УДК 004.056.57
Катерина КУЗЬМА
katushke@mail.ru
ORCID: 0000-0002-0937-7299
Василь ЗІВЕНКО
vazivenko@mail.ru
м. Миколаїв

АНАЛІЗ МЕТОДІВ ФІЛЬТРАЦІЇ ЕЛЕКТРОННОЇ ПОШТИ ВІД СПАМУ

У роботі досліджено методи фільтрації електронної кореспонденції, проаналізовано їх переваги та недоліки, рівень реалізації в поштових програмах, що дозволило визначити найбільш ефективні й перспективні з точки зору практичного застосування. Розглядаються такі методи фільтрації, як «чорні списки», авторизація поштових серверів, «сірі списки», статистичні методи фільтрації спаму, фільтри, що побудовані з використанням технологій штучного інтелекту. Наведено класифікацію методів спам-захисту, яка базується на двох основних підходах фільтрації – за способом посилки листа (формальні методи) та за його змістом (семантичні методи фільтрації). Визначено, що формальні методи вимагають постійного оновлення списків доступу, створюючи додаткове навантаження на користувача, завдяки чому вони були віднесені до малоефективних технологій фільтрації. Фільтри, побудовані з використанням семантичних методів, а саме з використанням нейромереж, вимагають навчання тільки на початковому етапі, довчаючись надалі самостійно, істотно знижуючи при цьому навантаження на користувача. В результаті проведеного дослідження встановлено, що розвиток, розробка семантичних методів, які базуються на використанні нейромереж, є актуальною задачею.

Ключові слова: фільтрація спаму, «чорні списки», «сірі списки», формальні методи фільтрації, семантичні методи фільтрації.

Постановка проблеми. Фахівці інформаційної безпеки змушені усе більше сил і часу приділяти боротьбі з SPAMом тому, що такі розсилання завдають серйозної шкоди інформаційним системам.

За останні роки було винайдено чимало способів боротьби з небажаною кореспонденцією. На жаль, зловмисники стежать за протидією поширенню SPAMа й винаходять всі нові прийоми для обходу фільтрів. До того ж нерідко фільтрація SPAM-листів приносить більше шкоди, чим користі: разом з рекламою не доходять до адресата й важливі ділові або особисті повідомлення. Таким чином, всі дослідження в області боротьби з незапитуваною кореспонденцією надзвичайно актуальні в цей час.

Аналіз останніх досліджень і публікацій. Аналіз, розробку методів розпізнавання спаму досліджено в роботах О. М. Певзнера, М.О. Семенової, А.М. Мироненка, Г. Робінсона, П. Грехама, Викас П. Дешпанде [1–6]. Кожен із методів має свої переваги та недоліки, для практичного застосування необхідно виконати їх класифікацію за функціональністю, сферою застосування, ефективністю тощо.

Постановка завдання. Метою роботи є аналіз та класифікація методів розпізнавання спама для виявлення переваг й недоліків існуючих підходів фільтрації електронної кореспонденції й визначення перспективних підходів антиспамового захисту, які знаходяться в розробці.

Виклад основного матеріалу. Комплексний захист від спаму складається з наступних етапів: аналіз відправника; використання фільтрів; аналіз змісту листа. Технічно дані етапи базуються на двох основних підходах фільтрації SPAM – фільтрація за формальними ознаками повідомлення (за способом посилки й оформленню) й за його змістом (семантичні методи фільтрації) (рис. 1). Формальні методи включають фільтрацію за списками (поштових адрес, IP-адрес) та за формальними ознаками листа (наявність багатьох відправників, відсутність одержувача, формат, розмір тощо).

Фільтрація з використанням «чорних списків» у цей час є, мабуть, найпоширенішим способом. «Чорні списки» діляться на два види: статичні й динамічні. Статичні «чорні списки» представляють собою списки адрес електронної пошти й доменів. При виявленні листа, відправленого з одного з таких адрес або доменів, сервер блокує даний лист. Багато програм фільтрації електронної пошти

ISSN 2520-2820 (online), ISSN 2524-0978 (print)

дозволяють поповнювати ці списки новими адресами, з яких ведеться розсилання SPAM-листів і підтримувати власні «чорні списки».

