МЕХАНІЗМИ ВИЯВЛЕННЯ ФІШИНГОВИХ РЕСУРСІВ У КОМП'ЮТЕРНИХ МЕРЕЖАХ НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Рязанов Є.О., Балагура Д.С.
Харківський національний університет радіоелектроніки, Харків, Україна

Фішингові атаки залишаються однією з найбільш деструктивних загроз у кіберпросторі через постійну еволюцію методів обходу статичних фільтрів. Сучасні зловмисники дедалі частіше використовують короткоживучі домени, динамічну генерацію URL-адрес та обфускацію контенту, що робить використання традиційних «чорних списків» недостатнім для забезпечення прийнятного рівня безпеки. Крім того, значну проблему становить швидкість поширення нових фішингових кампаній, коли час між створенням шкідливого ресурсу та його виявленням може становити лише кілька годин. Це зумовлює необхідність впровадження інтелектуальних систем, здатних виявляти аномалії в структурі вебресурсів у реальному часі, незалежно від наявності ресурсу у базах відомих загроз [1]. У цьому контексті особливого значення набувають підходи, що базуються на аналізі ознак та поведінкових характеристик, а не лише на сигнатурному пошуку.

Метою доповіді є обґрунтування структури та розробка концепції та механізмів детекції фішингових посилань на основі вилучення лексичних ознак URL-адрес та їх подальшої класифікації за допомогою методів машинного навчання[2]. Зазначений підхід має збільшити ефективність захисту від фішингових атак.

В основу підходу, що пропонується, покладено багатофакторний аналіз характеристик підозрілих ресурсів. Процес виявлення загрози поділяється на кілька етапів. На першому етапі здійснюється лексичний аналіз URL-адреси, що включає перевірку певних ознак, як: загальна довжина посилання, кількість субдоменів, наявність символів «@» та дефісів, а також використання IP-адрес замість доменних імен [3]. Для протидії атакам типу тайпосквотинг передбачається використання метрики відстані Левенштейна відносно переліку довірених брендів, що дозволяє ідентифікувати візуально схожі, але шкідливі домени. Математичне ядро може бути реалізуване із застосуванням ансамблевих методів, зокрема алгоритму випадкового лісу (Random Forest), що дозволить забезпечити стійкість до перенавчання та високу точність класифікації на зашумлених мережевих даних. Перевагою такого підходу є також інтерпретованість результатів, зокрема можливість оцінки важливості окремих ознак, що може бути використано для подальшої оптимізації моделі. Крім того, ансамблеві методи добре масштабуються та можуть бути адаптовані до змін у характері загроз шляхом періодичного донавчання моделі на нових даних.

З точки зору реалізації, зазначений підхід може бути запропонований у вигляді браузерного розширення, що функціонує як проміжний вузол контролю. Архітектура рішення включає аналітичний сервер (на базі

Flask/FastAPI), який прийматиме сформований вектор ознак, опрацьовуватиме його та повертатиме ймовірність належності ресурсу до шкідливих. Для мінімізації затримок можливе використання кешування результатів перевірок та асинхронної обробки запитів. У разі перевищення встановленого порогу ризику, розширення буде блокувати доступ до сторінки та виводити попередження для користувача. Такий підхід дозволить реалізувати проактивний захист, по суті не даючи зловмисним механізмам бути запущеними у системі, що атакується. Разом з тим, використання аналітичного серверу дозволить централізувати базу знань, тобто отримувати нові дані для навчання та, одночасно, знизити навантаження на обчислювальні ресурси клієнтської системи.

Важливим етапом розробки системи є оцінювання ефективності її функціонування. Зважаючи на неможливість проводити тестування у реальних умовах, основними механізмами тестування будуть синтетичні тести з використанням відкритих наборів даних (UCI Machine Learning Repository та PhishTank), а також крос-валідація для оцінки узагальнюючої здатності моделі. Для більш повної оцінки якості класифікації доцільно використовувати такі метрики, як precision, recall, F1-score та ROC-AUC, що дозволяють врахувати баланс між виявленням загроз і рівнем хибних спрацювань. Очікується, що використання ансамблевих моделей машинного навчання дозволить досягти високих показників точності при мінімальному рівні помилкових спрацювань (False Positives), що є критично важливим для збереження зручності користування системою з точки зору доступності доменів, що не несуть загроз.

Подальші дослідження можуть бути спрямовані на інтеграцію аналізу контенту сторінок за допомогою методів обробки природної мови (NLP). Також перспективним напрямком є використання гібридних моделей, які поєднують лексичний аналіз URL з поведінковим аналізом вебсторінок та характеристиками мережевої взаємодії, що дозволить суттєво підвищити рівень виявлення складних багаторівневих атак [4].

Список літератури

1. Jain A. K., Gupta B. B. A novel approach to detect phishing websites using self-structuring neural network. *EURASIP Journal on Information Security.* 2020. Vol. 2020, No. 4. P. 1–11.

2. Проблема недостатньої адаптивності традиційних систем мережевої безпеки до сучасних кіберзагроз: аналіз та шляхи вирішення / М. Надточий, Д. Балагура, П. Шулік, В. Просолов // *«Вісник Хмельницького національного університету.»* Том 359 № 6.1/2025, с. 63-69. https://doi.org/10.31891/2307-5732-2025-359-7

3. Verma R., Das A. What's in a URL? Fast Feature Extraction and Malicious URL Detection. *Proceedings of the 3rd ACM on International Workshop on Security And Privacy Analytics (IWSPA '17).* 2017. P. 55–63.

4. Дорофєєва, К., Сєвєрінов, О., Сидоренко, З., Сухотеплий, В. (2025). Застосування інструмента аналізу безпеки для виявлення критичних вразливостей у веб-додатках. Вісник Херсонського національного технічного університету, 3(4 (95)), 62-68.