УДК 004.75
DOI: 10.31891/2219-9365-2021-68-2-6

САВЕНКО Б. О., КАШТАЛЬЯН А. С.
Хмельницький національний університет

УДОСКОНАЛЕННЯ МЕТОДУ ЦЕНТРАЛІЗОВАНОГО ВИЯВЛЕННЯ РОЗПОДІЛЕНИХ АНОМАЛІЙ ЗА АЛГОРИТМОМ ПОШУКУ ГОЛОВНИХ КОМПОНЕНТ

В роботі удосконалено метод виявлення аномалії згідно методу головних компонент в комп'ютерних системах в мережі, який надав змогу застосовувати його не до однієї комп'ютерної станції, а до групи станцій, в яких встановлена самоорганізована розподілена система виявлення аномалій в комп'ютерних системах в мережі, що на відміну від відомих рішень, при застосуванні надав змогу скоротити обсяг даних і відповідно прискорити їх обмін між компонентами системи.

Практичне значення одержаних результатів полягає у розробленій архітектурі і компонентах розподіленої системи виявлення аномалій в комп'ютерних системах, в якій синтезовано вимоги самоорганізованості, централізованості, розподіленості, багаторівневості та на її основі створено самоорганізовану розподілену систему.

Проведені експериментальні дослідження з розробленою реалізацією самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах підтвердили ефективність запропонованих рішень і розробленої розподіленої системи щодо її функціонування в комп'ютерній мережі.

Ключові слова: розподілені системи, аномалії, ефективність, метод головних компонент.

B. SAVENKO, A. KASHTALIAN
Khmelnitskyi National University

IMPROVEMENT OF THE METHOD OF CENTRALIZED DETECTION OF DISTRIBUTED ANOMALIES IN THE ALGORITHM FOR SEARCHING MAIN COMPONENTS

The method of anomaly detection is improved according to the method of main components in computer systems in the network, which allowed to apply it not to one computer station, but to a group of stations in which a self-organized distributed system of anomaly detection in computer systems is installed. network, which, in contrast to known solutions, in application allowed to reduce the amount of data and, accordingly, speed up their exchange between system components.

The practical significance of the obtained results lies in the developed architecture and components of the distributed anomaly detection system in computer systems, which synthesizes the requirements of self-organization, centralization, distribution, multilevel and creates a self-organized distributed system.

Experimental studies with the developed implementation of a self-organized distributed system for detecting anomalies in computer systems have confirmed the effectiveness of the proposed solutions and the developed distributed system for its operation in the computer network.

Keywords: distributed systems, anomalies, efficiency, principal components method.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Комп'ютерні системи (КС) продовжують активно використовуватись в усіх сферах діяльності людини. Вони дозволяють суттєво підвищити продуктивність праці та автоматизувати багато складних процесів і це актуалізує перспективи їх використання в майбутньому.

Програмне забезпечення, яке використовується в КС, виконує дуже важливу роль та забезпечує ефективне розв'язання задач. Але із зростанням актуальності задач, при розв'язанні яких використовують КС, зловмисники спрямовують свої зусилля на взяття під контроль КС шляхом впливу на їх програмне забезпечення. Для цього вони застосовують різноманітні атаки на КС [1-4]. Частина з них стає успішною і цим створює проблеми користувачам КС. Для важливих критичних сфер [2] діяльності людини втрата контролю над КС стає серйозною проблемою з відповідними катастрофічними наслідками. Тому, проблемі побудови ефективних систем протидії зловмисним діям спрямованим на КС користувачів все більше приділяють уваги дослідники та розробники відповідних методів і засобів [2-9]. Такі системи потрібно наповнювати підсистемами виявлення аномальних чи зловмисних проявів, реалізовуючи в них відповідні методи.

Аналіз досліджень та публікацій

Дослідженню аномалій в комп'ютерних системах для виявлення зловмисного програмного забезпечення і комп'ютерних атак приділяють увагу багато дослідників. Вони розробили багато різних методів виявлення [10-18]. Для розробки нових чи покращення відомих рішень з виявлення аномалій в КС потрібно встановити переваги та недоліків відомих розроблених методів. Виділення нерозв'язаних підзадач, що впливають на досягнення ефективного результату з виявлення, та стратегії з усунення частини недоліків відомих методів дозволять підвищити достовірність та покращити ефективність виявлення. Розглянемо відомі

методи виявлення зловмисного програмного забезпечення і комп'ютерних атак в КС, які базуються на встановленні аномальних станів. До розгляду візьмемо ті з них, які запропоновані відомими дослідниками в цій галузі.

У роботі [10] представлений поглиблений аналіз чотирьох основних категорій методів виявлення аномалій, які включають класифікацію, статистику, теорію інформації та кластеризацію. Основна увага зосереджена на проблеми дослідження з наборами даних, що використовуються для виявлення вторгнень у мережу. Результати дослідження дають змогу пов'язати класифікацію, статистичну обробку даних та кластеризацію із поставленою науковою задачею в частині розробки застосовуваних методів при обробці вхідних даних з метою виявлення мережних аномалій.

Проблемі узгодження низьковимірних багатовимірних об'єктів до високомірних даних розглянуто в роботі [11] як з теоретичної, так і з обчислювальної точки зору. Оскільки набори даних стають більш неоднорідними та ускладненими, простори, які використовуються для їх апроксимації, повинні стати більш неоднорідними. В цій роботі відображеног результати роботи з переходами до змінених базисів, що є актуальним в розрізі поставленої задачі з виявлення аномалій і зменшення розмірності даних. Також, проаналізована обчислювальна складність таких перетворень для оцінки необхідних обчислювальних ресурсів.

З розповсюдженням Інтернету речей через бездротові сенсорні мережі генерується величезна кількість даних датчиків з безпрецедентною швидкістю, що призводить до дуже великої кількості явної або неявної інформації [12]. При аналізі таких даних датчиків особливо важливо точно та ефективно виявляти не тільки окремі аномальні поведінки, але й аномальні події (тобто моделі поведінки). Однак більшість попередніх робіт були зосереджені лише на виявленні аномалій, водночас ігноруючи співвідношення між ними. Навіть у підходах, що враховують кореляцію між аномаліями, більшість ігнорує той факт, що аномалія стану даних датчиків змінюється з часом. У цій статті [12] запропоновано безконтрольний метод виявлення аномалій контексту в Інтернеті речей за допомогою бездротових сенсорних мереж, який враховує як статус динамічної аномалії, так і кореляцію між аномаліями, заснованими в контексті на їх просторових та часових сусідах. А, також, досліджено в роботі ефективність запропонованого методу в моделі виявлення аномалії. Внесення в роботу відомостей про аномалії і аномальні прояви є важливим з точки зору врахування динаміки отримання даних.