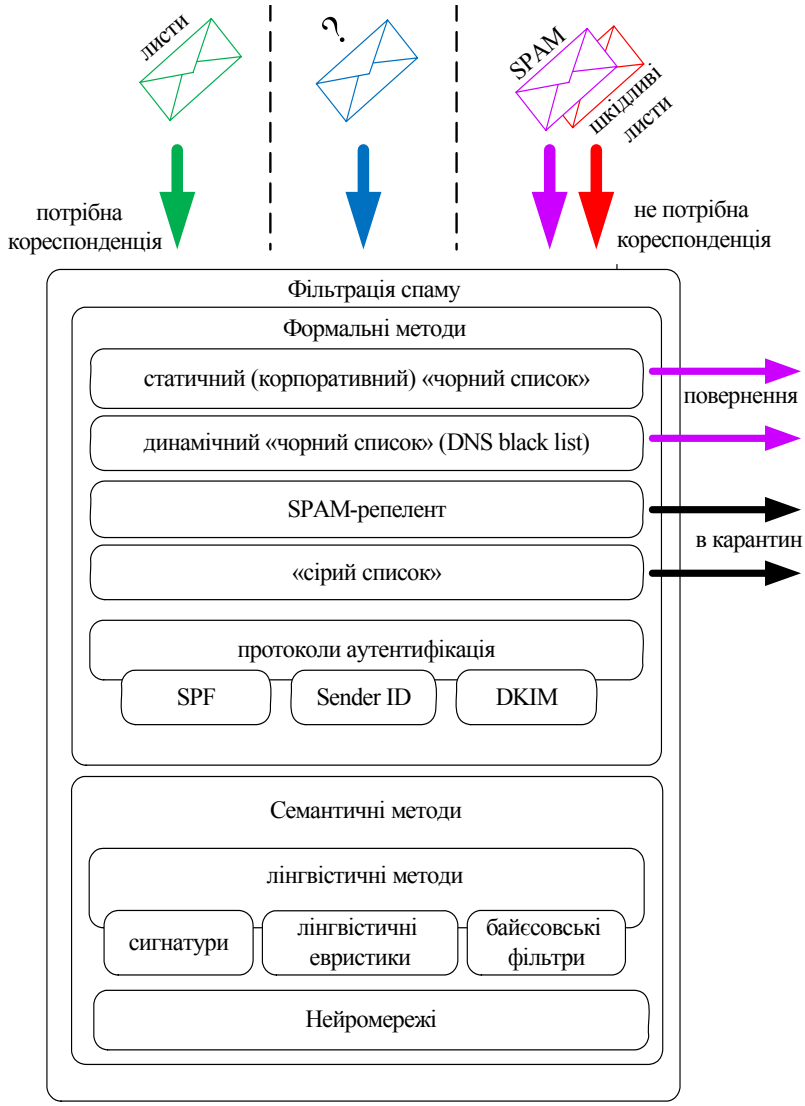

Рис. 1. Класифікація методів SPAM – фільтрації

Як правило, відновлення статичних «чорних списків» виконується вручну. Дана обставина є мінусом, тому що вимагає додаткових тимчасових ресурсів адміністратора.

Динамічні «чорні списки» (DNSBL – DNS black list) істотно відрізняються від статичних. DNSBL є мережною службою, яка надається провайдером «чорних списків». Ці провайдери відслідковують адреси IP (іноді й імена доменів), скомпрометовані зловмисниками. Поштові фільтри, що підтримують застосування динамічних «чорних списків», формують запит до провайдера «чорного списку», який містить адресу відправника, а також адреси поштових серверів, через які проходив лист шляхом до одержувача. Якщо при виконанні запиту з'ясується, що адреса є в «чорному списку» провайдера послуги, то висока ймовірність того, що лист є SPAMом. Звернення до служб динамічних «чорних списків» здійснюється через службу DNS для перевірки, чи не міститься в списках SPAMерів адреси IP, перерахованої в заголовку листа.

Для постійного поповнення «чорних списків», як правило, використаються наступні методи: поштові сервери-«пастки» (SPAM-traps). Ці сервери емулюють звичайний MTA (Mail transfer agent), записуючи при цьому IP-адреси комп'ютерів, що розсилають SPAM; скарги користувачів на SPAM-листи; автоматичне поповнення списків у результаті роботи інших фільтрів (наприклад, ймовірносно-статистичних фільтрів Байєса).

Серед переваг даного методу фільтрації листів варто виділити: практично будь-який MTA вміє звертатися до DNSBL, для цього не потрібна установка додаткового програмного забезпечення; системному адміністратору немає необхідності брати участь в обслуговуванні й підтримці «чорних списків».

