Розробка системи виявлення фішингових ресурсів на основі інтелектуального аналізу коду

УДК 004.056.53

Андрій Ковальжі
Національний університет «Одеська політехніка»,
9480575@stud.op.edu.ua

Фішингові атаки залишаються однією з найбільших загроз для безпеки інформаційних систем[1], незважаючи на розвиток технологій захисту. Їх еволюція супроводжується використанням обфускації, поліморфізму та методів соціальної інженерії, що значно ускладнює своєчасне виявлення загроз класичними методами, такими як чорні списки чи простий евристичний аналіз. Тому актуальним завданням є розробка нових підходів до автоматизованого виявлення фішингових ресурсів із високою точністю та здатністю до адаптації.

Метою цієї роботи є розробка програмного забезпечення, яке буде здатне правильно виявляти файли, що містять фішингові ознаки. У даній роботі запропоновано новий підхід до виявлення фішингових HTML-ресурсів шляхом візуалізації їх бінарного представлення[2] та подальшого машинного аналізу отриманих векторів ознак. Основна ідея полягає у перетворенні HTML-коду сторінки у байтовий потік із подальшим формуванням зображення у градаціях сірого розміром 128×128 пікселів. Кожен байт представляється як піксель певної яскравості, що дозволяє отримати характерні патерни структури файлу. Цей підхід не залежить від текстового наповнення сторінки та мови вмісту, що робить його стійким до різних способів обфускації коду.

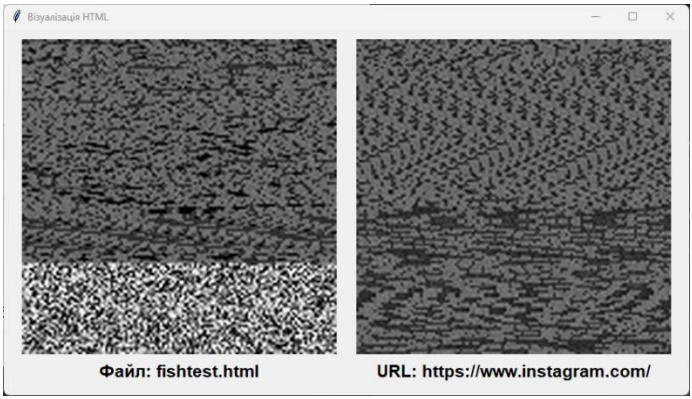

Рис. 1 – Приклад візуалізації фішингового та легітимного файлів

Для підвищення інформативності векторів ознак було розроблено метод поділу зображення на чотири логічні області (верхня, нижня, верхня середня та нижня середня частини). Для кожної області обчислюється нормалізована гістограма яскравості пікселів, що дозволяє зберігати локальні особливості

розподілу даних у структурі файлу. Отримані гістограми об'єднуються у єдиний вектор довжиною 1024 ознаки.

Для класифікації векторів було обрано алгоритм машинного навчання Support Vector Machine із використанням радіальної базисної функції (RBF)[3] як ядра. Процес налаштування гіперпараметрів моделі здійснювався методом GridSearchCV із перехресною валідацією. Навчання проводилося на датасеті, що складався з 3000 легітимних та 1700 фішингових HTML-файлів.

У результаті експериментів було встановлено, що точність класифікації складає 92.45%. Метрики Precision, Recall та F1-міра для обох класів склали близько 0.92, що свідчить про високу збалансованість моделі та її стійкість до помилок першого та другого роду. Було також побудовано матрицю конфузії та візуалізацію T-SNE, які підтвердили роздільну здатність моделі щодо кластеризації фішингових і легітимних векторів.

Розроблений програмний продукт має графічний інтерфейс, що дозволяє користувачеві завантажити локальний HTML-файл або ввести URL-адресу для автоматичної перевірки ресурсу на наявність ознак фішингової активності. У межах тестування система показала високу швидкість обробки запитів та стабільність роботи на реальних даних.

Запропонований підхід довів свою ефективність для задач виявлення фішингових атак за допомогою бінарної візуалізації та машинного навчання. Він демонструє потенціал для подальшого розвитку шляхом інтеграції глибших нейронних архітектур або самонавчальних методів, зокрема автоенкодерів чи контрастивного навчання, що дозволить ще краще пристосовувати систему до змін у природі загроз.

1. Sabillon R., Cano M. J., Serra-Ruiz J., Cavaller V. Cybercrime and Cybercriminals: A Comprehensive Study. International Journal of Computer Networks and Communications Security. 2016. Vol. 4. P. 165–176.

2. Baptista I., Shiaeles S., Kolokotronis N. A Novel Malware Detection System Based on Machine Learning and Binary Visualization : IEEE International Conference on Communications Workshops (ICC Workshops). Shanghai, China, 2019. P. 1–6.

3. Kumar A., Chatterjee J., Díaz V. A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. International Journal of Electrical and Computer Engineering (IJECE). 2020. Vol. 10, No 1. P. 486–493.

Розробка багаторівневих моделей захисту хмарної інфраструктури з використанням смарт-контрактів

УДК 004.056:343

Назар Козубаль[1], Ігор Пітух[2]
Західноукраїнський національний університет[1,2]
i.pitukh@wunu.edu.ua

Стрімкий розвиток хмарних обчислень спричинив значне зростання обсягів переданих та оброблюваних даних, а також масштабування інфраструктур IT-сервісів. Попри очевидні переваги хмарної моделі – гнучкість, масштабованість,