В роботі [13] процес виявлення несподіваних елементів або подій у наборах даних, які відрізняються від норми, розглядається як пошук аномалій. На відміну від стандартних задач класифікації, виявлення аномалій часто застосовується до немаркованих даних, беручи до уваги лише внутрішню структуру набору даних. Ця проблема відома як неконтрольоване виявлення аномалій і вирішується у багатьох практичних додатках, наприклад, у виявленні вторгнень у мережу, виявленні шахрайства, а також у галузі біології та медицини. У цій статті представлено результати здійсненої оцінки 19 різних алгоритмів виявлення аномалій на 10 різних наборах даних із декількох доменів додатків. Робота є важливою для досліджень безконтрольного виявлення аномалій.

Виявляти та обробляти аномалії для великих даних у режимі реального часу є складним завданням. Обсяг і швидкість даних у багатьох системах ускладнює типовим алгоритмам масштабування та збереження своїх характеристик у реальному часі [14]. Поширеність даних у поєднанні з проблемою, що багато існуючих алгоритмів враховують лише зміст джерела даних, а не контент. Запропоновані в роботі рішення визначають контекст виявлення аномалій. Він складається з двох різних кроків: виявлення вмісту та виявлення контексту. Детектор вмісту використовується для визначення аномалій у режимі реального часу. Детектор контексту використовується для обрізання результатів детектора вмісту, виявляючи ті аномалії, які вважаються як змістовими, так і контекстно аномальними. Детектор контексту використовує концепцію профілів, які є групами аналогічно згрупованих точок даних, що генеруються багатовимірним алгоритмом кластеризації. Дослідження було оцінено на основі проведених експериментів для двох реальних наборів даних датчиків. Резуьтати цієї роботи [14] важливі в контексті важливості обробки контенту датчиків.

В роботі [15] пропонується поступовий метод безконтрольного виявлення аномалії, який дозволяє швидко аналізувати та обробляти великі дані в режимі реального часу. Оцінка набору даних під час експерименту показує, що метод зближується зі своїм автономним аналогом для нескінченно зростаючих потоків даних.

В роботі [16] проаналізовано відомі рішення з виявлення аномалій, особливо для даних із великими розмірами та змішаними типами, де виявлення аномальних моделей чи поведінки є нетривіальною роботою. В результаті важливість даної роботи в проведених дослідженнях відомих підходів і їх порівнянні, що дозволить врахувати ці результати при виборі перспективних рішень.

В роботі [17] зосереджено увагу на ранньому виявленні несподіваних спостережень у фізичній інфраструктурі, що має велике значення для запобігання поломки системи та подальших втрат. Однак сучасна техніка для виявлення аномалій в існуючій платформі моніторингу інфраструктури головним чином залежить від методу фіксованого порогу. Очевидним недоліком методу є те, що він, як правило, призводить до високого рівня помилкового виявлення. У цьому дослідженні підхід до виявлення статистичних аномалій запроваджено

до моніторингу фізичної інфраструктури. В роботі розглядаються три важливі типи аномалій, які зустрічаються на платформі моніторингу інфраструктури, а саме наївні точкові аномалії, контекстні аномалії точок та зміщення рівня. В роботі пропонується до застосування розроблений метод, заснований на моделі Гаусса, для виявлення зазначених трьох аномалій. Оскільки запропонований метод може ефективно виявляти лише наївні точкові аномалії; запропоновано вдосконалений підхід, що поєднує результати статистичних випробувань на вихідних та першовідмінних даних моніторингу. Оцінюються результати запропонованих методів на реальному наборі даних. Результати показують, що оптимізований підхід до виявлення аномалій має хорошу точність і може значно знизити швидкість неправильного виявлення. Отримані результати дають розуміння обробки трьох типів аномалій.

Однією із сучасних проблем виявлення аномалій є здатність виявляти і розрізняти як точкові, так і колективні аномалії в межах послідовності даних або часових рядів [18]. В роботі [18] розроблено метод та засоби, щоб надати користувачам вибір методів виявлення аномалій, і, зокрема, забезпечує реалізацію нещодавно запропонованого сімейства алгоритмів виявлення аномалій. У статті [18] описуються реалізовані методи, а також висвітлюється їх застосування до модельованих даних, а також реальні приклади даних, що містяться в пакеті. Поділ на точкові і колективні аномалії, а також, методи їх виявлення є важливим в розвитку методології з виявлення аномалій.

Метою роботи [19] є швидке та точне виявлення ненормальних даних складного та складного промислового обладнання із датчиками. Завдяки стрімкому розвитку Інтернету речей, все більше обладнання обладнується датчиками, особливо більш складне та складне промислове обладнання встановлюється з великою кількістю датчиків. Для моніторингу роботи обладнання швидко збирається велика кількість даних моніторингу. Обробка таких даних, причому у великій кількості, представлена в роботі.

В роботі [20] представлено застосування такого методу, який називається однокласною машиною векторної підтримки, для пошуку аномальних шаблонів серед джерел, попередньо вибраних із середньо-інфрачервоного каталогу. Для створення моделі очікуваних даних в роботі описано результат тренувань алгоритму на наборі об'єктів зі спектроскопічними ідентифікаціями. Виявлення аномалій додає гнучкості автоматизованим процедурам поділу джерел та допомагає перевірити надійність та репрезентативність навчальних зразків. Таким чином, це слід розглядати як важливий крок у контрольованих схемах класифікації для забезпечення повноти та чистоти створених каталогів.

У роботі [21] описано загальний механізм аналітики, який забезпечує надійні попередження про зміни та аномалії в сенсорному потоці даних. У той час як більшість існуючих аналітичних реалізацій IoT вимагають припущень, що стосуються конкретних областей, в роботі надано значну інформацію через методи машинного навчання та вдосконалені статистичні тести без попередніх знань. Система виявлення аномалій мережі дозволяє контролювати комп'ютерну мережу, яка поводиться інакше, ніж мережевий протокол, і її багато застосовується в різних доменах. Проте, проблема виникає там, де різні домени застосунків мають різні визначальні аномалії у своєму середовищі. Вони ускладнюють вибір найкращих алгоритмів, які відповідають вимогам певних доменів. Крім того, проблема централізації, яка спричиняє руйнування мережевої системи, коли в систему вливається потужний зловмисний код. Тому в цій роботі показано результати проведеного експерименту із використанням контрольованого машинного навчання для системи виявлення аномалій мережі, яка мінімізує вартість зв'язку та пропускну здатність мережі, мінімізовану за допомогою набору даних для порівняння їх продуктивності в термінах їх точності та часу обробки для класифікатора для побудови моделі. В результаті, розподілений алгоритм вирішує проблему централізації з точністю та часом обробки, як і раніше, значним у порівнянні з централізованим алгоритмом, хоча є певна втрата точності та часу.