Недоліками підходу є: різні «чорні списки» мають різну ефективність. Щоб вибрати оптимальний набір таких списків, необхідний час на проведення досліджень; при зверненні до великої кількості «чорних списків» за допомогою DNS-запитів може різко зростати завантаженість каналу. На кожне вхідне з'єднання MTA буде відправляти один запит до кожного сервера «чорного списку», зазначеному у файлі конфігурації; деякі «чорні списки» вносять широкі діапазони адрес різних провайдерів і вимагають оплати за видалення; не виключена ймовірність помилок другого роду при аналізі листа.

Технологія фільтрації SPAM-репелент (spam-repellent) заснована на особливостях протоколу SMTP. В RFC 2821 сказано, що клієнт SMTP повинен чекати відгуку сервера при встановленні з'єднання протягом, принаймні, п'яти хвилин. У програмному забезпеченні розсилання SPAM-листів (SPAM-автомат) складно встановлювати такий тривалий період очікування. Використовуючи дану особливість SPAM-автомата, сервер-одержувач може навмисно поставити затримку перед відгуком, розраховуючи на те, що автомат перестане чекати або сервер, який ним використовується, встигне за час затримки потрапити в «чорні списки» DNSBL. У подібній ситуації SPAM-автомат може або завершити з'єднання, не дочекавшись відгуку, або почати передачу повідомлення негайно без очікування відгуку. В останньому випадку сервер-одержувач скидає з'єднання й ніяких даних не приймає. Основні труднощі полягають у виборі затримки. Дана технологія була реалізована в поштовому сервері «Kerio Mail Server», який дозволяє налаштувати величину затримки. Розроблювачі рекомендують вибирати максимальний час затримки рівний 30 с, при цьому сервер відфільтровує від 50 до 70 % SPAM-листів, що є перевагою даного методу.

Серед недоліків варто виділити наступне: не всі MTA підтримують технологію SPAM-репелент; SPAM-репелент не ефективний, якщо SPAM-листи пересилаються через відкритий поштовий ретранслятор, який перебуває в Інтернеті і є повноцінним MTA, і якщо в скриньку користувача перенаправляється пошта з іншої скриньки, не захищеної від SPAM-листів; не виключена ймовірність помилок другого роду при збільшенні затримки.

Принцип дії «сірих списків» (greylisting) як і SPAM-репелентів засновано на тактиці розсилання SPAM-листів: як правило, SPAM-листи розсилаються в дуже короткий час. Робота «сірих списків» полягає в навмисній затримці листів на короткий час. При цьому адреса й час пересилання заносяться в базу даних «сірого списку», а клієнтові SMTP відсилається повідомлення про тимчасову помилку. Згідно RFC 2821, клієнт SMTP повинен зберегти лист у черзі й повторити пересилання протягом п'яти днів. У загальному випадку інтервал очікування варто робити не меншим за 30 хвилин, однак підходи зі змінним часом очікування будуть давати переваги в тих випадках, коли клієнт SMTP може визначити причину невдачі передачі листа.

При використанні даної технології виникає досить серйозна проблема: «сірий список» може затримувати листи на інтервал часу аж до п'яти днів. Даний метод можна поліпшити, доповнивши його підтримкою одного із протоколів аутентифікації, наприклад, SPF, який дозволяє розпізнати підробку e-mail-адрес. Таким чином, ті листи, які задовольняють SPF можна не перевіряти на «сірі списки». Даний синтез дозволить скоротити час затримки більшої частини корисної кореспонденції. Переваги «сірих списків» – фільтрація до 50 % SPAMa.

Найпоширенішими технологіями аутентифікації на рівні SMTP є SPF (Sender Policy Framework), Sender ID і DKIM (DomainKeys Identified Mail).

SPF дозволяє власникові поштового домена сформулювати політики, тобто описати, у якому ступені він рекомендує довіряти тим або іншим серверам, що відправляють пошту від імені поштових адрес у цьому домені, а одержувачеві листа перевірити, чи відповідають два параметри: IP-адреса сервера відправника й адреса, зазначена в листі – політиці домена, якому ця адреса належить.

Суть технології полягає в наступному: в DNS-запис домена додається спеціальне поле типу .txt, що містить інформацію про сервери, яким дозволено або заборонено відправляти пошту від імені даного домена. Крім того, запис може містити опис політики поводження для всіх сторонніх серверів. В SMTP-сесії при одержанні mail приймаюча сторона робить DNS-запит, одержує SPF-запис і порівнює

ISSN 2520-2820 (online), ISSN 2524-0978 (print)

IP-адресу сторони, що посилає SMTP-сесії з SPF-політикою. Результатом є рекомендація із одержання або неодержання листа.

