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

КАШИЦЬКИЙ ОЛЕКСАНДР ВОЛОДИМИРОВИЧ

УДК

МЕТОДИ І ЗАСОБИ ВИЯВЛЕННЯ АНОМАЛІЙ В БАГАТОФАКТОРНИХ КОЛЕКЦІЯХ ДАНИХ

122 – Комп’ютерні науки

Автореферат
дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль

Роботу виконано на кафедрі комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

Керівник роботи: кандидат економічних наук,
доцент кафедри комп'ютерних наук
Струтинська Ірина Володимирівна,
Тернопільський національний технічний університет імені Івана Пулюя,

Рецензент: кандидат педагогічних наук,
доцент кафедри кібербезпеки
Кареліна Олена Володимирівна,
Тернопільський національний технічний університет імені Івана Пулюя,

Захист відбудеться __ грудня 2018 р. о $9^{.00}$ годині на засіданні екзаменаційної комісії №__ у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 702

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. На сьогоднішній день потреби в додатках для виявлення аномалій швидко зростають, особливо в Інтернет-безпеці. Відповідно до звіту про загрозу безпеці інтернет-безпеки Symantec в 2015 році, фішка атаки зі співом є однією з найсильніших атак, як правило, це повідомлення електронної пошти, які, ймовірно, надходять з надійного джерела, тоді як, натиснувши посилання наданих електронних листів, ваші особисті дані будуть вкрадений Зі зростанням ймовірності нападу, цей напад впливає на сфери продажу, фінанси та операції. Програма виявлення аномалій може бути використана для контролю над незвичною поведінкою користувачів та заблокувати ці атаки заздалегідь.

Мета і задачі дослідження. Метою даної дипломної роботи є розробка багатофакторної схеми виявлення аномалій потоку з детекторами аномалії точка та потоку для легкої інтеграції різних джерел даних. Також в цій роботі представлені деякі приклади реалізації різних алгоритмів, а також результати для різних установок.

Рамки, запропоновані в даній роботі, можуть бути використані в різних галузях промисловості, і ми прагнемо допомогти компаніям скорочувати споживання енергії та заощадити витрати на інфраструктуру, виявляючи невдачі вчасно та покращуючи конфіденційність та безпеку.

Об'єкт дослідження. Ми розпочали процес розробки структури, визначивши вимоги до структури та бажані інтерфейси між модулями.

Наукова новизна одержаних результатів. У цьому магістерському проекті розроблено тестову схему для тестування різних алгоритмів виявлення багатомірних потоків аномалій. Тестовий механізм складається з трьох різних частин: генераторів даних, детектори точкових аномалій, детектори потоку аномалій. Генератори даних використовуються для генерації різних джерел даних, детектори точкових аномалій використовуються для генерації балів аномалії з багатовимірних даних, і, нарешті, детектори аномалії потоку використовуються для виявлення аномалій відповідно до аномалій, отриманих раніше.

Практичне значення одержаних результатів. Генератори даних впроваджуються з джерелами даних з даних штучного моделювання, набору даних та даних часових рядів. Для точкових аномалійних детекторів використовуються Pyisc, локальний коефіцієнт викидів і векторна машина підтримки (кластер). Для детекторів аномального потоку реалізовані алгоритми DDM, CUSUM та FCWM.

Апробація результатів магістерської роботи. В цілому, з реалізації, ми бачимо, що мої рамки, запропоновані в моїй дипломній роботі, працюють дуже добре, всі класи та інтерфейси необхідні і легко бути розширені. Таким чином, користувачі могли зосередитись на конструкціях алгоритмів. Для оцінки користувачі можуть використовувати різні джерела даних, різні аномальні типи, щоб побачити, чи є алгоритми надійними чи ні. Також є сюжет, який показує результати аномалій в реальному часі та всі необхідні показники, для чого

простіше користувачеві настроювати параметри алгоритму та ефективно порівнювати з іншими алгоритмами.

Структура роботи. Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 8 частин, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 124 арк. формату А4, графічна частина – 15 слайдів презентації.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Робота присв'ячена потребам в додатках для виявлення аномалій, особливо в Інтернет-безпеці. Відповідно до звіту про загрозу безпеці інтернет-безпеки Symantec в 2015 році, фішка атаки зі співом є однією з найсильніших атак, як правило, це повідомлення електронної пошти, які, ймовірно, надходять з надійного джерела, тоді як, натиснувши посилання наданих електронних листів, ваші особисті дані будуть вкрадений Зі зростанням ймовірності нападу, цей напад впливає на сфери продажу, фінанси та операції. Програма виявлення аномалій може бути використана для контролю над незвичною поведінкою користувачів та заблокувати ці атаки заздалегідь.

