РЕЦЕНЗІЯ

офіційного рецензента – доктора фізико-математичних наук, професора
Горбачука Василя Михайловича
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

Методи та алгоритми штучного інтелекту (ШІ) вже проникли у всі прикладні сфери – від соцмереж та розваг, до медицини, економіки та бізнесу – забезпечуючи новий рівень автоматизації, гнучкості та зручності людино-машинної взаємодії. У сфері кібербезпеки використання ШІ стало не лише перспективним, а й необхідним. На сьогодні інфраструктура більшості комерційних інформаційних систем вже перенесена в хмарне середовище, з чим пов'язується широке застосування розподілених обчислень на основі інтеграції багатьох окремих інтернет-застосунків. Мережева орієнтованість таких систем відкриває безліч нових можливих векторів атаки. Одну з найбільш суттєвих загроз становлять шкідливі програмні комплекси з можливістю автономної роботи, віддаленого контролю та самостійного розповсюдження, також відомі як ботнети. Ботнети використовуються для крадіжки даних, розсилання спаму та є першоджерелами більшості DDoS-атак. Саме тому їх завчасне виявлення є першочерговим завданням кібербезпеки. Однак такі програми швидко вдосконалюються. Зловмисники використовують шифрування та спотворення даних, випадкову генерацію і швидку ротацію доменних імен та інші підходи для приховування своєї активності. У зв'язку з цим стає дедалі складніше розробляти традиційні системи виявлення, які зазвичай базуються на вивченні евристик поведінки й алгоритмічному співставленні шаблонів мережевих даних. Натомість моделі ШІ здатні самостійно навчатися та можуть виявляти складніші приховані закономірності. Це дозволяє протидіяти новим типам загроз, які мають спорідненість з прикладами, відомими з минулого.

Зв'язок дослідження з науковими програмами та темами

Тема дисертації та результати дослідження відповідають напрямку науково-дослідних робіт 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 рр., номер держреєстрації 0122U001164);

ВП.100.16 «Розробити формальні методи виявлення вразливостей програмних систем».

Рівень обґрунтованості наукових результатів

Наукові результати, представлені в дисертації, мають під собою надійне теоретичне й експериментальне підґрунтя. Використані у роботі підходи до створення та покращення систем виявлення мережевих загроз основані як на власних напрацюваннях автора, так і на теоретичних засадах попередніх наукових досліджень. Створені алгоритми представлені в роботі у вигляді псевдокодів, а для ключових обчислень наведені формули з відповідним математичним обґрунтуванням їх інтерпретації та коректності.

В ході дослідження здобувач використовував розроблену ним систему аналізу та класифікації мережевих даних для експериментального підтвердження працездатності та ефективності запропонованих алгоритмів і методів. Графіки та таблиці показників повноти виявлення дозволяють наочно простежити переваги запропонованих підходів на широкому спектрі даних. Таким чином, робота демонструє достатній рівень обґрунтованості наведених результатів та наукових положень.

Огляд змісту дисертації та відповідності вимогам до оформлення

Робота відповідає стандартним вимогам оформлення дисертації доктора філософії та містить вступну частину, основні розділи, висновки та додатки. В першому розділі автор проводить огляд найважливіших досліджень, споріднених за тематикою, – аналізу мережевих даних алгоритмами та моделями ШІ. Найбільш фундаментальні роботи представлені у хронологічному порядку, і читачу нескладно розуміти послідовність ключових етапів розвитку цієї області. Другий розділ роботи в деталях демонструє найважливіші аспекти побудови системи виявлення мережевої активності ботнетів та інших вірусних програм, а саме: виділення потоків трафіку, обчислення їх характеристичних атрибутів, обробка відкритих наборів мережевих пакетів, навчання та тестування моделей класифікації на отриманій базі. Третій розділ досліджує важливу проблему можливих змагальних впливів, спрямованих проти нейромереж-класифікаторів мережевих даних. Розділ присвячено вирішенню проблеми оцінки стійкості системи до спроб навмисного ухилення від виявлення шляхом внесення збурень у трафіку, а також підвищення стійкості нейромереж до можливих маніпуляцій такого роду. У четвертому розділі роботи досліджено можливості надання формальних гарантій коректності роботи класифікаторів на базі нейронних мереж. Основна увага зосереджена на проблемі верифікації критерію локальної стійкості класифікатора до збурень в мережевих даних. В розділі продемонстровано процес накладання

формальних обмежень на входи та виходи моделі з подальшим доведенням чи спростуванням їх виконуваності розв’язувачем для символьних логік Microsoft Z3.

Практична значимість роботи

Для демонстрації практичної значущості результатів роботи здобувачем самостійно розроблено прототип системи безпеки, здатний перехоплювати й аналізувати мережеві дані в режимі реального часу, а також виявляти сліди поведінки широкого спектру шкідливих програм. Тестування створеної системи проведено на досить репрезентативній вибірці, яка включала приклади активності 14 класів різних ботнетів і вірусів, а також приклади звичайного трафіку HTTP та P2P (peer-to-peer). Зокрема, при використанні штучної нейронної мережі у ролі класифікатора, система показала 81.6% точності виявлення, при цьому лише 0.14% результатів виявилися хибнопозитивними. Продемонстровано, що система має високу чутливість (> 75%) до трафіку таких ботнетів, як DonBot, RBot, TrickBot та Virut, а також штамів Emotet. Додатково прототип зміг виявити активність вірусу WannaCry, трояна Murlo та кількох інших типів шкідливого ПЗ з повнотою 30–60% на тестовій вибірці. Наведені результати свідчать про можливість практичного застосування створеної системи захисту, а її прототип отримав експериментальне впровадження у компанії ТОВ «НВП «Радікс»».

Окремо слід зазначити можливість ширшого прикладного застосування алгоритму формальної верифікації нейронних мереж, запропонованого в останньому розділі роботи. Універсальність алгоритму дозволяє використовувати його не лише в сфері кібербезпеки, але й в інших областях, де необхідно доводити математичні властивості повнозв’язних нейромереж з кусково-лінійними функціями активації.

Наукова новизна результатів дисертації

Крім практичної цінності, алгоритми та методи, запропоновані в дисертації, мають достатній рівень наукової новизни.

Автором запропоновано новий розвиток методу генерації штучних прикладів змагальних даних для оцінки стійкості моделей класифікації до спроб ухилення від виявлення, шляхом внесення обмежених збурень у розміри мережевих пакетів та міжпакетних інтервалів. Алгоритм знаходження необхідних збурень засновано на «швидкому методі знаку градієнту» (FGSM), який було вдосконалено з врахуванням природи характеристичних ознак мережевих даних і можливостей зловмисника впливати на них. Продемонстровано використання розробленого алгоритму для нарощування навчальної вибірки штучно-згенерованими прикладами, з

метою підвищення показників стійкості системи до можливих спроб ухилення від виявлення.

Іншим внесок здобувача є формалізація критеріїв стійкості класифікаторів мережевих потоків до флуктуацій та збурень у семантично-пов'язаних групах характеристичних ознак. Продемонстровано алгоритм символьного представлення нейронних мереж структурними елементами Microsoft Z3 та використання SMT-розв'язувача для формальної верифікації побудованих моделей.

Також суттєвим внеском роботи є створення нового методу спрощення символьного представлення графу обчислень нейронних мереж. Метод оснований на оптимізації складності моделі шляхом автоматичного доведення можливості вилучення нелінійностей з функцій активацій нейронів. Такий підхід вирішив проблему комбінаторного вибуху, спричинену кількістю розгалужень у SMT-формулах, і дозволив значно прискорити швидкість верифікації локальної стійкості нейромережі.

Повнота викладення результатів дослідження в публікаціях

Напрацювання, отримані в межах дисертаційного дослідження, знайшли відображення у 4 статтях, опублікованих у наукових виданнях, що входять до переліку фахових журналів України, 2 з яких належать до категорії «А». Частина результатів була також представлена на міжнародній науковій конференції та опублікована у вигляді матеріалів в CEUR. Опубліковані праці охоплюють ключові положення дисертації, відображають зміст її основних розділів та свідчать про належний рівень апробації проведеного дослідження.

Зауваженнята рекомендації

Водночас, у робота містить певні недоліки, на які варто звернути увагу:

1. Деякі з типів ботнетів, присутні в навчальних та тестових наборах, вже суттєво застаріли. Наприклад, основна активність DonBot припадала на 2009−2012 рр. Варто було б доповнити дослідження аналізом більш сучасних типів ботнетів, зокрема тих, які націлені на пристрої Інтернету речей, наприклад, Mirai та його похідних. Експериментальна база застаріла: основні набори даних (ISCX 2014, CTU-13) застарілі на понад 10 років; найновіші зразки malware датуються 2017−2018 рр.; відсутні сучасні IoT-ботнети, cryptominers, APT-related malware; не враховуються сучасні evasion techniques.
2. Експерименти у роботі засновані лише на аналізі наборів історичних мережевих даних. Було б добре розширити дослідження, провівши живе тестуванняв лабораторних умовах з залученням реальних екземплярів виконуваних файлів шкідливих програм та перехоплення трафіку,

породженого ними. Мережевий контекст обмежений: всі дані зібрані в лабораторних умовах; відсутній enterprise traffic з NAT, proxies, load balancers; не оцінено вплив мережевої інфраструктури на feature extraction.

3. В розділі 2 використання відносно простих архітектур моделей класифікації аргументовано вимогою до високої швидкості висновування, що необхідно для роботи системи у режимі реального часу. Однак, в тексті не вимірюється та не порівнюється швидкодія навчених моделей класифікації. Обмеження SMT-верифікації: працює лише з кусково-лінійними функціями активації; не масштабується на великі нейромережі (експоненційна складність); відсутня оцінка часової складності алгоритму верифікації. Неточності у математичній нотації: у розділі 4 потребується уточнення формулювань SMT-задач; недостатньо чітке визначення локальної стійкості для мережевого трафіку; потребується формалізація constraint satisfaction problem. Відсутній аналіз складності: не проаналізована часова та просторова складність алгоритмів; відсутній аналіз масштабованості SMT-верифікації; потребується оцінка memory footprint для обробки real-time.

4. Подібно до попереднього зауваження, у роботі не наведено оцінки пропускної здатності програмної реалізації реконструктора мережевих потоків. Попри те, що система позиціонується як прототип, було б доцільним оцінити кількість пакетів чи байтів, яку поточна реалізація могла б обробити за секунду. Performance-аналіз відсутній: не наведено метрики затримки для виявлення real-time; відсутні дані про memory footprint та CPU utilization; не оцінено throughput системи під навантаженням.

5. В розділі 3.5 використовується порогове значення класифікатора 0.51, однак відсутнє пояснення чому саме ця цифра була обрана. Слід було б обирати порогове значення, виходячи з міркувань, наведених у розділі 2, щодо збереження низького показника хибнопозитивного рівня. Тестові умови є нереалістичними: тестувальна вибірка містить 92.54% шкідливого трафіку vs реальні <1%, що робить метрики AUROC та точність нерепрезентативними; необхідна оцінка на realistic class distribution. Оцінювальні метрики обмежені: використання лише стандартних метрик без domain-specific; відсутня оцінка detection latency для сценаріїв real-time; потребується аналіз computational overhead та degradation під навантаженням. Статистична недостовірність результатів: експерименти на фіксованих train/test splits без cross-validation; відсутні довірчі інтервали та тести статистичної значущості; не враховується variance між запусками нейромереж; недостатньо описано hyperparameter tuning процес

6. У формулі 3.1 є неузгодженість між векторами та скалярами. Слід було зазначити, що це обчислення проводиться для кожної окремої i-тої компоненти вектору x, та замінити позначку градієнту на частинну похідну за відповідним x_i.

