УДК 004.056:004.8

Боцанюк І.М, магістрант,
Бродський Ю.Б., к.т.н., доцент
Державний університет «Житомирська політехніка»

АДАПТИВНА МОДЕЛЬ ПІДВИЩЕННЯ СТІЙКОСТІ СИСТЕМ ВИЯВЛЕННЯ ФІШИНГУ

Стрімка еволюція соціотехнічних атак формує потребу у створенні засобів кіберзахисту, здатних гнучко реагувати на нові техніки обходу фільтрації. Незважаючи на прогрес у сфері інформаційної безпеки, фішингові атаки залишаються одним із головних векторів початкового проникнення в корпоративні мережі [1], тоді як статичні моделі поступово втрачають ефективність унаслідок появи адаптивних методів маскування та контекстної маніпуляції [2]. Особливо вразливою залишається ділова електронна пошта, де зловмисники експлуатують семантику й поведінкові патерни, що ускладнює виявлення та призводить до значних економічних втрат [3]. Додаткову складність становить те, що сучасні атаки часто не містять шкідливих вкладень, а базуються на створенні переконливих підроблених сценаріїв комунікації. Основна проблема полягає у неможливості статичних систем своєчасно відображати зміни в поведінці атакуючих.

Метою роботи є розроблення адаптивної моделі виявлення фішингу, що поєднує гнучке зважування ризиків, аналіз поведінкових характеристик та механізми активного навчання. У ролі базової архітектури для оброблення тексту використано трансформерну модель DistilBERT, що забезпечує збалансоване співвідношення між точністю класифікації та швидкодією.

У межах дослідження створено архітектуру гібридної системи, яка інтегрує результати різних модулів аналізу в єдине адаптивне рішення. Центральним елементом є коефіцієнт адаптивної довіри - механізм, що коригує вагу прогнозу моделі відповідно до рівня впевненості; у разі низької визначеності посилюється роль евристичних перевірок та системних правил безпеки. Якщо в листі виявлено критичні індикатори ризику, такі як приховані символи, аномальна структура тексту чи ознаки соціальної інженерії, повідомлення автоматично маркується як підозріле.

Особливу увагу приділено модулю поведінкового аналізу, який формує персоналізований профіль користувача, враховуючи часові закономірності листування, історію попередніх комунікацій і сталість довірених доменів. Коли новий лист суттєво відхиляється від звичних патернів, система підвищує ризикову оцінку за рахунок додаткових

штрафних факторів. Це дозволяє виявляти технічно коректні, але поведінково аномальні повідомлення, які легко обходять класичні сигнатурні та контентні фільтри.

Для забезпечення безперервної адаптації впроваджено механізм активного навчання на основі локального зворотного зв'язку: виправлення користувача автоматично використовуються для донавчання моделі на пристрої. Такий підхід зменшує кількість хибнопозитивних спрацьовувань і забезпечує поступове удосконалення системи відповідно до реального середовища без передавання конфіденційних даних на зовнішні сервери.

Модуль очищення та нормалізації вхідного тексту додатково захищає систему від маніпулятивних впливів, виявляючи приховані символи та спроби прихованої модифікації контенту. Крім того, інтегрований модуль пояснення рішень формує інтерпретований звіт про причини блокування або маркування листа, підвищуючи прозорість і керованість процесу виявлення.

Запропонована адаптивна модель усуває основні обмеження статичних систем шляхом поєднання динамічного зважування ознак, поведінкового аналізу та локального активного навчання, що суттєво підвищує стійкість системи до нових технік атак і водночас забезпечує конфіденційність обробки даних. Перспективним напрямом розвитку є розроблення методів колективного захищеного навчання, що дозволить поширювати інформацію про загрози між користувачами без розкриття змісту приватної кореспонденції.

Список використаних джерел:

1. 2024 Data Breach Investigations Report [Електронний ресурс] / Verizon. – New York : Verizon, 2024. – URL: https://www.verizon.com /business/resources/reports/dbir/.

2. Lu J. Learning under Concept Drift: A Review / J. Lu, A. Liu, F. Dong // IEEE Transactions on Knowledge and Data Engineering. – 2018. – Vol. 31, № 12. – P. 2346–2363.

3. Internet Crime Report 2023 [Електронний ресурс] / Federal Bureau of Investigation. – Washington, D.C. : Internet Crime Complaint Center, 2023. – 33 p. –URL: https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Re port.pdf.