В першому розділі магістерської роботи розглянуто аномалії для ідентифікації випадків з великої кількості випадків, які в основному вважаються нормальними. Розробка багатофакторної схеми виявлення аномалій потоку з детекторами аномалії точка та потоку для легкої інтеграції різних джерел даних. Процес розробки структури на основі визначених вимог до структури та бажані інтерфейси між модулями. Фундамент реалізації в об'єктно-орієнтованій мові, що спрощує розширення інтерфейсів та реалізацію нових алгоритмів. Розробку основи, впровадження декількох алгоритмів, що дозволяють правильно відображати основні функції та продемонструвати гнучкість дизайну. Як оптимізувати параметри алгоритму для різних сценаріїв та алгоритмів.

В другому розділі магістерської роботи розглянуто пов'язані роботи. Описано незаконне виявлення точкових аномалій. Досліджено розрахунок аномалій. Проаналізовано кластеризація послідовного потоку Python (Pyisc). Описано модель підтримки векторної машини та підтримки векторної кластеризації. Розглянуто ідею локального коефіцієнту викиду (lof). Визначено зміни виявлення та потоку аномалій. Представлені основні ідеї декількох алгоритмів виявлення змін. Застосовано діагностику несправностей для потоку даних.

В третьому розділі магістерської роботи розглянуто генератор потоку. Описано детектор точкових аномалій. Досліджено поточний детектор аномалій. Проаналізовано основні потоки для тестової системи.

В четвертому розділі магістерської роботи розглянуто архітектура моєї структури. Описано генератор потоку. Досліджено пункт дефектоскопії аномалій. Проаналізовано розширення алгоритму DDM, CUSUM, FCWM. Виведено числові результати. Оцінено детекторів точкових аномалій.

В п’ятому розділі магістерської роботи розглянуто детекторів точкових аномалій. Описано оцінки DDM, CUSUM, FCWM. Виведено детекторів аномалії потоку.

В спеціальній частині показана оцінка алгоритмів виявлення потоків аномалій.

В розділі «Обґрунтування економічної ефективності» є дослідження методів і засобів виявлення аномалій в багатофакторних колекціях даних, що поділяються на декілька етапів, що дозволяє полегшити і структурувати виконання роботи.

В частині «Охорона праці та безпека в надзвичайних ситуаціях» описано мікроклімат робочої зони і небезпечні й шкідливі фактори при виконанні робіт за комп’ютером. Розглянуто підвищення стійкості роботи об’єктів господарської діяльності у воєнний час та оцінка дії радіоактивного забруднення місцевості після ядерного вибуху на виробничу діяльність муніципальних підприємств, установ та організацій.

В розділі «Екологія» розглянуті структура і тендеції розвитку світової енергетики та описано статистичний аналіз тендецій і закономірностей динаміки в екології.

У загальних висновках щодо дипломної роботи описано прийняті в дипломній роботі освітнього рівня «Магістр» наукові та технічні рішення і організаційно-технічні заходи, які забезпечують виконання завдання на проектування; оригінальні технічні рішення, прийняті автором в процесі роботи; технічні рішення роботи, які можуть бути впроваджені у виробництво.

В додатках до пояснювальної записки приведено ксерокопії тез доповідей.

В графічній частині подано тему, мету, об’єкт та предмет дослідження. Подано завдання до дипломної роботи. Наявні демонстраційні можливості для керування камерами, сценаріями та сценаріями обробки даного генератора та консорціуму даних, промислових даних та даних; аномальні детектори PYISC, SVM і Löf, а також детектори аномалій DDM, CUSUM та FCWM. Підключення до всіх апаратних пристроїв, призначених для детекції аномалійних дефектів, і PYISC, і LOF, забезпечують швидкий доступ до даних, а чоловіки SVM забезпечують безпеку та безпеку у всьому світі. Для виявлення аномальних детекторів DDM видно сандолітки, що відображає високу швидкість доступу, і CUSUM, яка відображає всі дані, що містяться у вашому додатку. FCWM відіграє важливу роль у вирішенні проблем.

ВИСНОВКИ

У цьому магістерському проекті розроблено тестову схему для тестування різних алгоритмів виявлення багатомірних потоків аномалій. Тестовий механізм складається з трьох різних частин: генераторів даних, детектори точкових аномалій, детектори потоку аномалій. Генератори даних використовуються для генерації різних джерел даних, детектори точкових аномалій використовуються для генерації балів аномалії з багатовимірних даних, і, нарешті, детектори аномалії потоку використовуються для виявлення аномалій відповідно до аномалій, отриманих раніше.

Генератори даних впроваджуються з джерелами даних з даних штучного моделювання, набору даних та даних часових рядів. Для точкових аномалійних детекторів використовуються Pyisc, локальний коефіцієнт викидів і векторна машина підтримки (кластер). Для детекторів аномального потоку реалізовані алгоритми DDM, CUSUM та FCWM.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ

1. Технологія доступу до інтернет ресурсів для осіб з особливими потребами / [Кашицький О.В. та ін.]. // Матеріали Міжнародної студентської науково-технічної конференції «Природничі та гуманітарні науки. Актуальні питання» Тернопільського національного технічного університету імені Івана Пулюя, (Тернопіль, 26 – 27 квітня 2018 р.). – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя – 2018. – С. 44.

АНОТАЦІЯ

Магістерська робота присв’ячена потребам в додатках для виявлення аномалій, особливо в Інтернет-безпеці. Відповідно до звіту про загрозу безпеці інтернет-безпеки Symantec в 2015 році, фішка атаки зі співом є однією з найсильніших атак, як правило, це повідомлення електронної пошти, які, ймовірно, надходять з надійного джерела, тоді як, натиснувши посилання наданих електронних листів, ваші особисті дані будуть вкрадений Зі зростанням ймовірності нападу, цей напад впливає на сфери продажу, фінанси та операції. Програма виявлення аномалій може бути використана для контролю над незвичною поведінкою користувачів та заблокувати ці атаки заздалегідь.

В першому розділі магістерська роботи розглянуто аномалії для ідентифікації випадків з великої кількості випадків, які в основному вважаються нормальними. Розробка багатофакторної схеми виявлення аномалій потоку з детекторами аномалії точка та потоку для легкої інтеграції різних джерел даних. Процес розробки структури на основі визначених вимог до структури та бажані інтерфейси між модулями. Фундамент реалізації в об’єктно-орієнтованій мові, що спрощує розширення інтерфейсів та реалізацію нових алгоритмів. Розробку основи, впровадження декількох алгоритмів, що дозволяють правильно відображати основні функції та продемонструвати гнучкість дизайну. Як оптимізувати параметри алгоритму для різних сценаріїв та алгоритмів.

В другому розділі магістерська роботи розглянуто пов’язані роботи. Описано незаконне виявлення точкових аномалій. Досліджено розрахунок аномалій. Проаналізовано кластеризація послідовного потоку Python (Pyisc). Описано модель підтримки векторної машини та підтримки векторної кластеризації. Розглянуто ідею локального коефіцієнту викиду (lof). Визначено зміни виявлення та потоку аномалій. Представлені основні ідеї декількох алгоритмів виявлення змін. Застосовано діагностику несправностей для потоку даних.

В третьому розділі магістерська роботи розглянуто генератор потоку. Описано детектор точкових аномалій. Досліджено поточний детектор аномалій. Проаналізовано основні потоки для тестової ситеми.

В четвертому розділі магістерська роботи розглянуто архітектура моєї структури. Описано генератор потоку. Досліджено пункт дефектоскопії аномалій. Проаналізовано розширення алгоритму DDM, CUSUM, FCWM. Виведено числові результати. Оцінено детекторів точкових аномалій.

В п’ятому розділі магістерська роботи розглянуто детекторів точкових аномалій. Описано оцінки DDM, CUSUM, FCWM. Виведено детекторів аномалії потоку.

Об’єкт дослідження: процес розробки структури, визначивши вимоги до структури та бажані інтерфейси між модулями

Предмет дослідження: фундамент реалізується в об’єктно-орієнтованій мові, що спрощує розширення інтерфейсів та реалізацію нових алгоритмів.

Мета роботи: є розробка багатофакторної схеми виявлення аномалій потоку з детекторами аномалії точка та потоку для легкої інтеграції різних джерел даних. Також в цій роботі представлені деякі приклади реалізації різних алгоритмів, а також результати для різних установок.

Основні результати: в цілому, з реалізації, ми бачимо, що мої рамки, запропоновані в моїй дипломній роботі, працюють дуже добре, всі класи та інтерфейси необхідні і легко бути розширені. Таким чином, користувачі могли зосередитись на конструкціях алгоритмів. Для оцінки користувачі можуть використовувати різні джерела даних, різні аномальні типи, щоб побачити, чи є алгоритми надійними чи ні. Також є сюжет, який показує результати аномалій в реальному часі та всі необхідні показники, для чого простіше користувачеві настроювати параметри алгоритму та ефективно порівнювати з іншими алгоритмами.

Ключові слова: БАГАТОВИМІРНИЙ, ВИЯВЛЕННЯ АНОМАЛІЙ ПОТОКУ, PYISC, SVM, LOF, DDM, CUSUM, FCWM.

ANNOTATION

Diploma work to the присв'ячена requirements in additions for the exposure of anomalies, especially in Internet-safety. In accordance with a report on a threat to safety of internet-safety of Symantec in 2015, a chip of attack with singing is one of the strongest attacks, as a rule, it is reports of e-mail, that, probably, come from a reliable source, while, pressing reference of the given electronic folias, your personal data will be stolen With the increase of probability of attack, this attack influences on the spheres of sale, finances and operations. The program of exposure of anomalies can be used for control above unusual behavior of users and to block these attacks in good time.

In the first division of diploma work anomalies are considered for authentication of cases from plenty of cases that are mainly considered normal. Development of multivariable chart of exposure of anomalies of stream with the detectors of anomaly point and to the stream for easy integration of different sources of data. A development of structure process on the basis of certain requirements to the structure and desirable interfaces is between the modules. Foundation of realization is in the object-oriented language that simplifies expansion of interfaces and realization of new algorithms. Development of basis, introduction of a few algorithms that allow correctly to represent

basic functions and show design flexibility. How to optimize the parameters of algorithm for different scenarios and algorithms.

In the second division diploma work the constrained works are considered. The illegal exposure of point anomalies is described. The calculation of anomalies is investigational. It is analysed clusterization of successive stream of Python (Pyisc). The model of support of vectorial machine and support of vectorial clusterization is described. The idea of local to the coefficient extrass (lof) is considered. The changes of exposure and stream of anomalies are certain. Presented basic ideas of a few algorithms of exposure of changes. The diagnosis of faults is applied for the flow of data.

The generator of stream is considered in the third division of diploma work. The detector of point anomalies is described. The current detector of anomalies is investigational. Basic streams are analysed for test ситеми.

In the fourth division of diploma work it is considered architecture of my structure. The generator of stream is described. The point of fault detection of anomalies is investigational. Expansion of algorithm of DDM is analysed, CUSUM, FCWM. Numerical results are shown out. It is appraised detectors of point anomalies.

It is considered in the fourth division of diploma wor architecture of my structure. The generator of stream is described. The point of fault detection of anomalies is investigational. Expansion of algorithm of DDM is analysed, CUSUM, FCWM. Numerical results are shown out. It is appraised detectors of point anomalies.

In the fourth division of diploma work it is considered architecture of my structure. The generator of stream is described. The point of fault detection of anomalies is investigational. Expansion of algorithm of DDM is analysed, CUSUM, FCWM. Numerical results are shown out. It is appraised detectors of point anomalies.

In the fifth division of diploma work it is considered detectors of point anomalies. The estimations of DDM are described, CUSUM, FCWM. It is shown out detectors of anomaly of stream.

Research object: development of structure process, defining requirements to the structure and desirable interfaces between the modules

Article of research: foundation will be realized in the object-oriented language that simplifies expansion of interfaces and realization of new algorithms.

Aim of work: there is development of multivariable chart of exposure of anomalies of stream with the detectors of anomaly point and to the stream for easy integration of different sources of data. Also in this work some examples of realization of different algorithms, and also results, are presented for different options.

Basic results: on the whole, from realization, we see that my scopes offered in my diploma work work very well, all classes and interfaces are needed and it easily to be extended. Thus, users could be concentrated on the constructions of algorithms. For an estimation users can use the different sources of data, different anomalous types, to see, or algorithms are reliable or not. Also there is a plot that shows the results of anomalies real-time and all necessary indexes, for what simpler to influence the parameters of algorithm an user and effectively to compare to other by algorithms.

Key words: MULTIDIMENSIONAL, EXPOSURE of ANOMALIES of STREAM, PYISC, SVM, LOF, DDM, CUSUM, FCWM.