Формулювання цілей статті

Таким чином, використання дослідження аномалій в КС є перспективним напрямом. При цьому важливим завданням постає досягнення зменшення розмірності досліджуваних характеристик для оперативної обробки результатів, що впливатиме на покращення ефективності виявлення та розподілених систем виявлення зловмисного програмного забезпечення та комп'ютерних атак.

Виклад основного матеріалу

Удосконалення методу централізованого виявлення розподілених аномалій згідно з алгоритмом пошуку головних компонент

Розглянемо використання самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах, яка надає можливість проводити пошук безпосередньо в одній комп'ютерній станції або одночасно в декількох. При цьому в обох випадках можна використати метод головних компонент в якості покрокового ітераційного алгоритму для отримання числових значень показників характерних ознак безпосередньо отриманих в одній комп'ютерній станції та декількох напротязі певного часу в деякому ковзному вікні. Також, отриману з вузлів в мережі інформацію про процеси, що протікають в них, самоорганізована розподілена система виявлення аномалій може досліджувати на предмет проявів аномалій, які відповідатимуть або зловмисному програмному забезпеченню або комп'ютерним атакам. Враховуючи складність виявлення зловмисного програмного забезпечення чи комп'ютерних атак через обмеженість кількості ознак за якими можна встановити аномальні прояви, а також, наявність надмірних обсягів

різнотипної та різнорідної інформації зібраної з вузлів в мережі, необхідним є удосконалення методу централізованого виявлення розподілених аномалій за алгоритмом пошуку головних компонент, який дозволив би зменшити розмірність інформації зібраної у вузлах в мережі без втрати її цінності та швидко обробити в єдиному центрі для забезпечення актуальності результату виявлення, що в результаті б покращило ефективність виявлення.

Для забезпечення виявлення розподілених аномалій з використанням методу централізованого виявлення розподілених аномалій за алгоритмом пошуку головних компонент, розробимо метод виявлення аномалій в одній з комп'ютерних станцій в мережі з врахуванням інтеграції цього методу в розроблену самоорганізовану розподілену систему з єдиним центром.

Збільшення активності в мережі до її вузлів виступає ознакою того, що це можуть бути зловмисні прояви і їх необхідно досліджувати. В реальному часі підвищена активність швидко змінюється на помірну, тому потрібні ефективні засоби і реалізовані в них методи, які б швидко реагували на такі події. Інакше, актуальність отриманої системою інформації про активність в мережі спрямовано до її вузла, а також, реакція на неї втрачатимуть необхідність. Основною ознакою, яку необхідно досліджувати першочергово в мережі, що фактично відповідає за підвищену активність, є обсяг трафіку. Відомо багато методів обробки трафіку мережі, причому з врахуванням різних топологій мереж та каналів входження в корпоративні чи локальні мережі. Зокрема, враховують також особливості трафіку при здійсненні розподілених атак і пошук самоподібності в його частинах.

Поняття мережного трафіку в постановці задачі дослідження аномалій включає дослідження кількості даних, що переміщуються в мережі протягом певного часу. Для правильного функціонування комп'ютерних мереж потрібно здійснювати в них контроль, аналіз, моделювання та управління відповідними спеціалізованими засобами. Особливо важливими в процесі виявлення аномалій є здійснення аналізу та вимірювання мережного трафіку, що включають моніторинг трафіку, зміни в ньому, тенденції, вимірювання кількості та виду трафіку. Отримання звітів різними спеціалізованими засобами про мережний трафік дає інформацію щодо запобігання зловмисним проявам та дозволяє забезпечити безпеку в мережі. Фрагменти обсягу трафіку даних протягом певного часу (три різних часових інтервали) в мережі Хмельницького національного університету зображено на рис. 1. Як видно із графіків в часових інтервалах обсяг трафіку змінюється і може суттєво відхилятись від середнього значення, що можна використати для встановлення аномальних проявів. Пересилання даних в комп'ютерних мережах здійснюється переважно в мережних пакетах. Ці пакети забезпечують навантаження в мережі. Варіантів передачі пакетів може бути багато і здійснюється за протоколами мережі. По прибуттю за місцем призначення в залежності від правил і протоколів пакети потребують здійснення перевірки наявності всіх, контролю цілісності і джерела надходження. Якщо розглядати трафік в магістральних лініях, то аномалії його обсягу можуть залишатися непоміченими через укрупнене представлення. Результати вимірювань можуть мати велику розмірність, в залежності від кількості ліній, але нормальні моделі трафіку знаходяться в підпросторі меншої розмірності. Виділення цього підпростору мережного трафіку, використовуючи метод головних компонент в трафіку, дає змогу ідентифікувати аномалії обсягу в підпросторі. Для аналізу аномальних проявів в мережному трафіку спочатку використаємо такі характеристичні параметри: коефіцієнт завантаження трафіку; типовий розмір пакету; середнє число фрагментованих пакетів. Для дослідження коефіцієнту завантаженості трафіку мережі розглянемо такі варіанти: мережний трафік протягом певного часу розкладається в часовий ряд; мережні трафіки порівнюється за певні часові періоди.

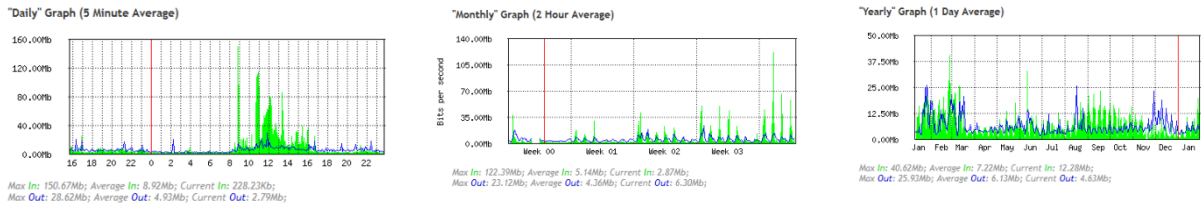

Рис. 1 - Зображення обсягу трафіку даних протягом певного часу (три різних часових інтервали) в мережі Хмельницького національного університету

Якщо мережний трафік отримується динамічно протягом певного часу, то розкладемо його в часовий ряд. Наприклад, нехай дано скінчену кількість векторів для його представлення $v_{1,j}, v_{2,j}, \dots, v_{f_j,j} \in R^u$, де вектори $v_{1,j}, v_{2,j}, \dots, v_{f_j,j}$ відповідають значенням ознак трафіку з j – того вузла в мережі, f_j – кількість ознак, тобто векторів. Для випадку представлення трафіку, приклад якого зображено графіками на рис. 1, через час його надходження та обсяг в конкретний момент часу, отримуємо, що значення f_j= 2. Тоді, пара векторів $v_{1,j}, v_{2,j}$ представлятиме трафік мережі протягом часу, який задано вектором $v_{1,j}$. Сумарний обсяг трафіку в конкретний момент часу представлятиметься вектором $v_{2,j}$ і вимірюватимемо в байтах для всіх з'єднань.

