УДК 004.056

РЕЗЕДЕНТ О.О.

КОНДРАТЕНКО Н.Р.

ЗАСІБ ВИЯВЛЕННЯ ФІШИНГОВИХ АТАК URL З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖЕВОГО ПІДХОДУ

Анотація

У роботі розглянуто проблему виявлення фішингових атак в інформаційних системах на основі аналізу URL-адрес. Запропоновано підхід, що базується на використанні методів машинного навчання, зокрема методу опорних векторів. Визначено основні ознаки фішингових URL, що дозволяють здійснювати їх класифікацію. Проведено аналіз ефективності моделі за метриками точності, прецизійності та чутливості, що підтверджує доцільність використання інтелектуальних підходів у задачах кібербезпеки.

Ключові слова: фішинг, URL-атаки, кібербезпека, машинне навчання, SVM, класифікація.

Abstract

The paper considers the problem of detecting phishing attacks in information systems based on URL analysis. An approach based on machine learning methods, in particular Support Vector Machine, is proposed. The main features of phishing URLs are identified, which allow their classification. The effectiveness of the model is evaluated using accuracy, precision, and recall metrics, confirming the feasibility of intelligent approaches in cybersecurity tasks.

Keywords: phishing, URL attacks, cybersecurity, machine learning, SVM, classification.

Вступ

У сучасному інформаційному середовищі спостерігається зростання кількості кіберзагроз, серед яких фішингові атаки є одними з найбільш поширених. Вони спрямовані на отримання конфіденційної інформації користувачів шляхом використання підроблених веб-ресурсів та маніпуляцій із URL-адресами [1]. Традиційні методи виявлення фішингових ресурсів, зокрема використання чорних списків, не забезпечують достатньої ефективності через динамічний характер атак та постійне оновлення шкідливих ресурсів [1]. У зв’язку з цим актуальним є застосування методів машинного навчання для аналізу ознак URL та виявлення фішингових сайтів [2]. Сучасні підходи, зокрема глибоке навчання, дозволяють автоматично виявляти складні закономірності у великих обсягах даних, що підвищує ефективність систем виявлення кіберзагроз [3].

Результати дослідження

У роботі розглянуто підхід до виявлення фішингових URL-атак, який базується на аналізі ознак веб-адрес та застосуванні методів машинного навчання. Запропонована модель включає етапи збору даних, вилучення ознак, їх обробки та подальшої класифікації веб-ресурсів на легітимні та фішингові [1].

У межах запропонованого підходу доцільно використовувати комбіновану систему оцінки, яка базується на методах косинусної подібності, коефіцієнті Жаккара та відстані Левенштейна. Кожен із цих методів розглядає текст URL як масив, векторне представлення або множину елементів, що дозволяє визначити ступінь близькості підозрілого посилання до легітимного шаблону. Зокрема, алгоритм Левенштейна розраховує мінімальну кількість операцій редагування для перетворення одного рядка в інший, що є критично важливим для виявлення візуально схожих доменів, які часто використовуються зловмисниками. Загальна ефективність виявлення визначається як зважена сума результатів цих методів, де найбільшу вагу має показник Левенштейна (0,4), що забезпечує стабільність результатів незалежно від глибини вкладеності URL.

Додатково модель враховує специфічні аномалії структури посилань, такі як наявність IP-адреси замість домену, використання символу подвійного слешу після протоколу для прихованого перенаправлення, а також маніпулятивне вставлення терміна «https» безпосередньо в доменне ім'я. Експериментальна перевірка на базі методу опорних векторів підтвердила перевагу лінійного ядра (SVM-linear), яке забезпечує точність на рівні 94,2% та чутливість 95%. Високі показники Accuracy та

Recall обумовлені лінійним характером залежностей між обраними ознаками, що дозволяє алгоритму ефективно будувати роздільну гіперплощину в багатовимірному просторі даних.

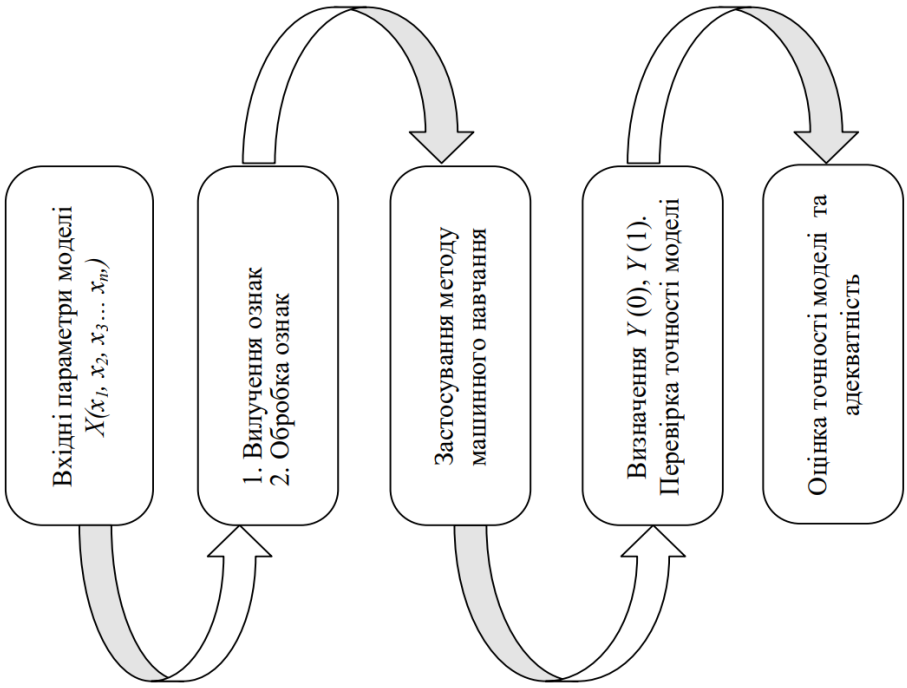

Рисунок 1 – Концептуальна модель виявлення фішингових атак

Для математичного обґрунтування оцінки підозрілості URL-адрес у межах запропонованої моделі використовується комбінований підхід, що базується на методах косинусної подібності, коефіцієнті Жаккара та відстані Левенштейна. Кожен із цих методів дозволяє розглядати текстову інформацію як масив, векторне представлення або множину елементів. Зокрема, відстань Левенштейна (алгоритм Дамерау-Левенштейна) визначає мінімальну кількість операцій редагування для перетворення одного рядка в інший, що є критично важливим для виявлення візуально схожих фішингових доменів.

Основними ознаками URL, що використовуються для виявлення фішингових ресурсів, є: довжина URL-адреси, наявність IP-адреси замість доменного імені, кількість піддоменів, використання спеціальних символів, а також характеристики безпеки з'єднання [1]. Аналіз цих параметрів дозволяє формувати вектор ознак для подальшої класифікації.

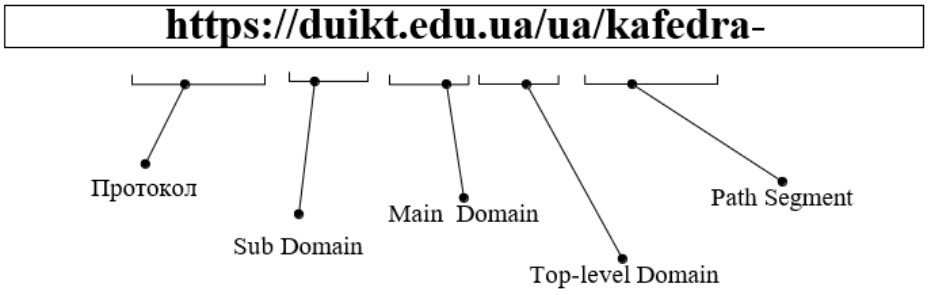

Рисунок 2 – Складові URL адреси

Для класифікації URL-адрес у роботі використано метод опорних векторів (SVM), який дозволяє ефективно розділяти дані у багатовимірному просторі ознак [1]. Основна ідея методу полягає у побудові гіперплощини, що максимально розділяє класи об'єктів. Крім класичних методів машинного навчання, у роботі розглянуто можливість застосування глибокого навчання для автоматичного вилучення ознак та підвищення точності класифікації. Нейронні мережі здатні самостійно виявляти складні залежності в даних та працювати з великими обсягами інформації, що робить їх ефективними для задач кібербезпеки [3].Альтернативні підходи до виявлення фішингових атак включають

використання інших алгоритмів машинного навчання, таких як логістична регресія, дерева рішень та ансамблеві методи, які дозволяють підвищити стійкість системи до нових типів атак [2]. Оцінювання ефективності запропонованого підходу здійснюється за допомогою стандартних метрик якості класифікації, зокрема точності (Accuracy), прецизійності (Precision) та повноти (Recall) [1]. Результати дослідження свідчать про високу ефективність використання інтелектуальних методів для виявлення фішингових URL-атак.

Після аналізу структури URL (рис. 2) необхідно деталізувати параметри класифікації посилань. Важливим етапом є ідентифікація аномалій: використання IP-адреси, наявність подвійного слешу після восьмої позиції для перенаправлення, маніпулятивне вставлення «https» у домен та застосування сервісів скорочення посилань для маскування цільового ресурсу.

Для аналізу цих ознак застосовано комбінований підхід на основі близькості текстових рядків. Відстань Левенштейна допомагає виявити візуально схожі домени через підрахунок операцій редагування. Косинусна подібність аналізує семантичний зв'язок через векторне представлення ASCII-кодів, а метод Жаккара оцінює схожість множин унікальних символів, що ефективно для великих обсягів даних.

Загальна ефективність виявлення фішингової атаки в межах запропонованої моделі розраховується як зважена сума результатів кожного з методів. Це дозволяє врахувати переваги кожного алгоритму та призначити їм вагу відповідно до їхньої релевантності в конкретному контексті. Математично Розрахунок загальної ефективності описується формулою [2]:

$$E_{general} = W_1 * E_1 + W_2 * E_2 + W_3 * E_3 \quad (1)$$

де $E_{general}$ – показник загальної ефективності виявлення аномалії; E_1, E_2, E_3 — значення ефективності, отримані методами відстані Левенштейна, косинусної подібності та коефіцієнта Жаккара відповідно; W_1 , W_2 W_3 – вагові коефіцієнти впливу кожного методу на кінцевий результат класифікації.

Висновки

У ході роботи розглянуто підхід до виявлення фішингових URL-адрес на основі аналізу їх ознак та використання методів машинного навчання. Визначено основні характеристики URL, які дозволяють відрізнити фішингові ресурси від легітимних. Для класифікації було використано метод опорних векторів, який показує хорошу ефективність при розділенні даних. Також встановлено, що застосування сучасних підходів, зокрема глибокого навчання, може підвищити точність виявлення фішингових атак. Отримані результати можуть бути використані для створення систем захисту інформації та покращення безпеки веб-ресурсів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Гайдур Г. І., Гахов С. О., Марченко В. В. Концептуальна модель виявлення фішингових атак на основі методів опорних векторів // Сучасний захист інформації. – 2024.
2. Toliupa S., Buchyk S., Shabanova A., Buchyk O. The Method for Determining the Degree of Suspiciousness of a Phishing URL // Taras Shevchenko National University of Kyiv. – Kyiv, Ukraine.
3. Глибоке навчання (Deep Learning): основи, принципи та застосування [Електронний ресурс]. – Режим доступу: https://foxminded.ua/deep-learning/

Резедент Олександр Олегович – студент групи 1БС-22б факультет інформаційних технологій та комп`ютерної інженерії, Вінницький національний технічний університет, м. Вінниця.

Кондратенко Наталія Романівна - к.т.н професор кафедри захисту інформації. Вінницький національний технічний університет.

Rezedent Oleksandr Olehovych – student of group 1BS-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia.

Kondratenko Natalia Romanivna - Ph.D., professor of Information Protection, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia