ВІДГУК

офіційного опонента – доктора технічних наук,
Гордєєва Олександра Олександровича
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»

Актуальність теми дослідження.

Системи виявлення мережевих загроз традиційно відіграють ключову роль у забезпеченні кібербезпеки та залишаються невід'ємною складовою комплексного захисту інформаційних систем від вторгнень. Спостерігається постійне протистояння між зловмисниками, які вдосконалюють свої методи атак, та розробниками засобів захисту, що змушує останніх адаптувати й розвивати механізми протидії новим загрозам. У цьому протистоянні класичні підходи до виявлення загроз дедалі частіше виявляються недостатньо ефективними через зростаючу складність та різноманітність поведінки сучасних типів шкідливого мережевого програмного забезпечення, яке активно використовується зловмисниками для реалізації широкого спектра атак. Одним з найбільш багатообіцяючих напрямків для вирішення цієї проблеми є застосування алгоритмів, моделей та методів штучного інтелекту для аналізу даних, які передаються комп'ютерними мережами. Зокрема, класифікатори машинного навчання вже активно використовуються в комерційних системах виявлення загроз, в особливості, у хмарному оточенні. Водночас, спектр видів мережевої активності, які ці системи здатні виявляти, є дещо обмеженим, а деталі їх реалізації залишаються закритими. Відповідно, дана область потребує подальших академічних досліджень, і тому тема дисертаційної роботи здобувача сьогодні є актуальною.

Рівень обґрунтованості наукових результатів, положень та висновків.

Наукові положення представлені в рамках дисертаційної роботи мають достатньо високий рівень обґрунтування. Твердження зроблені здобувачем базуються на існуючій теоретичній базі цієї області та мають експериментальне та емпіричне підтвердження. В дисертації міститься детальний опис постановки експериментів, і ключова інформація та числові результати наводяться у формі 24 таблиць в основній частині роботи та 5 таблиць у додатках. Отримані експериментальні дані обраховані на основі навчальних та тестувальних наборів мережевого трафіку, які знаходяться у

відкритому доступі, і в роботі містяться посилання на їх джерела. Для вирішення задач дослідження, в роботі активно використовуються алгоритми машинного навчання, моделі нейронних мереж, методи обробки мережевих даних, та методи формальної верифікації. Висновки зроблені здобувачем відповідають отриманим результатам та демонструють досягнення поставленої мети роботи.

Наукова новизна та практична значущість результатів роботи.

В дисертаційній роботі запропоновано ряд нових методів та вдосконалення існуючих алгоритмів, направлені на покращення якості систем виявлення слідів мережевої активності шкідливого програмного забезпечення моделями штучного інтелекту:

1. Продемонстровано створення систем аналізу мережевих даних на основі машинного навчання та нейронних мереж, навчених на розширеній вибірці. Зокрема, запропоноване розширення дозволило підвищити показники повноти виявлення TCP-трафіку шкідливих програм з 46% до 80.7%.
2. Запропоновано новий варіант розвитку швидкого методу знаку градієнту для генерації штучних змагальних прикладів мережевих потоків супроти нейронних мереж. Це вдосконалення дозволило імітувати можливі збурення в даних, внесені з метою ухилення від виявлення. Після навчання нейромережі на вибірці, доповненій штучними прикладами, продемонстровано зниження показників вразливості моделі з 8.5% до 2% при маніпуляціях довжинами міжпакетних інтервалів в межах 30%.
3. Формалізовано критерій локальної стійкості нейромереж-класифікаторів мережевих потоків до змагальних впливів з боку зловмисника. Розроблено метод формальної верифікації даного критерія з використанням нового алгоритму перевірки можливості спрощення кусково-лінійних функцій активації окремих нейронів за допомогою SMT-розв'язувача.

Отримані в ході роботи напрацювання набули прикладного застосування у формі прототипу системи аналізу мережевих даних для виявлення загроз. Екземпляр прототипу впроваджено у використання в компанії ТОВ «НВП Радікс», яка займається розробкою апаратного забезпечення для АЕС, що демонструє практичну значимість результатів роботи.

Повнота викладення результатів дослідження в публікаціях.

Результати та наукові положення роботи здобувач виклав у 4 статтях опублікованих у наукових журналах. Всі журнали включені до переліку фахових видань України і безпосередньо відповідають спеціальності

здобувача 122 «Комп'ютерні науки», або ж є суміжними з нею. Також частина напрацювань опублікована в одноосібних матеріалах однієї міжнародної конференції. У всіх опублікованих працях висвітлені найбільш вагомі результати дослідження і їх зміст відповідає матеріалам представленим в основних розділах дисертації.

Наукові програмами та теми, в рамках яких проводилось дослідження.

Дослідження виконувалось відповідно до теми науково-дослідницьких робіт відділу № 100 Інституту кібернетики імені В.М. Глушкова НАН України **ВП.100.17** «Розробити формальні методи виявлення зловмисної поведінки в мережі та хмарному оточенні на основі комбінації алгебраїчних методів та машинного навчання» (2022-2024, **номер держреєстрації 0122U001164**).

Оцінка змісту дисертації та відповідності вимогам до оформлення.

Вступна частина роботи висвітлюється актуальність теми дослідження, з урахуванням стрімкого розповсюдження ботнетів та інших мережевих загроз. Здобувачем чітко сформульована мета дослідження та визначено ряд завдань необхідних для її досягнення. У вступі визначаються предмет, об'єкт та методи дослідження, а також практичне значення результатів роботи.

В першому розділі проведено аналіз та систематизацію результатів споріднених робіт інших авторів в області виявлення трафіку зловмисних мережевих програм моделями штучного інтелекту. Розглядаються різні підходи до виділення характеристичних ознак з мережевих даних, зокрема способи агрегації та групування мережевих пакетів для можливості подальшого аналізу. Також зауважено основні невирішені проблеми в даній області: зокрема, гостра нестача та швидке застарівання існуючих наборів навчальних даних, та відсутність гарантій стійкості систем безпеки до можливих спроб ухилення від виявлення.

Другий розділ демонструє всі ключові етапи побудови системи виявлення шкідливого трафіку на основі класифікаторів нейронних мереж різних архітектур та машинного навчання. Описується реалізація алгоритму виділення мережевих потоків з перехоплених мережевих пакетів, та їх подальше використання, як окремих одиниць класифікації. Особливу увагу присвячено процесу обробки відкритих наборів трафіку ботнетів, а також навчанню та порівнюю показників ефективності різних моделей класифікації. Зокрема оцінювалась повнота виявлення та хибнопозитивний рівень. Також в розділі описано процес створення нового набору даних на основі вже

відомого ICSX Botnet 2014, шляхом його розширення зразками трафіку більш нових вірусних програм, таких як WannaCry, TrickBot та Emotet. Експериментально показана значна ефективність моделей на основі випадкових лісів та штучних нейронних мереж прямого розповсюдження.

В третьому розділі розглядається проблема змагальних атак зловмисників на нейромережі-класифікатори з метою приховування шкідливого трафіку від систем виявлення. Зокрема, розглядаються можливості маніпуляцій розмірами мережевих пакетів та інтервалами між послідовними пакетами у потоці. Здобувачем запропоновано розвиток швидкого методу знаку градієнту для генерації змагальних прикладів з урахуванням семантики простору ознак мережевих потоків. Також здійснено доповнення навчальної вибірки для підвищення стійкості нейромереж до змагальних атак.

В четвертому розділі роботи запропоновано метод для формальної верифікації властивостей повнозв'язних нейромереж. Метою є перевірка критерію стійкості класифікатора до можливих збурень в числових ознаках мережевих потоків. Цей критерій задається шляхом накладанням обмежень на вхідні параметри та значення вихідних нейронів. В розділі описано алгоритм верифікації виконуваності даних обмежень за допомогою SMT-розв'язувача Microsoft Z3.

Основний текст дослідження викладено на 128 сторінках, а повний обсяг роботи складає 167 сторінок. Розділи роботи містять 24 таблиці та 24 рисунки, які наглядно демонструють отримані здобувачем результати.

Робота Панчука Б.О. є завершеною науковою працею, а її оформлення задовольняє вимогам МОН України від 12.01.2017 № 40 «Про затвердження Вимог до оформлення дисертації». Тема дослідження розкрита у повній мірі і зміст роботи відповідає спеціальності 122 «Комп'ютерні науки».

Недоліки дисертації.

Водночас, у роботі присутні деякі неточності та недоліки, які варто зазначити:

1. У пункті 2.10.4 при описі розширення набору мережевих даних не вказано, які саме частини Malware Capture Facility Project використовувались. Необхідно було б дати посилання на конкретні складові цього проекту, які були включені у навчальну та тестову вибірки. Без цієї інформації може бути складно відтворити деякі отримані результати.
2. Не пояснено як саме обиралася оптимальна кількість рівнів, а також кількість нейронів на рівень для класифікаторів нейронних мереж.

3. В експериментах розділів 2 та 3 використовуються два різних інструменти виділення мережевих потоків зі схожим функціоналом: CICFlowMeter та NetWatcher, розроблений здобувачем. Слід було б уніфікувати підхід для отримання більш консистентних результатів.
4. При обчисленні характеристичних ознак мережевих потоків використовувати лише дані заголовків мережевого та транспортного рівнів моделі OSI. Було б доцільно також використати інформацію канального та прикладного рівнів.
5. Варто було б більше детально розглянути вже існуючі комерційні рішеннями, такі як Cisco Secure Network Analytics та Cisco Umbrella, і провести порівняння показників ефективності розробленої системи з альтернативами.

Проте, наявність зазначених недоопрацювань та недоліків не надто суттєво впливає на загальну наукову цінність та практичну значимість результатів дослідження.

Висновок.

Робота Панчука Б. О. «Виявлення мережевих загроз штучного інтелекту» демонструє достатній рівень кваліфікації автора та володіння навичками, необхідними для отримання ступеня доктора філософії. Аналіз рукопису дозволяє зробити висновок, що робота підготована здобувачем самостійно і не містить слідів порушення академічної доброчесності або фальсифікації результатів. Дисертація задовольняє 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», визначеного постановою Кабінету Міністрів України №44 від 12 січня 2022 року.

На підставі сказаного вище, вважаю, що автор дисертації Панчук Богдан Олександрович заслуговує на присудження йому ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

Офіційний опонент:
доктор технічних наук,
професор кафедри інженерії
програмного забезпечення
Луцького національного технічного
університету Олександр ГОРДЄЄВ

ПІДПИС ЗАСВІДЧУЮ:
Учений секретар
ЛУЦЬКОГО НАЦІОНАЛЬНОГО
ТЕХНІЧНОГО УНІВЕРСИТЕТУ
професор Алла ЗЕМКО