УДК 004.891
Р.І. Тимішак
Тернопільський національний технічний університет імені Івана Пулюя, Україна

АВТОМАТИЧНА ФІЛЬТРАЦІЯ ЯК ОДИН ІЗ ЗАСОБІВ БОРОТЬБИ ЗІ СПАМОМ

R.I. Tymishak

AUTOMATIC FILTERING AS A MEANS TO FIGHT SPAM

Рекламні листи, як правило, сильно відрізняються від звичайної кореспонденції; поширеним методом боротьби з ними стало відфільтровування їх з вхідного потоку пошти. На даний час цей метод – основний і найбільш широко використовується.

Існує програмне забезпечення для автоматичного визначення спаму (спам-фільтри). Воно може бути призначене для кінцевих користувачів або для використання на серверах. Це програмне забезпечення використовує два основні підходи.

Перший полягає в тому, що аналізується зміст листа і робиться висновок, спам це чи ні. Лист, класифікований як спам, відокремлюється від іншої кореспонденції: він може бути позначений, переміщений в іншу папку, видалений. Таке програмне забезпечення може працювати як на сервері, так і на комп'ютері клієнта. В останньому випадку користувач не бачить відфільтрованого спаму, але продовжує нести витрати, пов'язані з його прийомом, тому що фільтруюче програмне забезпечення отримує кожен лист і тільки потім вирішує, показувати його чи ні. З іншого боку, якщо програмне забезпечення працює на сервері, користувач не несе витрат з передачі його на свій комп'ютер.

Другий підхід полягає в тому, щоб, застосовуючи різні методи, визначити відправника як джерело спаму, не заглядаючи в текст листа. Це програмне забезпечення може працювати тільки на сервері, який безпосередньо приймає листи. При такому підході додатковий трафік витрачається тільки сервером на спілкування із відповідними поштовими програмами боротьби зі спамом (тобто на відмови приймати листи) і звертання до інших серверів (якщо такі потрібні) при перевірці.

Існують також спеціалізовані online-сервіси, наприклад, Лабораторія Касперського (сервіс Kaspersky Hosted Security) [1], СПАМОРЕЗ [2], які надають платний захист від спаму. Зміна MX-запису в доменному імені підприємства особливим чином дозволяє перенаправити пошту на спеціалізований поштовий сервер, де вона очищається від спаму і вірусів, а потім – на корпоративний поштовий сервер. Метод підходить для корпоративних користувачів і не годиться для власників поштових скриньок в публічних поштових системах.

Ще одна проблема автоматичної фільтрації в тому, що вона може помилково визначати як спам корисні повідомлення. Тому багато сервісів (наприклад, Yahoo!mail) не видаляють ті повідомлення, які фільтр визнав спамом, а поміщають їх в окрему папку.

Програми автоматичної фільтрації використовують статистичний аналіз вмісту листа для прийняття рішення, чи є воно спамом. Найбільшого успіху вдалося досягти за допомогою алгоритмів, в основу яких покладено теорему Байеса. Для роботи цих методів потрібно попереднє навчання фільтрів шляхом передачі йому розсортованих вручну листів для виявлення статистичних особливостей нормальних листів і спаму.

Метод дуже добре працює при сортуванні текстових повідомлень (в т.ч. HTML). Після навчання на досить великій вибірці вдається відсікти до 95-97% спаму. Для уникнення таких фільтрів виробники спаму іноді поміщають змістовну частину в

картинку, вкладену в лист, текст же або відсутній, або випадковий, що не дозволяє фільтру скласти статистику для розпізнавання таких листів. У цьому випадку необхідно користуватися програмами розпізнавання тексту (більшість сучасних поштових програм цього не підтримують), або використовують інші методи.

Запорука надійної роботи байєсівського методу - постійне донавчання фільтра і вказування йому на вчинені помилки. У поштових програмах для цього вводиться можливість ручної позначки повідомлення спам/не спам, а в поштових сервісах в мережі Інтернет – клавіша поскаржитися на спам [3].

Багато програм і поштових сервісів в мережі Інтернет дозволяють користувачеві задавати власні фільтри. Такі фільтри можуть складатися зі слів або, рідше, регулярних виразів, в залежності від наявності або відсутності яких повідомлення потрапляє або не потрапляє в сміттєву папку. Однак така фільтрація трудомістка і негнучка, крім того, вимагає від користувача певної міри знайомства з комп'ютерами. З іншого боку, вона дозволяє ефективно відсіяти частину спаму, і користувач точно знає, які повідомлення будуть відсіяні і чому [4].

Поряд з автоматичною фільтрацією для контролю поштових розсилок можуть використовуватися такі методи:

– загальні посилення вимог до листів і відправників, наприклад – відмова у прийомі листів з неправильною зворотною адресою (листи з неіснуючих доменів), перевірка доменного імені за IP-адресою комп'ютера, з якого йде лист тощо;

– сортування листів за змістом полів заголовка листа дає можливість позбутися від деякої кількості спаму;

– системи типу «виклик-відповідь» дозволяють переконатися, що відправник – людина, а не програма-робот. Використання цього методу вимагає від відправника виконання певних додаткових дій, часто це може бути небажано.

– системи визначення ознак масовості повідомлення, такі як Razor і Distributed Checksum Clearinghouse.

– розробка Міністерства оборони США – кожен представник цього міністерства має «типову карту доступу» побудовану на основі смарт-карти з вбудованим мікропроцесором, в якій записані цифрові сертифікати PKI з інформацією про користувача. Така карта використовується як посвідчення особи, для аутентифікації і доступу до комп'ютерних мереж, друку і сканування документів і в тому числі для підтвердження повноважень при відправці листів. Кожен житель США має Social Security number, який складається з 9 цифр, тому при бажанні CAC можна легко використовувати для аутентифікації користувача в Інтернет.

Література.

1. Kaspersky Hosted Email Security Protects Russia’s Largest Department Store // Kaspersky.lab. – Режим доступу: http://www.kaspersky.ru/gum. – Дата доступу: листопад 2016 року. – Заголовок з екрану.

2. Спаморез // Защита от СПАМА. – Режим доступу: https://spamorez.ru/. – Дата доступу: листопад 2016 року. – Заголовок з екрану.

3. Спам, види спаму і боротьба зі спамом // Безкоштовні антивіруси і антивірусні програми для пк, кпк, нетбуків та мобільних телефонів. – Режим доступу: http://best-free-soft.at.ua/publ/spam_vidi_spamu_i_borotba_zi_spamom/1-1-0-33. – Дата доступу: листопад 2016 року. – Заголовок з екрану.

4. Спам // Знаймо разом. – Режим доступу: http://znaimo.com.ua/Спам. – Дата доступу: листопад 2016 року. – Заголовок з екрану.