гіперплощиною PL_t, ортогональною до циклу в точці $\xi(t)$

Джерело: розроблено авторами

У розглянутому тривимірному просторі (n=3) для побудови рішення $V(t)$ рівняння (1) будемо використовувати наступний сингулярний розклад:

$$V(t) = \lambda_1(t) v_1(t) v_1^T(t) + \lambda_2(t) v_2(t) v_2^T(t) + \lambda_3(t) v_3(t) v_3^T(t),$$

де $\lambda_1(t) \ge \lambda_2(t) \ge \lambda_3(t)$ – власні значення, а $v_1(t)$, $v_2(t)$, $v_3(t)$ – власні вектори матриці $V(t)$. З умови (3) виходить, що для будь-якого t матриця $V(t)$ є виродженою і розподіл точок перетину випадкових траєкторій складної системи (зірки на рис. 1) зосереджений на площині PL_t. А це, в свою чергу, означає, що $\lambda_3(t)=0$ і відповідний власний вектор

$v_3(t)=r(t)/||r(t)||$ є дотичними до циклу. Тому розклад матриці *V(t)* має наступний вигляд:

$$V(t)=\lambda_1(t)v_1(t)v_1^T(t)+\lambda_2(t)v_2(t)v_2^T(t). \tag{5}$$

Тут *V(t)* задається векторами $v_1(t)$, $v_2(t)$ та скалярними функціями $\lambda_1(t)$, $\lambda_2(t)$. Функції $\lambda_1(t)$, $\lambda_2(t)$ у випадку невироджених шумів строго позитивні і визначають при будь-якому *t* дисперсію випадкових траєкторій циклу вздовж векторів $v_1(t)$, $v_2(t)$. Значення $\lambda_1(t)$, $\lambda_2(t)$ задають розмір, а $v_1(t)$, $v_2(t)$ – напрямок осей еліпса розсіювання точок перетину випадкових траєкторій складної системи з площиною PL_t. Рівняння цього еліпса у площині PL_t має вигляд:

$$\left(x-\xi(t)\right)^T W^+(t)\left(x-\xi(t)\right)=2k^2,$$

де *k* – довірча ймовірність $P=1-e^{-k}$.

Позначимо через $u_1(t)$, $u_2(t)$ ортонормований базис площини PL_t. Цей базис може бути знайдений за відомим τ-періодичним розв'язком ξ(t). Власні вектори $v_1(t)$,$v_2(t)$ можна отримати обертанням базису $u_1(t)$, $u_2(t)$ на деякий кут ϕ(t):

$$v_1(t)=u_1(t)cos\phi(t)+u_2(t)sin\phi(t), \tag{6}$$

$$v_2(t)=-u_1(t)sin\phi(t)+u_2(t)cos\phi(t). \tag{7}$$

В результаті рівняння (5)–(7) дозволяють виразити невідоме рішення системи (1)–(3) через три скалярні функції $\lambda_1(t)$, $\lambda_2(t)$, ϕ(t).

Позначимо:

$$P_1(t)=v_1(t)v_1^T(t),\ P_2(t)=v_2(t)v_2^T(t).$$

Відзначено, що $P_i(t)$ (i=1, 2) є проекційними матрицями:

$$P_iv_i=v_i,\ P_iv_i=0(i\neq j),\ P=P_1+P_2. \tag{8}$$

Представлено розклад (5) у вигляді:

$$V(t)=\lambda_1(t)P_1(t)+\lambda_2(t)P_2(t).$$

Зроблено наступне припущення. Для ортонормованих векторних функцій $v_i(t)$ і проекційних матриць $P_i(t)=v_i(t)v_i^T(t)$ (i=1,2) справедливі наступні тотожності:

$$v_1^T(t)\widehat{P}_1(t)v_1(t)\equiv 0, \tag{9}$$

$$v_1^T(t)\widehat{P}_2(t)v_1(t)\equiv 0, \tag{10}$$

$$v_2^T(t)\widehat{P}_1(t)v_2(t)\equiv 0, \tag{11}$$

$$v_2^T(t)\widehat{P}_2(t)v_2(t)\equiv 0, \tag{12}$$

$$v_1^T(t)\widehat{P}_1(t)v_2(t)=\phi(t)+K_1^T(t)u_2(t). \tag{13}$$

$$v_1^T(t)\widehat{P}_2(t)v_2(t)=\phi(t)+K_1^T(t)u_2(t). \tag{14}$$

Довести вказане припущення можна таким чином – тотожність (9)

