ВІДГУК

офіційного опонента – кандидата технічних наук,

Ізоніна Івана Вікторовича

на дисертаційну роботу **Панчука Богдана Олександровича**

на тему **«Виявлення мережевих атак алгоритмами штучного інтелекту»,**

подану на здобуття наукового ступеня доктора філософії

за спеціальністю 122 Комп'ютерні науки

1. Актуальність теми дисертації.

За останнє десятиліття стає дедалі помітніше стрімке збільшення кількості мережевих атак на інтернет ресурси, що завдає суттєвих збитків приватним особам, компаніям та державним установам. Цей явище пов'язано з появою нових типів шкідливих програм, здатних автономно функціонувати та самостійно розповсюджуватись у мережевому середовищі. Для протидії таким загрозам необхідно завчасно виявляти потенційні джерела атаки. Однак, протоколи функціонування даних типів шкідливих програм постійно змінюються та вдосконалюються і в експертів, зазвичай, немає доступу до їх вихідного коду, що суттєво ускладнює їх виявлення класичними методами на основі співставлення сигнатур мережевого трафіку. Методи штучного інтелекту дозволяють вивчати високорівневі шаблони поведінки програм на основі історичних даних про їх мережеву активність та виділяти неочевидні залежності, що дозволяє виявляти цілі класи споріднених зловмисних програм та завчасно запобігати їх розповсюдженню. Водночас, моделі виявлення на основі штучного інтелекту, зокрема на основі штучних нейронних мереж, можуть бути вразливими до спроб навмисного ухилення від виявлення через внесення збурень у мережеві дані. Це обумовлює необхідність вдосконалення методів підвищення стійкості систем виявлення до впливів такого роду. Не менш важливим є підвищення ступеню довіри до систем кібербезпеки, що можна досягнути шляхом розробки та застосування методів формальної верифікації властивостей стійкості моделей класифікації мережевих даних. Перераховані проблеми потребують подальшого дослідження, що свідчить про важливість та актуальність теми дисертаційної роботи Панчука Богдана Олександровича.

2. Ступінь обґрунтованості наукових результатів, положень, та висновків.

Наукові результати наведені в роботі здобувача викладені структуровано та послідовно. Всі розроблені методи реалізовано програмним шляхом. Основні положення, які виносяться на захист обґрунтовано коректним використанням математичного апарату, підкріплені успішною реалізацією, практичним впровадженням результатів дисертаційних досліджень, яке продемонструвало збігання теоретичних досліджень з реальними результатами.

Достовірність та обґрунтованість розроблених методів і програмних засобів підтверджується результатами експериментальних досліджень. Їх ефективність та дієвість емпірично продемонстрована через ряд експериментів із використанням відкритих наборів даних для навчання та тестування моделей штучного інтелекту. Це дозволяє відтворювати та проводити порівняння отриманих результатів із спорідненими роботами в цій області досліджень.

Наукові положення, висновки та рекомендації, сформульовані в дисертаційній роботі, є коректними, науково виваженими та можуть бути рекомендовані для використання при розробленні й впровадженні інтелектуальних систем виявлення мережевих атак, зокрема ботнетів, з підвищеною стійкістю до змагальних впливів.

3. Зв'язок роботи з науковими програмами, планами та темами.

Дисертаційна робота здобувача виконана у відповідності до плану наукових досліджень відділу № 100 Інституту кібернетики імені В.М. Глушкова НАН України у рамках науково-дослідницької теми «Розробити формальні методи виявлення зловмисної поведінки в мережі та хмарному оточенні на основі комбінації алгебраїчних методів та машинного навчання» (ВП.100.17, 2022-2024, за номером держреєстрації 0122U001164).

4. Наукова новизна результатів дослідження.