Таким чином, кожна точка графіку, що представляє обсяг мережного трафіку, задається парою значень. За певний інтервал часу самоорганізованою розподіленою системою виявлення аномалій з певними сталими періодами часу здійснюється збір цих пар точок. Нумерація точок розпочинається з першої отриманої пари і продовжується до тієї точки, яка є останньою з очікуваних точок. Після того, як зібрано задану кількість пар, система здійснює їх централізацію. Представлені таким чином данні є двомірними.

Здійснимо представлення пари векторів для визначеної кількості q точок спостереження так:

$$\begin{pmatrix} v_{1,j.1} & v_{1,j.2} & \dots & v_{1,j.q} \\ v_{2,j.1} & v_{2,j.2} & \dots & v_{2,j.q} \end{pmatrix} \quad (1)$$

Після отримання q пар системою і обробкою, синхронно обсяг трафіку мережі продовжує відображатись системою в подальших парах. Самоорганізована розподілена система після обробки певного набору даних, які задано формулою (1), отримує частину даних фактично оновлених і залишає з оброблених даних пари, які надійшли останніми. Перші пари точок, після обробки, видаляються з подальших обчислень. Кількість таких видалених пар залежить від часу обробки системою та часу, який витрачений на пересилання даних. Якщо витрачений час більше, ніж час витрачений на збір q нових пар системою, тоді ці зібрані нові пари втрачаються, бо розпочнеться новий набір наступних пар. Для вирішення цієї проблеми необхідно збільшувати інтервал збору сусідніх пар векторів. Розрахунок інтервалу часу між сусідніми парами здійснимо так:

$$t_{int,1} = \frac{t_{obr,1}+t_{dos,1}+t_{dod,1}}{q}, \quad (2)$$

де $t_{int,1}$ – інтервал часу між отриманням даних обсягу трафіку, тобто між сусідніми значеннями $(v_{1,j,i-1}, v_{2,j,i-1})$ та $(v_{1,j,i}, v_{2,j,i})$; $t_{obr,1}$ – час, який витрачено на обробку даних з матриці (1); $t_{dos,1}$ – час, який витрачено на переміщення даних до центру самоорганізованої розподіленої системи; $t_{dod,1}$ – додаткові часові втрати, які пов'язані з затримками в обробці даних через вищу пріоритетність інших задач.

Додаткові часові втрати оцінюються експериментально і встановлюються значенням, яке є максимальним зі всіх досліджуваних. Але ці додаткові часові втрати не можуть перевищувати час обробки даних $t_{obr,1}$, тобто $t_{obr,1} \geq t_{dod,1}$. Тому, з самого початку цей час може бути заданий, як такий що дорівнює часу обробки, тобто $t_{dod,1} = t_{obr,1}$. На кожному етапі обробки даних з матриці (1) центр системи фіксуватиме значення додаткових часових втрат $t_{dod,1}$ та усереднюватиме його з попередніми значеннями, якщо з самого початку перше таке значення дорівнювало $t_{obr,1}$. Аналогічно проводимо оцінку для часу, який витрачено на переміщення даних до центру самоорганізованої розподіленої системи $t_{dos,1}$. Приймемо з самого початку його таким що дорівнює часу обробки, тобто $t_{dos,1} = t_{obr,1}$. В подальшому усереднюватимемо його зі всіма попередніми значеннями і отримаємо оцінку цього часу. Результатом такого підходу буде визначення початкового інтервалу часу між отриманням даних обсягу трафіку, тобто між сусідніми значеннями $(v_{1,j,i-1}, v_{2,j,i-1})$ та $(v_{1,j,i}, v_{2,j,i})$, який визначатиметься в залежності від часу обробки так:

$$t_{int,1} = \frac{3*t_{obr,1}}{q}. \quad (3)$$

Наслідком з формули (3) є те, що при значенні інтервалу часу між отриманням даних обсягу трафіку суттєво меншому за значення $\frac{3*t_{obr,1}}{q}$ за певний час надходження мережного трафіку частина значень не буде врахована при обчисленні та очікувані результати обчислень втратять актуальність, бо процес надходження мережного трафіку є динамічним. Ці обчислення стосуються дослідження мережного трафіку за обсягами, що надходять. Якщо значення інтервалу часу відповідає вимогам, тоді переходимо до обчислення, зокрема першочергово до централізації отриманих даних. Централізація потрібна, щоб в подальшому оперувати з середніми значеннями близькими до нуля або взагалі нульовими. Для здійснення централізації даних знаходимо середнє значення серед всіх значень кожної з характеристичних ознак і від кожного значення із отриманого набору (формула (1)) віднімає середня значення, що відноситься до його характеристичної ознаки. Геометричний зміст централізації означає переміщенню центра координат в нову точку, таким чином, що вся вибірка буде розміщена в межах цього нового центру. Це перепредставлення даних з матриці (1) дає можливість не тільки зменшити розрядність в числах, але і відображає розсіювання вибірки. Тому, знаходимо середні значення вибірки з даних представлених в формулі (1) так:

$$v_{1,j,s} = \frac{1}{q}\sum_{i=1}^{q} v_{1,j,i}, \; v_{2,j,s} = \frac{1}{q}\sum_{i=1}^{q} v_{2,j,i}, \quad (4)$$

де $v_{1,j,s}$ – середньоарифметичне значення аргументів першого вектора $v_{1,j}$ з j – ої комп'ютерної станції; $v_{2,j,s}$ – середньоарифметичне значення аргументів першого вектора $v_{2,j}$ з j – ої комп'ютерної станції; q – кількість значень.

Згідно значень формул (1) та (4) здійснимо централізацію значень векторів так:

$$v_{1,j,i,c} = v_{1,j,i} - v_{1,j,s}, v_{2,j,i,c} = v_{2,j,i} - v_{2,j,s}, \quad (5)$$

де $v_{1,j,i,c}$ – централізоване значення першого вектора $v_{1,j}$ з j – ої комп'ютерної станції; $v_{2,j,i,c}$– централізоване значення першого вектора $v_{2,j}$ з j – ої комп'ютерної станції; $i = 1, 2, \dots, q$; q – кількість значень.

Для отриманих централізованих даних будуємо коваріаційну матрицю, яка описує сумісне чередування декількох змінних, зокрема для розглядуваного випадку – двох змінних. Головна діагональ матриці коваріацій містить дисперсії ознак, а інші елементи містять коваріації один з одним. Дисперсію ознак визначаємо так:

$$s^2_{v_{1,j}} = \frac{1}{q-1} \sum_{i=1}^{q} (v_{1,j,i,c})^2, s^2_{v_{2,j}} = \frac{1}{q-1} \sum_{i=1}^{q} (v_{2,j,i,c})^2, \quad (6)$$

де $s^2_{v_{1,j}}$ – квадрат дисперсії значень вектора $v_{1,j}$ з j – ої комп'ютерної станції;

$s^2_{v_{2,j}}$ – квадрат дисперсії значень вектора $v_{2,j}$ з j – ої комп'ютерної станції; $i = 1, 2, \dots, q$; q – кількість значень.