Недоліки використання SPF для фільтрації пошти: аналіз SPF-політики для прийнятої пошти вимагає модифікації ПЗ поштових серверів; при буквальному виконанні SPF-політики можуть бути відкинуті легітимні листи.

Технологія Sender ID була розроблена в Microsoft на основі технології SPF і є сумісною з нею. Даний метод аутентифікації відправників E-mail реалізований у сервісах Microsoft: Hotmail і MSN. Спочатку дана технологія мала назву Caller ID. Sender ID, як і SPF, використає ті ж самі txt-записи ресурсів DNS. Основне розходження між SPF і Sender ID полягає в тому, що Sender ID перевіряє адреси «from», що містяться в тілі повідомлення e-mail, а не тільки адреси відправника рівня SMTP.

Ще однією технологією для аутентифікації e-mail відправників є DKIM (DomainKeys Identified Mail), яка є об'єднанням розробок DomainKeys фірми «Yahoo!» і Internet Identified Mail фірми «Cisco». DKIM працює як стандартна криптосистема. Власник поштового сервісу (відправник) генерує пару криптоключів (відкритий і закритий). При цьому допускається генерація декількох криптопар. Відкритий ключ публікується в записах DNS, а закритий зберігається на поштовому сервері, що підтримує DKIM. Кожне вихідне повідомлення забезпечується підписом, що зберігається в його заголовку. MTA на стороні одержувача за допомогою публічного ключа перевіряє, чи дійсно підпис було згенеровано доменом, зазначеним в адресі відправника. Дана технологія в цей час застосовується компанією «Google» для позначення всієї вихідної кореспонденції цифровим підписом DomainKeys.

Недоліками DKIM є: необхідність серверів і для вихідної і для вхідної пошти; додаткові накладні витрати на обробку поштових повідомлень; збільшення частоти перегляду записів DNS.

Семантичні методи передбачають розпізнавання за змістом листа (словосполучення, евристики, статистика) або розпізнавання за зразками листів (за сигнатурами). Для роботи семантичних методів використовуються фільтри здатні до самонавчання, при цьому необхідно здійснювати їх початкове навчання, тобто розсортовувати вручну листи для виявлення статистичних особливостей нормальних листів і SPAM. Таким чином, задача фільтрації SPAM, розглядається як задача класифікації – визначення належності об'єкта (електронного повідомлення) до одного з заздалегідь виділених класів (спам і «не спам») на підставі аналізу сукупності ознак, що характеризують даний об'єкт.

Теорема Байєса лежить в основі багатьох сучасних систем штучного інтелекту, призначених для роботи в умовах невизначеності. Такі системи дають ймовірнісну оцінку, тому звичайно не заміняють експерта, а забезпечують підтримку прийняття рішення.

Нехай $F_S(W_i)$ – кількість SPAM-листів, у яких зустрілося слово W_i, а $F_{NS}(W_i)$ – кількість корисних листів, у яких зустрілося слово W_i; H_S – гіпотеза про те, що лист є SPAMом, H_{NS} – корисний лист. Тоді ймовірність того, що поява слова W_i у листі означає SPAM, обчислюється за формулою:

$$P(W_i \mid H_s) = \frac{F_s(W_i)}{F_s(W_i) + F_{NS}(W_i)},$$

а ймовірність того, що слово W_i не вказує на SPAM у листі:

$$P(W_i \mid H_{NS}) = \frac{F_{NS}(W_i)}{F_s(W_i) + F_{NS}(W_i)}.$$

Якщо вектор W включає всі m слів нового листа, то ймовірність того, що він SPAM, обчислюється за формулою Байєса таким чином:

$$P(H_S \mid W) = \frac{\prod_{j=1}^{m} P(W_j \mid H_s)}{\prod_{j=1}^{m} P(W_j \mid H_s) + \prod_{j=1}^{m} P(W_j \mid H_{NS})}.$$

Віднесення листа до SPAMу або корисних листів виконується з врахуванням заданого програмістом, адміністратором, користувачем почтової програми спам-фільтрації значення ймовірності, яке становить 0,6–0,8. Після ухвалення рішення щодо класифікації листа в базі даних обновляються ймовірнісні бази для слів, які входять до нього.

В основі фільтра лежить список ознак, за якими проводиться аналіз повідомлення і обчислюється умовна ймовірність спамності за кожного ознакою. Загальна ймовірність спаму повідомлення

визначається за одним з методів: 1) об'єднуються всі ймовірності за теоремою Байєса; 2) ймовірності комбінуються і перевіряються на скільки отримана множина схожа з випадковою (метод Фішера).

