МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ І ПРОГРАМНОЇ ІНЖЕНЕРІЇ

ПАПЕРОВСЬКИЙ БОГДАН ОЛЕГОВИЧ

УДК 004.415.5

Аналіз сучасних методів виявлення аномалій та можливих втручань в роботу комп'ютерної системи

Спеціальність 125 «Кібербезпека»

Автореферат

дипломної роботи на здобуття освітньо-кваліфікаційного рівня «магістр»

Тернопіль — 2018

Роботу виконано на кафедрі кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

Керівник роботи:	кандидат технічних наук, зав. кафедри кібербезпеки Загородна Наталія Володимирівна Тернопільський національний технічний університет імені Івана Пулюя,
Рецензент:	кандидат технічних наук, доцент Золотий Роман Захарійович Тернопільський національний технічний університет імені Івана Пулюя

Захист відбудеться 26 грудня 2018 р. о 9.00 годині на засіданні екзаменаційної комісії №32 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 806.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність дослідження. Аномальні структури в наборі даних, які не відповідають більшості даних, зазвичай називають відхиленнями або аномаліями. У багатьох сферах, таких як виявлення шахрайства, моніторинг навколишнього середовища та медична діагностика, одним з головних завдань є виявлення таких випадків або їх усунення . Двома основними припущеннями багатьох існуючих методів виявлення викидів є особливість викидів та відмінність між такими викидами та звичайними даними.

Дана робота є актуальною, так як:

- на ринку інформаційних технологій існує потреба в виявленні раніше невідомих атак, так як вони з'являються з великою швидкістю;
- системи виявлення вторгнення стають все більш популярними і їх застосування повинно бути досліджено;
- методи виявлення аномалій на основі машинного навчання, мають велику кількість помилкових спрацювань, а отже потребують оптимізації та вдосконалення.

Мета і завдання дослідження. Провести порівняльний аналіз методів машинного навчання виявлення аномалій з метою підвищення ефективності роботи систем виявлення втручань в роботу комп’ютерної системи. Запропонувати удосконалену методику виявлення аномалій з метою зменшення кількості хибних спрацювань, а отже підвищення точності роботи системи виявлення втручань.

Об’єктом дослідження є аномалії роботи комп’ютерних систем

Предметом дослідження є методи виявлення аномальної поведінки.

Методи дослідження. В процесі дослідження використано загальнонаукові методи пізнання: порівняння, системний аналіз, моделювання. Досліджувалися методи виявлення аномалій на основі статичного аналізу, машинного навчання, індуктивного висновку, нечіткої логіки, генетичних алгоритмів та нейронних мереж.

Наукова новизна роботи: Вдосконалення методу виявлення аномалій з метою зменшення кількості помилкових спрацювань та підвищення ефективності роботи системи.

Практичне значення дослідження полягає наданні рекомендацій щодо покращення роботи системи виявлення аномалій.

Апробація результатів дипломної роботи. Основні положення дослідження доповідалися й обговорювалися на науково-практичних конференціях: на VII Міжнародній науково-технічній конференції молодих учених та студентів “Актуальні задачі сучасних технологій” (Тернопіль, 28-29 листопада 2018 року) та VI Науково-технічна конференція "Інформаційні моделі, системи та технології" (Тернопіль, 12-13 грудня 2018 року).

Структура роботи. Дипломна робота складається із вступу, семи розділів, висновків, списку використаних джерел із найменувань. Робота містить рисунків 34, 3 таблиці і 39 формул. Обсяг основного тексту становить 137 сторінок, перелік використаних джерел 3 сторінки, додатки 7 сторінок. Загальний обсяг дипломної роботи складає 145 сторінок.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність проблеми, визначено об’єкт і предмет дослідження, сформульовано його мету, завдання, розкрито теоретичну та методологічну основу, методи дослідження; висвітлено наукову новизну, практичне значення роботи;.