безпосередньо випливає з наступного співвідношення:

$$v_1^T \widehat{P}_1 v_2 = v_1^T \left[v_1 v_1^T \right] v_1 = v_1^T \left[v_1 v_1^T + \widehat{v}_1 v_1^T \right] v_1 = v_1 v_1^T + \widehat{v}_1 v_1^T = \left[v_1 v_1^T \right] \equiv 0. \tag{15}$$

Тотожність (12) доводиться аналогічно. Тотожність (10) отримується з:

$$v_1^T \widehat{P}_2 v_2, = v_1^T \left[v_2 v_2^T + \widehat{v}_2 v_2^T \right] v_1 = v_2 v_1^T v_1 v_2^T + v_2 v_1^T v_1 v_2^T \equiv 0. \tag{16}$$

Тотожність (11) доводиться аналогічно.

Використаємо наступні рівності:

$$v_1 = u_1 \cos\varphi + u_2 \sin\varphi + v_2 \varphi,$$

$$u_1^T u_1 \equiv 0, u_2^T u_2 \equiv 0, -\left(u_1^T u_2 \right) = u_1^T u_2.$$

Після цього можна сформувати такий вираз:

$$v_1^T \widehat{P}_1 v_2 = v_1^T \left[v_1 v_1^T + \widehat{v}_1 v_1^T \right] v_2 = v_2 v_1^T = \left(u_1^T \cos\varphi + u_2^T \sin\varphi + v_2^T \varphi \right) v_2 = \left(u_1^T \cos\varphi + u_2^T \sin\varphi \right) v_2 + \varphi =$$

$$= -\left(u_1^T u_1 \cos\varphi \sin\varphi + u_1^T u_2 \cos^2\varphi - u_2^T u_1 \sin^2\varphi + u_2^T u_2 \cos\varphi \sin\varphi + \varphi \right) = u_1^T u_2 + \varphi \right).$$

З цих співвідношень слідує (13). Тотожність (14) доводиться аналогічно.

Доведене припущення надає набір тотожностей для ортонормованих векторних функцій і проекційних матриць. Ці тотожності дозволяють суттєво спростити обчислення, пов'язані з аналізом чутливості циклів і взаємодій між проекціями векторів у складній ком'ютерні системі. Оскільки матриці $P_i(t)$ є проекційними та ортогональними, їх використання допомагає розділяти аналіз векторів у різних напрямках, що робить можливим більш ефективне моделювання та прогнозування станів складної комп'ютерної системи.

Використання проекційних матриць і ортогональних векторів для вивчення поведінки компонентів складної комп'ютерної системи є відмінним підходом від існуючих моделей аналізу аномалій, таких як кореляційний аналіз або аналіз на основі кластеризації. Модель (9)–(16) надає можливість оцінювати зміни в ортогональності векторів, що дозволяє виявляти складні взаємодії між компонентами.

Ортогональні проекції є основою для виявлення аномальних відхилень, які можуть вказувати на нові або несподівані зв'язки між процесами. Наприклад, втрачена ортогональність вказує на порушення незалежності компонентів, що може бути ознакою вторгнення або несправності.

Також модель дозволяє оцінювати, як різні компоненти системи взаємодіють один з одним. Наприклад, нормальний стан системи передбачає, що певні компоненти (такі як використання CPU та мережева активність) слабо корелюють. Використовуючи модель, можна визначити наявність небажаних взаємодій між компонентами системи через аналіз проекційних матриць та їх поведінки з часом. Поява кореляцій між векторами, які мають бути ортогональними, може сигналізувати про спробу несанкціонованого доступу або впровадження нового шкідливого процесу, що впливає на кілька частин системи одночасно.

На рис. 2 представлені результати роботи моделі для виявлення аномалій у складній комп'ютерній системі. Графік відображає розподіл нормального стану складної комп'ютерної системи (позначено синіми точками) та аномалій (позначено червоними точками) у 3D просторі.

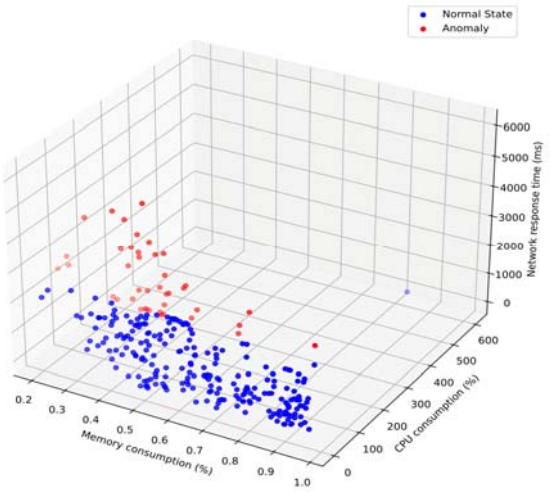


Рисунок 2 – Аттрактор поведінки складної комп'ютерної системи,
отриманий на основі моделі ВАЗККС

Джерело: розроблено авторами

В табл. 1. наведені значення досліджених показників точності виявлення аномалій у складній комп'ютерні системі

Таблиця 1 – Показники точності виявлення аномалій у складній комп'ютерній системі з використанням запропонованої моделі, отримані в результаті моделювання

№	Метрика	Значення	Час виконання
1	Правильність (Accuracy)	0,84	0,0023 мс
2	Точність (Precision)	0,87	
3	Повнота (Recall)	0,74	
4	Зважена середня точність (F1-Score)	0,78	

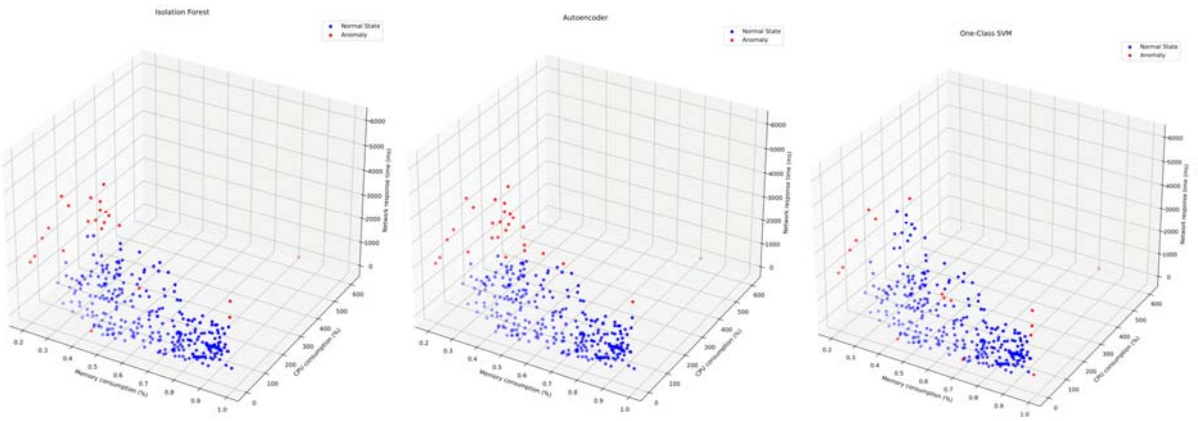
Джерело: розроблено авторами

З рис. 2 та табл. 1 видно, що запропонований метод виділяє деякі аномальні точки, але вони значною мірою переплітаються з нормальними, що свідчить про наявність деяких помилок при визначені аномалії. Це може бути наслідком непідготовленості даних, отриманих з датасету CSE-CIC-IDS 2018 [11, 12]. Але слід зазначити, що використання розробленої моделі разом з моделями машинного навчання має перспективу в покращенні показників точності.

Для перевірки можливості покращення розробленої моделі за допомогою выдомих методів машинного навчання слід зауважити, що основною моделлю оцінки даних на аномальність залишається розроблена модель виявлення аномальних зв'язків між компонентами складної комп'ютерної системи.

На рис. 3. представлені 3D-графіки результатів експериментів з використанням розробленої моделі у поєднанні з іншими моделями машинного навчання, а саме з Isolation Forest, Autoencoder та One-Class SVM [13–16].

На графіку, що ілюструє результати поєднання розробленої моделі з моделлю Isolation Forest (рис. 3а), можна помітити, що запропонована модель виявляє аномалії у складній комп'ютерній системи, розподілені по всіх осях, з меншою концентрацією в порівнянні з нормальними точками. Однак точність цієї комбінованої моделі не дуже висока (~0.88). Тобто деякі аномалії ця комбінована модель не виявляє, а частина нормальних даних може бути помилково класифікована як аномалії.



а *б* *в*

Рисунок 3 – 3D-графіки результатів експериментів з використанням розробленої моделі у поєднанні з іншими моделями машинного навчання (а – Isolation Forest, б – Autoencoder, в – One-Class SVM)

Джерело: розроблено авторами

Графік, що ілюструє результати поєднання розробленої моделі з моделлю Autoencoder (рис. 3б), показує, що аномалії розподілені досить рівномірно, але ця комбінована модель пропускає багато аномальних точок. Метрики ефективності також не показали високих значень. Це може свідчити про те, що Autoencoder не повністю навчився особливостям системи і, можливо, потребує подальшого налаштування гіперпараметрів або більшого обсягу навчальних даних, що в умовах реального часу є складним завданням.

Графік, який ілюструє результати поєднання розробленої моделі з моделлю One-Class SVM (рис. 3в) показує, що ця комбінована модель краще справляється в деяких аспектах, проте, з рисунку видно, що є області, в яких модель або пропускає аномалії, або помилково класифікує нормальні дані як аномальні.

У табл. 2 представлені значення показників точності запропонованих комбінованих моделей визначення аномалій у складних комп’ютерних системах.

Таблиця 2 – Показники точності виявлення аномалій у складній комп’ютерній системі з використанням запропонованих комбінованих моделей, отримані в результаті моделювання

№	Метрика	Значення	Час виконання
Розроблена модель у комбінації з моделлю Isolation Forest			
1	Правильність (Accuracy)	0,95	787,36 мс
2	Точність (Precision)	0,89	
3	Повнота (Recall)	0,65	
4	Зважена середня точність (F1-Score)	0,78	
Розроблена модель у комбінації з моделлю Autoencoder			
1	Правильність (Accuracy)	0,93	33020,25 мс
2	Точність (Precision)	0,87	
3	Повнота (Recall)	0,73	
4	Зважена середня точність (F1-Score)	0,78	

Продовження таблиці 2

Розроблена модель у комбінації з моделлю One-Class SVM			
1	Правильність (Accuracy)	0,99	5,11 мс
2	Точність (Precision)	0,99	
3	Повнота (Recall)	0,76	
4	Зважена середня точність (F1-Score)	0,88	

Джерело: розроблено авторами

Як видно з результатів, наведених у табл. 1-2, найкращі часові характеристики показав розроблена комбінована модель з використанням One-Class SVM. Це говорить про можливість його використання й в складних комп'ютерних системах реального часу. При цьому розроблений модель має певні переваги в часі виконання, а точність виявлення аномалій співставна з моделями машинного навчання.

Таким чином, була запропонована модель виявлення аномальних зв'язків між компонентами складної комп'ютерної системи, яка є вдосконаленим інструментом для математичного аналізу поведінки комп'ютерних систем з метою захисту їх від дестабілізуючих факторів. Розроблена модель використовує ортогональні векторні функції та проекційні матриці для аналізу аномалій. Це дозволяє створювати просторові розклади, які дають можливість виявляти складні взаємозв'язки між компонентами складної комп'ютерної системи, використовуючи тільки значення і вектори. Такий підхід поєднує математичний формалізм з реальними технічними даними, що робить його ефективним для глибокого аналізу аномалій у складній комп'ютерній системі. Крім того, дана модель дозволяє описати поведінку складної системи як геометричний еліпс у просторі, де параметри еліпсу відповідають рівням завантаженості процесора, пам'яті та інших ресурсів, а це забезпечує можливість детального аналізу взаємозв'язків між компонентами складної комп'ютерної системи.