Основним недоліком Байєсівського класіфікатора є припущення, що події, які відповідають наявності того чи іншого слова в електронному листі або повідомленні, є незалежними по відношенню один до одного, тобто всі слова статистично незалежні. Це спрощення в загальному випадку є невірним для природних мов таких як англійська, де ймовірність виявлення прикметника підвищується за наявності, наприклад, іменника.

Висновки і перспективи досліджень. Максимальний результат, досягнутий байєсовськими фільтрами складає близько 95 % відфільтрованого спаму. Існують безліч модифікацій, які дозволяють збільшити ефективність фільтра: метод градуйованої фільтрації «спаму», який забезпечує підвищення якості оцінок даних за рахунок врахування наступних параметрів – кількості листів, в яких зустрічалися слова певної категорії, частоти використання слів у листах певної категорії, використання слів, що вперше зустрілися в листі і не існували до цього в базі [6]; побудова фільтра на основі багатошарового персептрона, що дозволяє враховувати семантичні зв'язки автоматично [4].

Перевага нейромережевого підходу перед наївним Байєсовським класифікатором полягає в тому, що не робиться ніяких попередніх припущень про характер небажаних повідомлень, а семантичні зв'язки враховуються автоматично. Малодослідженим залишається питання використання нейромереж, що добре зарекомендували себе в задачах розпізнавання образів, окремим випадком яких є фільтрація спаму. Таким чином, розвиток нейромережевого підходу стосовно фільтрації небажаних повідомлень є актуальним завданням.

Список використаних джерел

1. Graham P. A Plan for Spam / P. Graham, 2002. — Режим доступу: http://www.paulgraham.com/spam.html.
2. Robinson G. A Statistical Approach to the Spam Problem / G. Robinson // Linux Journal, 2003. — Issue #107. — Режим доступу: http://www.linuxjournal.com/article/6467.
3. Vikas P. Deshpand. An Evaluation of Naive Bayesian Anti-Spam Filtering Techniques / Vikas P. Deshpande, Robert F. Erbacher, Chris Harris // Proceedings of the 2007 IEEE Workshop o n Information Assurance United States Military Academy, West Point, 2007. — NY 20—22 June. — Режим доступу: http://digital.cs.usu.edu/~erbacher/publications/Bayes-Vikas2.pdf.
4. Мироненко А. Н. Алгоритм контентной фильтрации спама на базе совмещения метода опорных векторов и нейронных сетей : автореф. дис. на соискание науч. степени канд. техн. наук: спец. 05.13.19 «Методы и системы защиты информации, информационная безопасность» / А. Н. Мироненко. — СПб., 2012. — 18 с.
5. Певзнер О. М. Моделювання та аналіз ефективності зниження спам-ризиків за допомогою марківської фільтрації // Матеріали VI Всеукраїнської науково-практичної конференції «Комп'ютерне моделювання та інформаційні технології в науці, економіці та освіті» (Кривий Ріг, 26—28 квітня 2005 р.). — С. 156—164.
6. Семенова М. А. Модель и метод градуированной фильтрации «СПАМА»: автореф. дис. на соискание науч. степени канд. техн. наук: спец. 05.13.19 «Методы и системы защиты информации, информационная безопасность» / М. А. Семенова — СПб., 2009. — 20 с.

Kateryna KUZMA, Vasyl ZIVENKO
Mykolaiv

ANALYSIS OF METHODS OF EMAIL FILTERING FROM SPAM

In this work the methods of filtering e-mail, their strengths and weaknesses, the level of implementation in the e-mail programs have been analysed, that allowed to define the most effective and practical of them for use. Filtering techniques such as the «black list», the authorization of mail servers, «grey listing», statistical methods of filtering spam, filters, based on intellectual technologies are considered.

The classification of spam-protection methods based on two main approaches of filtering – way of sending a letter (formal methods) and analyzing the contents of the letter (semantic filtering methods) have been proposed. During research was determined that formal methods require constant updating access lists, creating an additional burden on the user, thus allowed to assign them to ineffective filtering technologies.

Filters based on neural networks require training initially, proceeding learning independently in a future, significantly lowering the burden on the user. In a result was proved that development of semantic methods based on the use of neural networks is an actual scientific research.

Key words: spam filtering, «black list», «grey listing», formal methods of filtering, semantic filtering methods.

Екатерина КУЗЬМА, Василий ЗИВЕНКО
г. Николаев

АНАЛИЗ МЕТОДОВ ФИЛЬТРАЦИЯ ