УДК 004.056
В. Дячун; В. Ваврічен
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

МЕТОДИ ВИЯВЛЕННЯ СПАМУ

UDC 004.056
V. Diachun; V. Vavrichen

METHODS OF SPAM DETECTION

Спам – це непотрібна, часто рекламна або шахрайська інформація, що надходить через електронну пошту, соціальні мережі або інші канали комунікації. Спам часто містить шкідливі посилання або вкладення, що можуть привести до зараження комп'ютерів вірусами, шкідливим програмним забезпеченням чи фішингом. Велика кількість спам-повідомлень засмічує поштові скриньки, що може ускладнювати роботу та знижувати продуктивність. Спам несе із собою нав'язливу рекламу, що може займати важливі ресурси пристроїв, наприклад, інтернет-трафік або пам'ять.

Боротьба зі спамом є однією з найстаріших проблем комп'ютерної безпеки, і, водночас, важливою складовою забезпечення безпеки та ефективності цифрового середовища сьогодення. Існує кілька підходів до виявлення та боротьби зі спамом. Традиційні методи виявлення спаму діляться на дві великі групи:

1. Методи, що базуються на аналізі повідомлення. Один з найпоширеніших методів – це фільтрація на основі ключових слів і шаблонів. Програми автоматично аналізують текст повідомлення на наявність певних слів або фраз, характерних для спаму. Крім того, аналіз контенту листа може здійснюватись з використанням сигнатур в оновленій базі даних, із застосуванням статистичних методів на основі теореми Байєса, за допомогою використання SURBL ((Spam URL Real-time Block Lists)
2. Репутаційні схеми. Їхнє завдання – виявляти розсилки однотипних листів серед великої кількості користувачів. Цей метод оцінює репутацію відправника, базуючись на історії його діяльності. Якщо адреса відправника була раніше пов'язана з великим обсягом спаму, система може позначити його повідомлення як спам.
3. Методи, засновані на визнанні відправника спамером. Ці методи спираються на різні чорні списки IP-адрес та адрес електронної пошти. Відомі адреси або IP-адреси, з яких часто надходить спам, можуть бути внесені в чорні списки, що допомагає блокувати подібні повідомлення ще на етапі їх надходження.
4. Методи, засновані на перевірці адреси електронної пошти та доменного імені відправника.

Крім цього фільтрація спам-листів може ґрунтуватись на аналізі поведінки користувачів. Якщо користувач отримує спам-повідомлення від незнайомих адрес чи в незвичний час, система може попередити про ймовірний спам. Це свого роду задача виявлення аномалій.

Сьогодні для фільтрації електронної пошти активно розробляються методи, що навчаються, або інтелектуальні методи, засновані на алгоритмах машинного навчання та техніках глибокого навчання. Підвищення точності роботи цих методів є актуальним науковим завданням.

Література

1. Saadat Nazirova. Survey on Spam Filtering Technique. Communications and Network, 2011, No. 3, P.153-160. doi:10.4236/cn.2011.33019
2. J, Rajesh & G, Mahalakshmi & P, Sudarshan. Email Spam Detection using Machine Learning Techniques. IARJSET, 2020, No 8, P.189-193.
3. AbdulNabi, Isra’a & Yaseen, Qussai. Spam Email Detection Using Deep Learning Techniques. Procedia Computer Science. 2021, No. 184 (2), P. 853-858