У першому розділі — *“Огляд сучасних методів виявлення аномалій в комп’ютерних системах”* — проаналізовано визначення поняття аномалії, коротко описано причини їх виникнення. Розгляну три типи аномалій: поодинокі (точкові), контекстуальні та групові аномалії. Також описано на основі яких характеристик відбувається виявлення аномальної поведінки. Розглянуто методи виявлення аномалій на основі кількості відомої інформації про вибірку даних.

У другому розділі — *“Системи виявлення вторгнень на основі машинного навчання”* — було описано основну системи виявлення вторгнень, сфери застосування таких систем та їх класифікацію. Розглянуто особливості розробки: вузлових, мережевих та систем виявлення вторгнень у роботу веб-додатку. Був проведений аналіз методів машинного навчання для виявлення аномалій. Також було розглянуто проблеми застосування таких методів у справжніх системах та робота таких систем після виявлення аномальної активності.

У третьому розділі − “***Вдосконалення методу SVM з метою зменшення кількості хибних спрацювань*** ”− детально описано та проаналізовано процес однокласової класифікації, розглянуто метод SVM та OCSVM. Представлено покращений метод на базі OC SVM та розказано про принцип його роботи. Були показані результати порівняння OC SVM та покращеного OC SVM. Проаналізовано переваги та певні обмеження, які властиві вдосконаленому методу.

У спеціальній частині – проведено аналіз технічних рішень та надано рекомендації для ефективної розробки системи виявлення аномалій.

В частині «Обґрунтування економічної ефективності» проведено розрахунки капітальних та поточних витрат, рентабельність інвестицій, обчислено прогнозований ефект від впровадження системи та термін окупності інвестицій.

В частині «Охорона праці та безпека в надзвичайних ситуаціях» розглянуто важливість психологічного стану та місця роботи для людей, що працюють із комп’ютером.

В частині «Екологія» розглянуто формування бази екологічних банків даних та описано сучасні системи керування такими даними.

У загальних висновках щодо дипломної роботи наведено короткий опис основної частини; сформульовано основні результати, отримані в роботі та сформульовано рекомендації для організацій, що мають справу з інтелектуальною власністю.

В додатках до пояснювальної записки приведено тези.

Та описані характеристики ознак ефективності роботи методів виявлення аномалій.

ВИСНОВКИ

У дипломній роботі запропоновано актуальне рішення до проблеми знаходження аномалій у множені даних, що стосується вдосконалення методу SVM, який використовують для виявлення відхилень. В ході вирішення поставленого завдання були отримані наступні аналітичні та практичні результати:

1) Проведено огляд джерел в області дослідження;
2) виконано аналіз сучасних методів виявлення з використанням різних підходів заснованих на щільності, природі, та інформативності набору даних. Проведено порівняння існуючих методів, яке показало, що у великій кількісті алгоритмів присутня проблема високого рівня помилкових спрацювань. У зв'язку з цим є необхідність в покращенні алгоритмів детектування аномалій, що дозволяє зменшити кількість хибних виявлень.
3) запропоновано шляхи підвищення роботи систем виявлення аномалій;
4) запропоновано вдосконалений метод поліпшення якості виявлення аномалій у наборі даних;
5) проаналізовано архітектуру та принцип роботи систем виявлення аномалій;
6) протестовано ефективність роботи вдосконаленого методу;
7) проаналізовано сучасні технологічні рішення для розробки систем виявлення аномалій;
8) доведено економічну доцільність розробки і впровадження розробленого алгоритму.

З практичної точки зору запропонований алгоритм може бути використаний для інтегрування в існуючу систему детектування виявлення аномалій для зменшення хибних виявлень. Аналіз аномалій засобами машинного навчання має великий простір подальших досліджень, особливо в області кібербезпеки та медицини. Вдосконалений алгоритм має певні обмеження, що і може бути предметом подальших досліджень.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ

1. Паперовський Б.О. Аналіз сучасних методів виявлення аномалій та можливих вторгнень в роботу комп'ютерної системи/ Тези доповіді на VII Міжнародній науково-технічній конференції молодих учених та студентів “Актуальні задачі сучасних технологій” 28-29 листопада 2018 року. – Том II. – Тернопіль, 2018. – С. 146 ст.