Висновки. В результаті дослідження розроблено математичну модель виявлення аномальних зв'язків між компонентами складної комп'ютерної системи, яка на відміну від інших використовує геометричний підхід, де аномалії виявляються через зміну взаємної ортогональності між компонентами. Це дозволило зменшити час виявлення аномалій стану комп'ютерної системи до 10%. При цьому точність виявлення аномалій залишилась на заданому рівні.

Проведено дослідження використання розробленої моделі у комплексі з моделями Isolation Forest, Autoencoder, One-Class SVM. Результати досліджень показали суттєві (до 10 разів) збільшення швидкості виявлення аномалій поведінки комп'ютерної системи, при незначному зниженні точності цієї операції. Це дає можливість зробити висновки про доцільність використання розробленої моделі для виявлення аномалій поведінки високонавантажених складних комп'ютерних систем в режимі реального часу.

Список літератури

1. S. Yu, H. Jiang, S. Huang, X. Peng & A. Lu, "Compute-in-Memory Chips for Deep Learning: Recent Trends and Prospects," in IEEE Circuits and Systems Magazine, vol. 21, no. 3, pp. 31-56, thirdquarter 2021, doi: 10.1109/MCAS.2021.3092533. https://www.scimagojr.com/journalsearch.php?q=26004&tip=sid&clean=0
2. S. Kumar, S. Gupta and S. Arora, "Research Trends in Network-Based Intrusion Detection Systems: A Review," in IEEE Access, vol. 9, pp. 157761-157779, 2021, doi: 10.1109/ACCESS.2021.3129775. https://www.scimagojr.com/journalsearch.php?q=21100374601&tip=sid&clean=0
3. Lu, P.-J.; Lai, M.-C.; Chang, J.-S. A Survey of High-Performance Interconnection Networks in High-Performance Computer Systems. Electronics 2022, 11, 1369. https://doi.org/10.3390/electronics11091369

4. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Y., Yevstrat, D., Chyrva, Y., Kuchuk, H. (2022). Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples. Eastern European Journal of Enterprise Technologies, 6 (4 (120)), 40–49. doi: https://doi.org/10.15587/1729¬4061.2022.269128
5. Meleshko, Y., Raskin, L., Semenov, S., & Sira, O. (2019). Methodology of probabilistic analysis of state dynamics of multi¬dimensional semi¬Markov dynamic systems. Eastern-European Journal of Enterprise Technologies, 6(4 (102), 6–13. https://doi.org/10.15587/1729-4061.2019.184637
6. Semenov, S., Zhang, L., Cao, W., Bulba, S., Babenko, V., & Davydov, V. (2021). Development of a fuzzy GERT-model for investigating common software vulnerabilities. Eastern-European Journal of Enterprise Technologies, 6(2 (114), 6–18. https://doi.org/10.15587/1729-4061.2021.243715
7. Meleshko, Y.V., Yakymenko, M., & Semenov, S. (2021). A Method of Detecting Bot Networks Based on Graph Clustering in the Recommendation System of Social Network. International Conference on Computational Linguistics and Intelligent Systems.
8. Semenov, S., Gavrilenko, S. and Chelak, V. (2016), "Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on basis of economic test", Actual problems of economics, Kyiv, Vol. 4(178), pp. 451-459.
9. Devaney, Robert. (2021). An Introduction to Chaotic Dynamical Systems. 10.1201/9780429280801.
10. Göcs, László & Johanyák, Zsolt. (2023). Identifying Relevant Features of CSE-CIC-IDS2018 Dataset for the Development of an Intrusion Detection System. 10.48550/arXiv.2307.11544.
11. Göcs, László & Johanyák, Zsolt. (2023). Identifying Relevant Features of CSE-CIC-IDS2018 Dataset for the Development of an Intrusion Detection System. 10.48550/arXiv.2307.11544.
12. IDS 2018 | Datasets | Research | Canadian Institute for Cybersecurity | UNB." [Online]. Available: https://www.unb.ca/cic/datasets/ids-2018.html. [Accessed: 05-Nov-2024].
13. Mohammed, Mahmood & Telek, Miklos. (2023). Anomaly Detection using combination of Autoencoder and Isolation Forest. 25 - 30. 10.3311/WINS2023-005.
14. Ribeiro, D.; Matos, L.M.; Moreira, G.; Pilastri, A.; Cortez, P. Isolation Forests and Deep Autoencoders for Industrial Screw Tightening Anomaly Detection. Computers 2022, 11, 54. https://doi.org/10.3390/computers11040054
15. Gavrylenko , S. Y. , & Sheverdin, I. V. (2021). DEVELOPMENT OF METHOD TO IDENTIFY THE COMPUTER SYSTEM STATE BASED ON THE «ISOLATION FOREST» ALGORITHM . Radio Electronics, Computer Science, Control, (1), 105–116. https://doi.org/10.15588/1607-3274-2021-1-11
16. Gavrylenko S., Semenov S., Sira O., Kuchuk N. Identification of the state of an object under conditions of fuzzy input data. Eastern-European Journal of Enterprise Technologies, 2019, Vol. 1, No. 4 (97), pp. 22–29. DOI: 10.15587/1729-4061.2019.157085