Коефіцієнт коваріації ознак векторів $v_{1,j}$ та $v_{2,j}$ між собою визначимо за формулою так:

$$cov(v_{1,j}, v_{2,j}) = \frac{1}{1-q} \sum_{i=1}^{q} (v_{1,j,i,c} \cdot v_{2,j,i,c}), \quad (7)$$

де $cov(v_{1,j}, v_{2,j})$ – коваріація ознак векторів $v_{1,j}$ та $v_{2,j}$ між собою.

Коефіцієнт парної кореляції ознак векторів $v_{1,j}$ та $v_{2,j}$ визначаємо за формулою так:

$$r_{v_{1,j},v_{2,j}} = \frac{cov(v_{1,j},v_{2,j})}{s_{v_{1,j}} \cdot s_{v_{2,j}}}, \quad (8)$$

де $r_{v_{1,j},v_{2,j}}$ – коефіцієнт кореляції.

Побудуємо коваріаційну матрицю $cov(K)$ так:

$$cov(K) = \begin{pmatrix} s^2_{v_{1,j}} & s_{v_{1,j},v_{2,j}} \\ s_{v_{2,j},v_{1,j}} & s^2_{v_{2,j}} \end{pmatrix}, \quad (9)$$

де $s_{v_{2,j},v_{1,j}}$ та $s_{v_{1,j},v_{2,j}}$ - коваріації між значеннями компонентів векторів; $s_{v_{1,j},v_{1,j}} = s^2_{v_{1,j}}$; $s_{v_{2,j},v_{2,j}} = s^2_{v_{2,j}}$; $s_{v_{2,j},v_{1,j}} = s_{v_{1,j},v_{2,j}}$.

Для оцінки внеску головних компонент в загальну мінливість знайдемо власні значення. Дисперсія вздовж власних векторів є пропорційною їх власним значенням. Власних векторів буде стільки, скільки початкових змінних, тобто для розглядуваного прикладу їх буде два. Власні вектори будуть перпендикулярними між собою і вони задають напрями осей головних компонент. Вздовж першого вектора буде відкладена максимальна дисперсія даних, в вздовж наступного власного вектора – максимальна дисперсія з тих значень що залишились.

За допомогою власних значень і власних векторів знаходимо в просторі ознак нові осі, вздовж яких буде максимальне розсіювання точок. В результаті знаходимо нові координати точок в новому координатному просторі. Власні значення та власні вектори знаходимо з характеристичного рівняння так:

$$det(cov(K) - \lambda \cdot E) = 0, \quad (10)$$

де $det(cov(K) - \lambda \cdot E)$ – це детермінант матричного виразу $(cov(K) - \lambda \cdot E)$; λ – набір власних значень.

В розглядуваному випадку кількість значень λ буде два. Якщо врахувати третю ознаку в мережному

трафіку, яка вказує на кількість мережних пакетів, то кількість значень λ буде три і, відповідно, власних векторів буде три. Аналогічно, в залежності від кількості ознак факторів, кількість значень λ може бути більшою.

Знаходження власних значень і власних векторів за формулою (10) дає змогу зменшити розмірність та визначити найбільш суттєві ознаки факторів. За даними збору мережного трафіку, які представлено в матриці, що відображають данні часового ряду. Здійснимо центрування даних матриці для отримання нульового середнього значення. Кількість рядків матриці відображатиме кількість об'єктів, з яких отримуються данні. Причому кожен рядок матриці відповідатиме вектору вимірювань для всіх ліній за один крок вимірювань за певним часом. Збір даних здійснюється самоорганізованою розподіленою системою в центр, в якому відбувається обробка. Нехай скінчена кількість векторів $v_{1,j}, v_{2,j}, \dots, v_{f_j,j} \in R^u$, де вектори $v_{1,j}, v_{2,j}, \dots, v_{f_j,j}$ відповідають значенням мережного трафіку з магістральних ліній, данні про які зібрано в j – ту комп'ютерну станцію, а f_j – кількість магістральних ліній, данні з який отримуються в j – ту комп'ютерну станцію. Тоді, вектор $v_{i,j}$ позначає вектор вимірювань так:

$$v_{i,j} = v_{i,j}(t), \tag{11}$$

де t – час, коли відбулось вимірювання.

Самоорганізована розподілена система здійснює в центрі усереднення отриманих значень виконує метод головних компонент для отриманих даних та здійснює обчислення за методом головних компонент в кожній своїй компоненті, що розміщена в різних комп'ютерних станціях. Отримана інформація з комп'ютерних станціій про результати обробки надсилається в центр системи і там приймається рішення про наявність аномальних проявів. На відміну від класичного підходу з одним центром прийняття рішення, в такому варіанті з'являються варіації даних, що пов'язано з можливою неповнотою даних, що отримуються в усі комп'ютерні станції. Це надає змогу здійснити обчислення результатів в різних компонентах самоорганізованої розподіленої системи і обробити отримані результати з більшою точністю достовірності. Схема основних кроків удосконаленого методу представлена на рис. 2.

В результаті обробки даних ознак зібраних у місці призначення мережного трафіку, встановлено низьку внутрішню розмірність магістралей. Базові нормальні потоки трафіку ефективно знаходяться в низькому *l*-мірному підпросторі, який називається нормальним підпростором трафіку. Ті (*n* - *l*) головні компоненти, які залишаються, складають аварійний підпростір мережного трафіку. Виявлення аномалій обсягу мережного трафіку покладається на розкладання потоку трафіку $v_j = v_j(t)$, який отримується і обробляється в j – комп'ютерній станції, в будь-який час на компоненти, які поділимо на нормальні та аварійні. Представимо їх так:

$$v_j = v_{j,n} + v_{j,a}, \tag{12}$$

де $v_{j,n}$ - відповідає змодельованому нормальному трафіку, тобто отримана проєкція v_j на нормальний підпростір; $v_{j,a}$ - відповідає залишковому трафіку, тобто отримана проєкція v_j на аварійний підпростір.

Обчислення значень $v_{j,n}$ та $v_{j,a}$ здійснюємо з використанням методу головних компонент, вибираючи при цьому перші *k* основних компонентх ті, які отримують домінуючу відмінність в даних. Аномалія обсягу переважно призводить до суттєвої зміни $v_{j,a}$. Тобто за результатами такого розрахунку отримується сигнал про аномалію обсягу. При цьому враховується порогова статистична величина, яка розраховується за певного визначеного довірчого рівня. Для реалізації запропонованого удосконаленого методу виявлення аномалії за методом головних компонент в комп'ютерних системах в мережі використаємо розподілене відстеження мережного трафіку. Застосування методу головних компонент не до однієї комп'ютерної станції, а до групи станцій, в яких встановлена самоорганізована розподілена система виявлення аномалій в комп'ютерних системах в мережі. Враховуючи великий обсяг даних, що надходить і потребує оперативного аналізу, необхідно у вузлових компонентах самоорганізованої розподіленої системи застосовувати метод головних компонент, щоб скоротити обсяг даних, які кожна компонента системи обробляє окремо і які надсилає в центр системи. При цьому необхідно, щоб в центрі системи та у її компонентах дійсно здійснювалось виявлення аномалій. Точність відстеження мережного трафіку не є

обов'язковим, а може бути приблизним після застосування методу головних компонент. Головне завдання імплементованого в систему методу полягає в виокремленні стану в момент аномального прояву, тобто відстежувати стан дуже точно не потрібно за наявності нормальних умов. Це надає змогу скоротити обсяг даних і відповідно прискорює їх обмін між компонентами системи.

Результати експериментальних досліджень

Для підтвердження ефективності удосконаленого методу реалізуємо його в самоорганізованій розподіленій системі. Самоорганізовану розподілену систему виявлення аномалії в комп'ютерних системах згідно розробленої її архітектури, методу підтримки цілісності та методу виявлення аномалій реалізуємо проміжним програмним забезпечення, що об'єнуватиме в одне ціле комп'ютерні станції в мережі, з двома типами інтерфейсу: для компоненти, в якій міститься центр прийняття рішень верхнього рівня ієрархії: для компонент, в яких міститься центр прийняття рішень нижнього рівня ієрархії. Зображення віконної форми інтерфейсу компоненти розподіленої системи, в якій міститься центр верхнього рівня ієрархії, представлено на рис. 4.1.

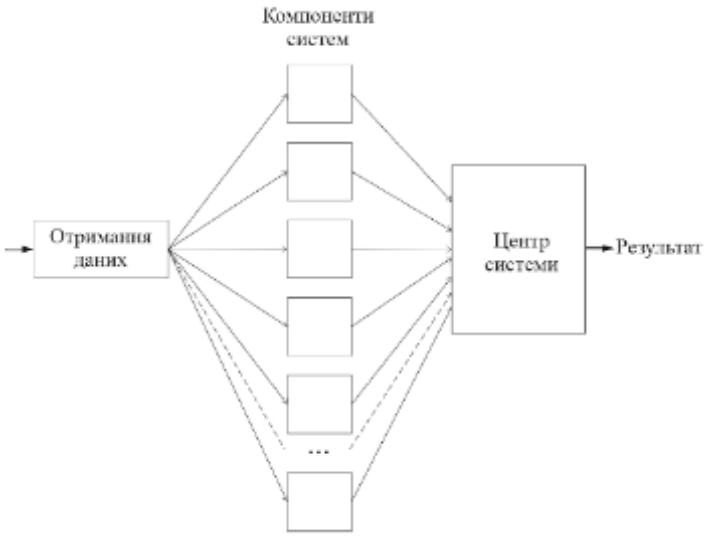

Рис. 2. Схема основних кроків удосконаленого методу

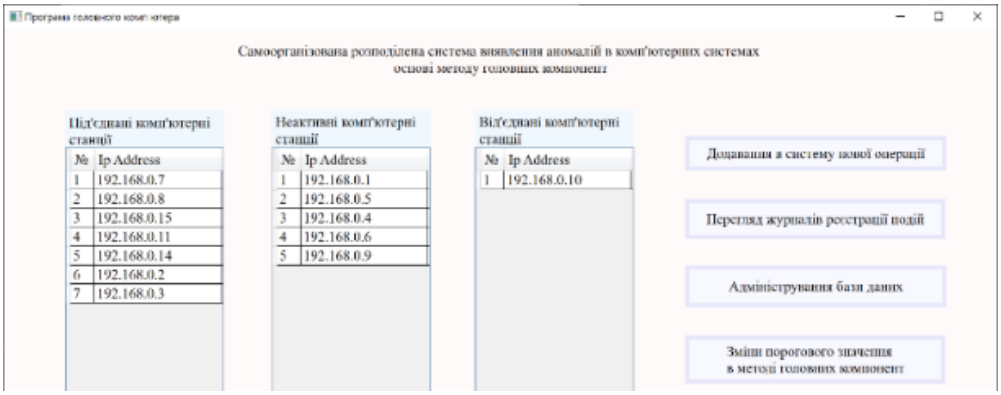

Рис. 3. Віконна форма інтерфейсу розробленої розподіленої системи

Достовірність виявлення аномалії в комп'ютерних системах потребує дослідження для встановлення можливості її використання в реальних умовах. Частина розподіленої системи, яка відповідає за виявлення аномалії, імплементована, як відповідний метод в неї, і її тестування дозволить оцінити достовірність з виявлення аномалії в комп'ютерних системах. Тому, розглянемо тестування розподіленої системи спочатку з відімкненим модулем, що відповідає за виявлення аномалії. Показники, які досліджуватимемо згрупуємо за такими характеристиками: час комунікації між окремими компонентами; час комунікації між компонентами системи і компонентою, в якій розміщено центр прийняття рішень вищого рівня ієрархії; час комунікації між компонентами в залежності від кількості активних компонентів, які формують систему; час, який витрачено на роботу, коли в системі здійснено розподілення центру, тобто тестування з розподіленим центром, і час, який витрачений, коли центр не розподіляється між рівнями ієрархії, а знаходиться повністю в одній компоненті.

Експериментальні дослідження з розробленою самоорганізованою розподіленою системою проведено в локальній комп'ютерній мережі, створеній за технологією Ethernet з швидкістю передачі даних 1 Гб/с між вузлами в мережі. Оскільки повідомлення для передачі між компонентами системи містять дуже невелику кількість інформації, то вони формуються в короткі пакети і вважатимемо, що кожне з них передавалось одним пакетом обсягом 64 байти. Експериментальні дослідження проводились протягом 50 діб окремо для випадку, коли центр прийняття рішень був розподілений між всіма компонентами з врахуванням двох рівнів ієрархії і так само 50 діб для випадку, коли центр прийняття рішень був розміщений тільки в одній компоненті. На протязі всього часу експерименту було здійснено фіксування часу відправки повідомлень, їх номеру та фіксування часу отримання. Це здійснювалось для того, щоб провести дослідження витрат часу на комунікацію в середині самої системи. **Результати експериментальних досліджень, представлені в табл. 1, підтверджують збільшення кількості переданих повідомлень** для випадку, коли час витрачено на роботу системи з розподіленням центру, тобто тестування з розподіленим центром, порівняно для випадку, коли час витрачено на роботу системи без розподілення центру між компонентами.