2. Паперовський Б.О. Аналіз сучасних методів виявлення аномалій та можливих вторгнень в роботу комп'ютерної системи/ Тези доповіді VI Науково-технічна конференція "Інформаційні моделі, системи та технології", 12-13 грудня 2018 року. – Том II. – Тернопіль, 2018. – С. ст.

АНОТАЦІЇ

Ключові слова: АНОМАЛІЇ, МЕТОДИ ВИЯВЛЕННЯ АНОМАЛІЙ, СИСТЕМИ ВИЯВЛЕННЯ ВТОРГЕНЬ, МЕТОД ОПОРНИХ ВЕКТОРІВ, МЕТОД ЗМЕНШЕННЯ ЙМОВІРНОСТІ ХИБНИХ ВИЯВЛЕНЬ

Об'єкт дослідження: аномалії роботи комп’ютерних систем

Мета роботи (проекту): Провести порівняльний аналіз методів машинного навчання виявлення аномалій з метою підвищення ефективності роботи систем виявлення втручань в роботу комп’ютерної системи. Запропонувати шляхи підвищення роботи таких систем. Оцінити точність роботи систем виявлення аномалій. Запропонувати удосконалену методику виявлення аномалій з метою зменшення кількості хибних спрацювань, а отже підвищення точності роботи системи виявлення втручань. Перевірити ефективність роботи вдосконаленого алгоритиму на тестових даних.

Методи дослідження: аналіз, системний підхід, методи: статистичні, контрольованого навчання, неконтрольованого навчання, на основі щільності розподілу набору даних, кластеризації та класифікації.

У спеціальній частині проведено аналіз сучасних технологічних рішень для швидкої та ефективної побудови систем виявлення аномалій. Сформульовано формальні вимоги до вибору методів детектування. Запропоновано метод поліпшення якості виявлення аномалій на основі однокласових опорних векторів. Розроблено алгоритм зменшення помилкових спрацювань системи виявлення аномалій.

В економічному розділі визначено економічну ефективність від розробки і реалізації запропонованого алгоритму.

Практичне значення роботи полягає в можливості інтеграції розробленого алгоритму в існуючу систему виявлення аномалій для покращення ефективності аналізу та розпізнавання даних.

Результати проведених в дипломній роботі досліджень можуть бути використані для подальшої роботи над удосконаленням алгоритмів виявлення аномалій.

Наукова новизна дослідження полягає в удосконаленні методу однокласових опорних векторів, для з метою зменшення кількості хибних спрацювань, а отже підвищення точності роботи системи виявлення втручань.

Keywords: ANOMALIES, METHODS OF DETECTING ANOMALIES, INTRUDION DETECTION SYSTEMS, METHOD SUPORT VECTOR MACHINE, FALSE ALARM REDUCTION METHOD.

Project objectives: computer system anomalies.

Project purpose: To conduct a comparative analysis of methods of technical training to detect abnormalities with the effective work of systemic detection of interventions in the work of the system. Suggest ways to work such systems. Evaluate the accuracy of the system of detection of anomalies. We propose to improve the technique of detecting anomalies with local reductions in the number of hybrid species, as well as to determine

the accuracy of the system for detecting interventions. Check the effectiveness of the improved algorithm on the test data.

Research methods: analysis, systematic approach, methods: statistical, supervised learning, unsupervised learning, based on the density of data set distribution, clustering and classification.

In the technical part, existing methods of detecting anomalies are analyzed and an improved SVM algorithm. Analysis of existing modern detection methods which use various recognition approaches and clustering methods is carried out. Requirements for approaches and clustering methods is carried out. Requirements fo an algorithm for reducing false alarms has been developed.

In the economic section, economic efficiency of proposed algorithm is assesed.

Practical value of this work is the ability to integrate developed algorithm into existing system for detecting anomalies in different sets of data.

The results of the research can be used for further work on improving the algorithms for detecting anomalies.

The scientific novelty of the research is to improve the method of one-class support vector machine, in order to reduce the number of false positives alarms, and thus increase the accuracy of the system of detection of interventions.