References

1. S. Yu, H. Jiang, S. Huang, X. Peng and A. Lu, "Compute-in-Memory Chips for Deep Learning: Recent Trends and Prospects," in IEEE Circuits and Systems Magazine, vol. 21, no. 3, pp. 31-56, thirdquarter 2021, doi: 10.1109/MCAS.2021.3092533. https://www.scimagojr.com/journalsearch.php?q=26004&tip=sid&clean=0
2. S. Kumar, S. Gupta and S. Arora, "Research Trends in Network-Based Intrusion Detection Systems: A Review," in IEEE Access, vol. 9, pp. 157761-157779, 2021, doi: 10.1109/ACCESS.2021.3129775. https://www.scimagojr.com/journalsearch.php?q=21100374601&tip=sid&clean=0
3. Lu, P.-J.; Lai, M.-C.; Chang, J.-S. A Survey of High-Performance Interconnection Networks in High-Performance Computer Systems. Electronics 2022, 11, 1369. https://doi.org/10.3390/electronics11091369
4. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Y., Yevstrat, D., Chyrva, Y., & Kuchuk, H. (2022). Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples. Eastern European Journal of Enterprise Technologies, 6 (4 (120)), 40–49. doi: https://doi.org/10.15587/1729¬4061.2022.269128
5. Meleshko, Y., Raskin, L., Semenov, S., & Sira, O. (2019). Methodology of probabilistic analysis of state dynamics of multi¬dimensional semi¬Markov dynamic systems. Eastern-European Journal of Enterprise Technologies, 6(4 (102), 6–13. https://doi.org/10.15587/1729-4061.2019.184637
6. Semenov, S., Zhang, L., Cao, W., Bulba, S., Babenko, V., & Davydov, V. (2021). Development of a fuzzy GERT-model for investigating common software vulnerabilities. Eastern-European Journal of Enterprise Technologies, 6(2 (114), 6–18. https://doi.org/10.15587/1729-4061.2021.243715
7. Meleshko, Y.V., Yakymenko, M., & Semenov, S. (2021). A Method of Detecting Bot Networks Based on Graph Clustering in the Recommendation System of Social Network. International Conference on Computational Linguistics and Intelligent Systems.

