УДК 004.91
А. М. Стефанів, Н. В. Загородна канд. техн. наук, доц., Р.О. Козак, канд. техн. наук, доц.
Тернопільський національний технічний університет імені Івана Пулюя, Україна

КЛАСИФІКАЦІЯ МЕТОДІВ ВИЯВЛЕННЯ ФІШИНГУ В ІНТЕРАКТИВНИХ МУЛЬТИМЕДІЙНИХ ВИДАННЯХ

A.M. Stefaniv, N.V. Zagorodna, Ph.D., Assoc. Prof., R.O. Kozak, Assoc. Prof.

CLASSIFICATION OF METHODS OF PHISHING DETECTION IN INTERACTIVE MULTIMEDIA EDITIONS

Із розвитком інформаційних технологій набули великої популярності електронні інтерактивні мультимедійні видання, які надають доступ до актуальної інформації усім бажаючим. Інтерактивні мультимедійні електронні видання вирізняються з-поміж іншої електронної літератури низкою особливостей. Зокрема, вони містять матеріал, що функціонує у вигляді гіпертексту, вузли якого можуть поєднувати текстові документи, графічні зображення, відеозаписи, аудіозаписи тощо. Такі гіпертекстові посилання можуть вказувати не лише на оригінальні зовнішні ресурси, але, і на підроблені.

Видання відкритого типу, як от Вікіпедія, містять велику кількість статей з посиланнями, які додаються великою кількість користувачів з усього світу. Ці видання можуть бути використані для отримання зловмисниками персональної інформації читачів видання шляхом розміщення достовірної інформації з гіпертекстовими посиланнями на фейкові веб-сайти. Користувачі будучи переконаними, що користуються оригінальним сайтом, можуть ввести персональні дані, тим самим подарувати їх зловмисникам.

Наприклад, стаття про банківські акції, може містити посилання на сторінку отримання бонусу від банку з обов'язковою авторизацією. В даному випадку, якщо користувач перейде на фейковий веб-сайт, який візуально матиме вигляд оригінального, і введе дані необхідні для авторизації, тим самим передавши їх у руки шахраїв. Даний вид атаки називається "фішинг" [1].

Для реєстрації даного виду атак та ведення обліку фішингових веб-сайтів було засновано організацію Anti-Phishing Working Group (APWG). APWG була заснована в 2003 році компанією Tumbleweed Communications, фінансовими установами та постачальниками електронної комерції. У червні APWG 2004 була зареєстрована як незалежна корпорація, яка контролюється її радою директорів-засновників, керівниками та керівним комітетом [2].

Починаючи з 2005 року організація Anti-Phishing Working Group реєструє та перевіряє посилання на веб-сайти, які можуть виявитися фейковими. Дана інформація за попередні роки доступна на сайті організації [3]. Гістограма на рисунку 1 зображує кількість зареєстрованих нових унікальних фішингових сайтів кожного року. Дані станом на липень 2019 року.

Очевидно, що з 2005 року щороку з'являються нові фейкові сайти. Кількість сайтів в межах навіть однієї інформаційної системи може сягати такого рівня, що ручна перевірка всіх посилань в межах статті буде забирати значну кількість часу. Виникає необхідність задіяння, крім ручної модерації текстів та сторонніх посилань, автоматизованих методів виявлення загроз.

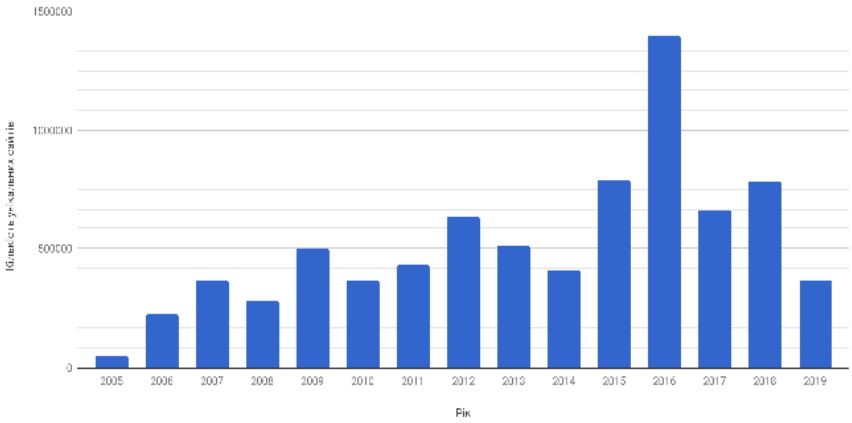

Рисунок 1. Гістограма кількості зареєстрованих нових унікальних підроблених сайтів

Одним зі способів виявлення таких атак є аналіз гіпертекстових посилань за допомогою традиційних способів – використання чорних або білих списків сайтів [4]. Висока точність є перевагою методу фільтрування по списках. Недоліком ж такого підходу є необхідність постійного оновлення списків до актуальної версії.

Іншим способом виявлення посилань на фейкові сайти є використання класифікаторів машинного навчання. Система може класифікувати нові фішингові сайти з використанням моделі, розробленої за допомогою навчальних наборів даних, що містять відомі атаки. Однією з головних проблем такої моделі є те, що дуже мало наборів даних з фішинговими адресами доступні у відкритому доступі.

Автори статті [5] навели порівняння загальних алгоритмів машинного навчання для класифікації URL-адрес, таких як SVM, класифікатор Naïve Bayes, дерево рішень та нейронна мережа. Однією з головних проблем класифікаторів дерев рішень є надмірність. Як правило, дерево рішень дуже добре класифікує дані навчального набору, але дає слабкі результати за допомогою тестового набору даних. Для підвищення точності та зменшення надмірності необхідно проводити “обрізання” дерев.

Крім того, точність може бути підвищена за допомогою ансамблю дерев. Нейронні мережі покази найнижчу точність для набору даних, що досліджувався в [5].

Література

1. Ramzan, Zulfikar. "Phishing attacks and countermeasures". – In Stamp, Mark, Stavroulakis, Peter (eds.). Handbook of Information and Communication Security. Springer. – 2010 – ISBN 978-3-642-04117-4.

2. About the APWG – Електронні видання [Електронний ресурс] – Режим доступу: https://apwg.org/about-us/ – Дата доступу: 10.11.2019. – APWG | About us

3. APWG Reports – Електронні видання [Електронний ресурс] – Режим доступу: https://apwg.org/trendsreports/ – Дата доступу: 10.11.2019. – APWG | Phishing Activity Trends Reports

4. Bergholz, A., Chang, J. H., Paass, G., Reichartz, F., & Strobel, S. (2008, August). Improved Phishing Detection using Model-Based Features. InCEAS

5. Detecting Phishing Emails [Електронний ресурс] – Режим доступу: https://meu.edu.jo/uploads/1/590422b4d5dd8_1.pdf – Дата доступу: 10.11.2019 – Detecting Phishing Emails Using Machine Learning Techniques.