7. Робота не згадує актуальної динамічної галузі ботнетів для БПЛА.
8. На с. 19−20 у підрозділі «Зв’язок роботи з науковими програмами, планами, темами» не вказано «Теоретичні та експериментальні дослідження з розробки кейс-орієнтованих методів оцінювання, інтерфейсів та технології верифікації ПСКЗ» (0123U102106).
9. Критичні обмеження системи NetWatcher на GitHub, які суттєво обмежують практичну застосовність системи: аналіз лише TCP-трафіку, виключення UDP та ICMP; сучасні ботнети активно використовують UDP (DNS tunneling, P2P).
10. Основна частина (розділи 2, 3, 4) займають 94 сторінки (с. 47−140 з рисунками і таблицями) із загальних 167 сторінок дисертації, тобто менше 57%.
11. Після заголовків 1.2.2, 2.4.2, 2.10.3, 4.6 текст починається на наступній сторінці, а не на сторінці заголовку. Заголовок рис. 2.15 розташований на наступній сторінці, а не на сторінці рисунку.
12. Запозичені (англомовні) рис. 2.1 (с. 53), 2.2, 2.3 (с. 55), рис. 2.4 (с. 56), рис. 2.5 (с. 62), рис. 2.6 (с. 69, 73), рис. 2.7 (с. 76) без посилань.
13. Якщо рис. 2.6 розміщено на с. 69, то наступний рис. на с. 73 бути 2.7, а не 2.6.
14. На с. 10, 158 дисертації невірно вказана назва фахового періодичного видання, де опублікована стаття дисертанта: замість «Проблеми кібербезпеки та інформаційних технологій» (такого видання не існує) має бути «Проблеми керування та інформатики». Цю назву слід також виправити для ЄДЕБО.
15. Недостатнє обґрунтування вибору методу генерації змагальних прикладів: у розділі 3.2 автор обирає FGSM без порівняльного аналізу з PGD, C&W, DeepFool; відсутнє обґрунтування придатності FGSM для специфіки мережевих потоків; потребується експериментальне порівняння ефективності різних методів генерації; не показано, чи враховуються семантичні обмеження мережевих протоколів при генерації. Адаптація FGSM обмежена: заявлена адаптація потребує глибшого теоретичного обґрунтування; відсутній аналіз реалістичності згенерованих змагальних прикладів; потребується формалізація constraints для мережевого домену.
16. Проблема contamination експериментальних даних: у наборі ISCX-CTU-Extended частина даних з CTU-13 потрапляє у навчальну та тестувальну вибірки, що призводить до data leakage та завищених показників якості; потребується чітке розділення джерел даних для train/test split.
17. Baseline-порівняння обмежені: відсутнє порівняння з сучасними NIDS (Snort ML, Suricata, Zeek); не включено state-of-the-art academic methods (2020−2024); базові моделі (логістична регресія, RF) застарілі для сучасних стандартів; потребується порівняння з ensemble-методами та transformer-

архітектурами. Недостатнє позиціонування відносно існуючих рішень: відсутнє детальне порівняння з Cisco Secure Network Analytics, Darktrace, ExtraHop; потребується аналіз переваг запропонованого підходу над commercial solutions; обмежений огляд recent academic works (2020−2024).

Висновок.

Дисертаційна робота Панчука Б. О. «Виявлення мережевих атак алгоритмами штучного інтелекту» свідчить про високий рівень підготовки здобувача, його здатність самостійно проводити наукові дослідження та досягати результатів, що мають наукову новизну і практичну значущість. Структура, зміст і оформлення дисертації відповідають вимогам, встановленим у «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженому постановою Кабінету Міністрів України № 44 від 12 січня 2022 року.

Хоча робота містить деякі недоліки, однак проведене дослідження є завершеним, а отримані наукові результати мають прикладне значення та можуть бути використані для створення та вдосконалення систем виявлення мережевих загроз. Вважаю доцільним присудження Панчуку Богдану Олександровичу наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

Офіційний рецензент
доктор фізико-математичних наук, професор,
завідувач відділу інтелектуальних інформаційних технологій
Інституту кібернетики НАН України Василь ГОРБАЧУК

Підпис В. Горбачук
ЗАСВІДЧУЮ
Зав. канц.
02.06.2025