8. Semenov, S., Gavrilenko, S. & Chelak, V. (2016), "Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on basis of economic test", Actual problems of economics, Kyiv, Vol. 4(178), pp. 451-459.
9. Devaney, Robert. (2021). An Introduction to Chaotic Dynamical Systems. 10.1201/9780429280801.
10. Göcs, László & Johanyák, Zsolt. (2023). Identifying Relevant Features of CSE-CIC-IDS2018 Dataset for the Development of an Intrusion Detection System. 10.48550/arXiv.2307.11544.
11. Göcs, László & Johanyák, Zsolt. (2023). Identifying Relevant Features of CSE-CIC-IDS2018 Dataset for the Development of an Intrusion Detection System. 10.48550/arXiv.2307.11544.
12. IDS 2018 | Datasets | Research | Canadian Institute for Cybersecurity | UNB." [Online]. Available: https://www.unb.ca/cic/datasets/ids-2018.html. [Accessed: 05-Nov-2024].
13. Mohammed, Mahmood & Telek, Miklos. (2023). Anomaly Detection using combination of Autoencoder and Isolation Forest. 25 - 30. 10.3311/WINS2023-005.
14. Ribeiro, D.; Matos, L.M.; Moreira, G.; Pilastri, A.; Cortez, P. Isolation Forests and Deep Autoencoders for Industrial Screw Tightening Anomaly Detection. Computers 2022, 11, 54. https://doi.org/10.3390/computers11040054
15. Gavrylenko , S. Y. , & Sheverdin, I. V. (2021). DEVELOPMENT OF METHOD TO IDENTIFY THE COMPUTER SYSTEM STATE BASED ON THE «ISOLATION FOREST» ALGORITHM . Radio Electronics, Computer Science, Control, (1), 105–116. https://doi.org/10.15588/1607-3274-2021-1-11
16. Gavrylenko S., Semenov S., Sira O., Kuchuk N. Identification of the state of an object under conditions of fuzzy input data. Eastern-European Journal of Enterprise Technologies, 2019, Vol. 1, No. 4 (97), pp. 22–29. DOI: 10.15587/1729-4061.2019.157085

Yelyzaveta Meleshko, Prof., DSc., **Mykola Yakymenko,** Assoc. Prof., PhD phys.&math. sci.,
Volodymyr Mikhav, PhD, **Yaroslav Shulika,** post- graduate
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

A Mathematical Model of Detecting Anomalous Connections Between Components of a Complex Computer System

The object of the research is the process of detecting anomalies in high-load complex computer systems (HLCSS). The practical application area includes HLCSS such as banking transaction servers and cloud platforms, where stable operation must be ensured under heavy request loads. The problem addressed in the research is the lack of real-time anomaly detection models in HLCSS with a specified accuracy under resource constraints. Anomalies may signal system malfunctions, process mismatches, or potential cyberattacks. High-load systems are particularly sensitive to even minor disruptions, as they can cause significant delays or complete service unavailability for many users simultaneously, leading to financial losses and loss of customer trust.

Modern web services face issues such as attempts of DDoS attacks, significant fluctuations in user requests, database connection problems, memory leaks, and the impact of unexpected changes in network or hardware configuration. Therefore, ensuring timely anomaly detection is a critical aspect of high-load system security. Automating the analysis process and promptly identifying potential problems in real time help reduce risks and minimize downtime. Developing a mathematical model for anomaly detection in high-load web services aims not only to improve detection accuracy but also to ensure a timely response to potential threats under resource constraints. This helps avoid significant financial losses and enhances user trust in such services.

The paper creates and investigates a mathematical model for detecting anomalous connections between components of a complex computer system (HLCSS). The testing results of the model showed the following performance metrics: accuracy – 84%, precision – 87%, recall – 74%, F1-Score – 78%. The positive results of the study are explained by the following prerequisites: the HLCSS model uses projection matrices and orthogonal vector functions for anomaly analysis. This allows for the creation of spatial decompositions that reveal complex interconnections between components of a complex computer system using only eigenvalues and vectors. Thus, the model can be applied for operational data analysis and anomaly detection in resource-constrained environments.

high-load complex computer systems, anomaly detection, mathematical models, dynamic chaos, complex networks

Одержано (Received) 18.11.2024 *Прорецензовано (Reviewed) 17.12.2024*
Прийнято до друку (Approved) 23.12.2024