Таблиця 1.

Результати експериментальних досліджень
щодо ефективності функціонування самоорганізованої розподіленої системи

№ з. п.	Характеристика часової величини	Інтервал часу для випадку, коли час витрачено на роботу системи з розподіленням центру, тобто тестування з розподіленим центром, с	Інтервал часу для випадку, коли час витрачено на роботу системи без розподілення центру між компонентами, с
1	Час комунікації між окремими компонентами	1,42 – 2,61	1,41 – 2,56
2	Час комунікації між компонентами системи і компонентою, в якій розміщено центр прийняття рішень вищого рівня ієрархії	1,81 – 2,87	1,67 – 2,23
3	Час комунікації між компонентами в залежності від кількості активних компонентів, які формують систему. Випадки: 3.1. Кількість компонент 2-4. 3.2. Кількість компонент 5-8.	 1,41 – 2,51 1,83 – 2,81	 1,27 – 2,39 1,72 – 2,43
4	Кількість переміщених пакетів для досліджуваної події 1	7546	5788
5	Кількість переміщених пакетів для досліджуваної події 2	12458	8386

Також, за результатами проведеного експерименту було встановлено, що коли центр прийняття рішень розподілений між рівнями ієрархії, то ефективність за часом краща, бо обробка на нижньому рівні ієрархії скорочує час обробки події порівняно з використанням одного центру. Достовірність при обробці аномалії в комп'ютерних системах, які вводились штучно, становить 0,8356, що є задовільним результатом і підтверджує достатню ефективність запропонованих рішень.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

В роботі удосконалено метод централізованого виявлення розподілених аномалій за алгоритмом пошуку головних компонент, який на відміну від відомих імплементований в самоорганізовану розподілену систему з центром прийняття рішень та надає змогу визначати аномальні прояви на основі обробки даних в центрі та компонентах системи одночасно з подальшим їх усередненням. В результаті такого застосування методу в частини даних буде зменшено розмірність з моменту отримання, а в другої частини даних після надсилання в центр. Але їх обробка стане уточненням оброблених в центрі даних отриманих після першої обробки в компонентах системи.

Розроблене програмне забезпечення для забезпечення функціонування самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах підтверджує можливість реалізації запропонованих рішень. Проведені експериментальні дослідження з розробленою реалізацією самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах згідно отриманих коефіцієнтів підтверджують ефективність запропонованих рішень і розробленої розподіленої системи щодо її функціонування в комп'ютерній мережі.

Література

1. Savenko O. Metamorphic Viruses' Detection Technique Based on the Equivalent Functional Block Search / O. Savenko, S. Lysenko, A. Nicheporuk, B. Savenko // CEUR-WS, ISSN: 1613–0073. – 2017. – Vol. 1844. – Pp. 555–569.

2. Lysenko S., Bobrovnikova K., Matiukh S., Hurman I., Savenko O. Detection of the botnets' low-rate DDoS attacks based on self-similarity. International Journal of Electrical and Computer Engineering, Vol. 10, Issue 4, 2020, Pages 3651-3659. DOI: http://doi.org/10.11591/ijece.v10i4.pp3651-3659.

3. Pomorova O. Metamorphic Viruses Detection Technique based on the the Modified Emulators [Text] / O. Pomorova, O. Savenko, S. Lysenko, A. Nicheporuk // CEUR-WS. – 2016. – Vol. 1614. – PP.375-383, ISSN: 1613-0073

4. Wawryn, K., Widuliński P. Detection of anomalies in compiled computer program files inspired by immune mechanisms using a template method. Journal of Computer Virology and Hacking Techniques. 2020. https://doi.org/10.1007/s11416-020-00364-w

5. Zeng J., Tang W. (2015) Negative Selection Algorithm Based Unknown Malware Detection Model. In: Gong M., Linqiang P., Tao S., Tang K., Zhang X. (eds) Bio-Inspired Computing - Theories and Applications. BIC-TA 2015. Communications in Computer and Information Science, vol. 562. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-49014-3_53

6. Корченко А. О. Методи ідентифікації аномальних станів для систем виявлення вторгнень: автореф. дис. ... д-ра техн. наук: 05.13.21, Київ, 2019, 40 с.

7. Лукова-Чуйко Н. В. Методологічні основи забезпечення функціональної стійкості розподілених інформаційних систем до кібернетичних загроз: автореф. дис. ... д-ра техн. наук: 05.13.06, Київ, 2018, 40 с.

8. B. Savenko, S. Lysenko, K. Bobrovnikova, O. Savenko, G. Markowsky. Detection DNS Tunneling Botnets // Proceedings of the 2021 IEEE 11th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IDAACS'2021, Cracow, Poland, September 22-25, 2021.

9. Savenko, O., Nicheporuk, A., Hurman, I., Lysenko, S. - CEUR-WS. – 2019. – Vol. 2393. – P.633-643, ISSN: 1613-0073.

10. Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu. A survey of network anomaly detection techniques. Journal of Network and Computer Applications Vol. 60, January 2016. P. 19-31.

11. Bernadette J. Stolz, Jared Tanner, Heather A. Harrington, Vidit Nanda Geometric anomaly detection in data. Proceedings of the National Academy of Sciences Aug 2020, 117 (33) 19664-19669; DOI: 10.1073/pnas.2001741117

12. Xiang Yu, Hui Lu, Xianfei Yang, Ying Chen, Haifeng Song, Jianhua Li, Wei Shi An adaptive method based on contextual anomaly detection in Internet of Things through wireless sensor networks. International Journal of Distributed Sensor Networks 2020. Vol. 16(5) DOI: 10.1177/1550147720920478

13. Goldstein M., Uchida S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. 2016. PLOS ONE 11(4): e0152173. https://doi.org/10.1371/journal.pone.0152173

14. Hayes, M.A., Capretz, M.A. Contextual anomaly detection framework for big sensor data. Journal of Big Data 2, 2. 2015. https://doi.org/10.1186/s40537-014-0011-y

15. Liu, L., Hu, M.; Kang, C., Li, X. Unsupervised Anomaly Detection for Network Data Streams in Industrial Control Systems. Information 2020, 11, 105. https://doi.org/10.3390/info11020105

16. Xiaodan Xu, Huawen Liu, Minghai Yao. Recent Progress of Anomaly Detection. Complexity, vol. 2019, Article ID 2686378, 11 pages, 2019. https://doi.org/10.1155/2019/2686378

17. Jianwen Huang, Zhen Chai and Hailong Zhu. Detecting anomalies in data center physical infrastructures using statistical approaches. Journal of Physics: Conference Series, Volume 1176, Issue 2. Jianwen Huang et al 2019 J. Phys.: Conf. Ser. 1176 022056