В роботі вперше розроблено метод проведення формальної верифікації властивостей нейронних мереж за допомогою SMT-розв'язувача, в основі якого покладено спрощене представлення обчислювального графу мереж з кусково-лінійними функціями активації. Автором вперше формалізовано критерій стійкості класифікаторів мережевого трафіку до потенційних збурень у вхідних даних та проведено автоматизовану верифікацію, що дозволило більш точно оцінити стійкість системи виявлення загроз до можливих змагальних впливів з боку зловмисника. Набуло подальшого розвитку використання методу генерації штучних прикладів та доповнення наборів даних шляхом адаптації швидкого методу знаку градієнту до простору ознак мережевих потоків, що дозволило оцінити та підвищити стійкість систем виявлення шкідливого трафіку до змагальних атак з урахуванням семантики можливих дій зловмисника.

5. Практичне значення отриманих результатів.

Отримані в дисертаційній роботі результати можуть стати основою для розроблення компонентів системи виявлення шкідливих програм шляхом аналізу мережевих даних у режимі реального часу із використанням моделей машинного навчання. Використання модифікації методу генерації змагальних прикладів мережевих потоків дозволяє оцінювати вразливість навчених моделей машинного навчання до змагальних впливів. Окрім цього, цей метод забезпечує можливість аугментації навчальної вибірки даних для підвищення

стійкості моделі машинного навчання до можливих спроб ухилення від виявлення. На основі отриманих результатів, створено прототип програми виявлення мережевих загроз, який впровадженого у компанії НВП "Радікс".

6. Повнота викладення наукових положень та висновків в опублікованих працях.

Основні результати дисертаційної роботи опубліковано в 4 статтях, які на момент публікації включено до переліку фахових видань України за спеціальністю 122. Два видання відносяться до фахових видань України категорії А. В публікаціях висвітлено матеріали подані в першій половині розділу 2, а також в розділах 3 та 4 дисертації. Також опубліковані праці містять частину оглядових матеріалів описаних у першому розділі роботи. Результати роботи апробовано на міжнародій науковій конференції, матеріали якої індексуються наукометричною базою Scopus. Таким чином, результати дослідження та наукові положення, що виносяться на захист, достатньою мірою відображені в опублікованих здобувачем наукових працях.

7. Оцінка змісту дисертації та її завершеності.

Повний обсяг дисертації становить 167 сторінок і включає 24 таблиці та 24 рисунки. Основний текст дисертації, викладений на 128 сторінках, у повній мірі розкриває тему дослідження, демонструє вирішення всіх поставлених завдань, і достатньо розгорнуто описує процес досягнення встановленої мети роботи.

У вступі здобувач аргументує актуальність теми дисертації, визначає мету, та формулює перелік дослідницьких завдань, необхідних для її досягнення. Тут також подано новизну та практичну цінність роботи.

Перший розділ роботи має оглядовий характер. В ньому праведно аналіз споріднених досліджень за темою класифікації мережевих даних методами машинного навчання з метою виявлення активності зловмисних програм, в особливості ботів та ботнетів. Здобувачем виділено найбільш значущі роботи, стисло представлена їх суть, особливості та внесок у розвиток застосування штучного інтелекту в області мережевої кібербезпеки.

В другому розділі роботи описано процес побудови системи виявлення шкідливого мережевого трафіку із використанням моделей класифікації на основі машинного навчання та штучних нейронних мереж. Зокрема, детально описаний метод агрегації мережевих пакетів у мережеві потоки та виділення характеристичних числових ознак, на базі яких виконується класифікація даних. Окрім цього, продемонстровано результати навчання та тестування різних архітектур класифікаторів на кількох відкритих наборах даних, зокрема розглянуто ICSX Botnet 2014, CTU-13, а також їх розширення прикладами трафіку новіших зразків вірусного ПЗ отриманих з Malware Capture Facility Project.

Третій розділ присвячено задачі підвищення стійкості систем виявлення мережевих загроз до змагальних атак. Описано адаптацію методу швидкого знаку градієнту для генерації штучних прикладів мережевих потоків, з метою оцінки стійкості моделі класифікації до змагальних впливів. Окрім цього, у розділі продемонстровано ефективність використання аугментації навчальної вибірки згенерованими змагальними даними і експериментально показано можливість підвищення стійкості класифікатора на основі штучної нейронної мережі до збурень у характеристиках розмірів корисного навантаження пакетів.

Четвертий розділ дисертації присвячено формальній верифікації властивостей стійкості нейронних мереж до можливих флуктуацій у вхідних даних. В цьому розділі формалізується критерій стійкості класифікатора та пропонується метод застосування SMT-розв'язувача для автоматичного доведення або спростування цього критерію для штучних нейронних мереж з кусково-лінійними функціями активації.

У висновках до роботи підсумовуються отримані науково-практичні результати у відповідності до завдань та мети дослідження.

Робота Панчука Б. О. задовольняє вимогам щодо оформлення дисертації затвердженого МОН України від 12.01.2017 № 40 "Про затвердження Вимог до оформлення дисертації" та відповідає стандарту та напряму дослідження відповідно до спеціальності та освітньої програми 122 Комп'ютерні науки.

8. Недоліки та зауваження до дисертаційної роботи.

1) У розділі 2 проводилось навчання класифікаторів мережевих даних на основі як нейронних мереж, так і методів машинного навчання, таких як випадковий ліс. Проте, у розділі 3 розглядаються виключно штучні нейронні мережі. Відповідно проблема змагальних атак розкрита лише для одного типу класифікаторів, із тих, які використовувались в практичній частині роботи. І хоча теоретично недиференційовні моделі класифікації є дійсно менш вразливими до змагальних атак, для яких необхідне обчислення градієнту, однак варто було би включити перевірку їх вразливості до трансферних атак, використовуючи вже навчені нейронні мережі у якості сурогатних моделей.

2) Схоже зауваження можна зробити і до розділу 4, де наводиться опис методу формальної верифікації лише для повнозв'язних нейронних мереж. Відповідно, для системи виявлення загроз описаної у роботі не надається формальних гарантій стійкості при використанні саме класифікаторів на основі методів машинного навчання, а також рекурентних та згорткових нейронних мереж, які згадуються у розділі 2.

3) В розділі 3 розглядаються лише змагальні атаки типу "white box", для виконання яких, зловмиснику має бути відома архітектура та всі параметри моделі класифікації, що є не зовсім реалістичним. Більш

практичним було б використання методів "чорної скриньки" (black box), де зловмисник не має доступу до моделі, однак може спостерігати за результатами їх функціонування на різних вхідних даних.

4) У Таблиці 2.3. варто було б навести показники хибнопозитивного рівня класифікаторів на основі рекурентних та згорткових нейромереж, оскільки в роботі наголошується важливість цієї метрики для задач кібербезпеки.

5) В Таблиці 4.1 надано результати формальної верифікації критерію стійкості класифікатора до збурень лише в статистиках міжпакетних інтервалів та тривалості мережевих потоків. Варто було б також навести результати для статистик розмірів пакетів, можливість збурення яких розглядалась у розділі 3.

Однак, зазначені зауваження не є принциповими, істотно не впливають на зміст дисертаційної роботи та не знижують її наукової та практичної цінності.

9. Висновок про відповідність дисертації вимогам.

Дисертаційна робота Панчука Б. О. "Виявлення мережевих атак алгоритмами штучного інтелекту" є завершеною науковою працею, яка містить нові науково обґрунтовані результати. Актуальність, наукова новизна, практична цінність, обсяг та якість подання матеріалу відповідають всім вимогам зазначеним у "Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії", затвердженим Постановою Кабінетом Міністрів України № 44 від 12 січня 2022 р.

Вважаю, що здобувач Панчук Богдан Олександрович заслуговує на присудження ступеня доктора філософії за спеціальністю 122 "Комп'ютерні науки".

Офіційний опонент:
кандидат технічних наук,
доцент кафедри систем штучного інтелекту,
Інституту комп'ютерних наук та інформаційних технологій,
Національного університету
"Львівська політехніка". _____ Іван ІЗОНІН

Підпис Ізоніна І.В. засвідчую:
Вчений секретар
Національного університету
"Львівська політехніка"
к. т. н., доцент _____ Роман БРИЛИНСЬКИЙ