18. Fisch, A., Grose, D., Eckley, I.A., Fearnhead, P., & Bardwell, L. (2020). anomaly: Detection of Anomalous Structure in Time Series Data. arXiv: Applications.arXiv:2010.09353

19. Lu, X., Wang, S., Kang, F., Liu, S., Li, H., Xu, X. and Cui, L. (2019). An anomaly detection method to improve the intelligent level of smart articles based on multiple group correlation probability models. International Journal of Crowd Science, Vol. 3, № 3. P. 333-347. https://doi.org/10.1108/IJCS-09-2019-0024

20. Solarz A., Bilicki M., Gromadzki M., Pollo A. , Durkalec A., Wypych M. Automated novelty detection in the WISE survey with one-class support vector machines. Published online: 05 October 2017. DOI: 10.1051/0004-6361/201730968

21. Mukrimah Nawir, Amiza Amir, Naimah Yaakob, Ong Bi Lynn. Effective and efficient network anomaly detection system using machine learning algorithm. Bulletin of Electrical Engineering and InformaticsVol.8, No.1, March 2019, pp. 46~51ISSN: 2302-9285, DOI: 10.11591/eei.v8i1.1387

References

1. Savenko O. Metamorphic Viruses Detection Technique Based on the Equivalent Functional Block Search / O. Savenko, S. Lysenko, A. Nicheporuk, B. Savenko // CEUR-WS, ISSN: 1613–0073. – 2017. – Vol. 1844. – Pp. 555–569.

2. Lysenko S., Bobrovnikova K., Matiukh S., Hurman I., Savenko O. Detection of the botnets low-rate DDoS attacks based on self-similarity. International Journal of Electrical and Computer Engineering, Vol. 10, Issue 4, 2020, Pages 3651-3659. DOI: http://doi.org/10.11591/ijece.v10i4.pp3651-3659.

3. Pomorova O. Metamorphic Viruses Detection Technique based on the the Modified Emulators [Text] / O. Pomorova, O. Savenko, S. Lysenko, A. Nicheporuk // CEUR-WS. – 2016. – Vol. 1614. – PP.375-383, ISSN: 1613-0073

4. Wawryn, K., Widuliński P. Detection of anomalies in compiled computer program files inspired by immune mechanisms using a template method. Journal of Computer Virology and Hacking Techniques. 2020. https://doi.org/10.1007/s11416-020-00364-w

5. Zeng J., Tang W. (2015) Negative Selection Algorithm Based Unknown Malware Detection Model. In: Gong M., Linqiang P., Tao S., Tang K., Zhang X. (eds) Bio-Inspired Computing - Theories and Applications. BIC-TA 2015. Communications in Computer and Information Science, vol. 562. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-49014-3_53

6. Korchenko A. O. Metody identyfikatsii anomalnykh staniv dlia system vyiavlennia vtorhnen: avtoref. dys. ... d-ra tekhn. nauk: 05.13.21, Kyiv, 2019, 40 s.

7. Lukova-Chuiko N. V. Metodolohichni osnovy zabezpechennia funktsionalnoi stiikosti rozpodilenykh informatsiinykh system do kibernetychnykh zahroz: avtoref. dys. ... d-ra tekhn. nauk: 05.13.06, Kyiv, 2018, 40 s.

8. B. Savenko, S. Lysenko, K. Bobrovnikova, O. Savenko, G. Markowsky. Detection DNS Tunneling Botnets // Proceedings of the 2021 IEEE 11th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IDAACS2021, Cracow, Poland, September 22-25, 2021.

9. Savenko, O., Nicheporuk, A., Hurman, I., Lysenko, S. - CEUR-WS. – 2019. – Vol. 2393. – P.633-643, ISSN: 1613-0073.

10. Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu. A survey of network anomaly detection techniques. Journal of Network and Computer Applications Vol. 60, January 2016. P. 19-31.

11. Bernadette J. Stolz, Jared Tanner, Heather A. Harrington, Vidit Nanda Geometric anomaly detection in data. Proceedings of the National Academy of Sciences Aug 2020, 117 (33) 19664-19669; DOI: 10.1073/pnas.2001741117

12. Xiang Yu, Hui Lu, Xianfei Yang, Ying Chen, Haifeng Song, Jianhua Li, Wei Shi An adaptive method based on contextual anomaly detection in Internet of Things through wireless sensor networks. International Journal of Distributed Sensor Networks 2020. Vol. 16(5) DOI: 10.1177/1550147720920478

13. Goldstein M., Uchida S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. 2016. PLOS ONE 11(4): e0152173. https://doi.org/10.1371/journal.pone.0152173

14. Hayes, M.A., Capretz, M.A. Contextual anomaly detection framework for big sensor data. Journal of Big Data 2, 2. 2015. https://doi.org/10.1186/s40537-014-0011-y

15. Liu, L., Hu, M.; Kang, C., Li, X. Unsupervised Anomaly Detection for Network Data Streams in Industrial Control Systems. Information 2020, 11, 105. https://doi.org/10.3390/info11020105

16. Xiaodan Xu, Huawen Liu, Minghai Yao. Recent Progress of Anomaly Detection. Complexity, vol. 2019, Article ID 2686378, 11 pages, 2019. https://doi.org/10.1155/2019/2686378

17. Jianwen Huang, Zhen Chai and Hailong Zhu. Detecting anomalies in data center physical infrastructures using statistical approaches. Journal of Physics: Conference Series, Volume 1176, Issue 2. Jianwen Huang et al 2019 J. Phys.: Conf. Ser. 1176 022056

18. Fisch, A., Grose, D., Eckley, I.A., Fearnhead, P., & Bardwell, L. (2020). anomaly: Detection of Anomalous Structure in Time Series Data. arXiv: Applications.arXiv:2010.09353

19. Lu, X., Wang, S., Kang, F., Liu, S., Li, H., Xu, X. and Cui, L. (2019). An anomaly detection method to improve the intelligent level of smart articles based on multiple group correlation probability models. International Journal of Crowd Science, Vol. 3, № 3. P. 333-347. https://doi.org/10.1108/IJCS-09-2019-0024

20. Solarz A., Bilicki M., Gromadzki M., Pollo A. , Durkalec A., Wypych M. Automated novelty detection in the WISE survey with one-class support vector machines. Published online: 05 October 2017. DOI: 10.1051/0004-6361/201730968

21. Mukrimah Nawir, Amiza Amir, Naimah Yaakob, Ong Bi Lynn. Effective and efficient network anomaly detection system using machine learning algorithm. Bulletin of Electrical Engineering and InformaticsVol.8, No.1, March 2019, pp. 46~51ISSN: 2302-9285, DOI: 10.11591/eei.v8i1.1387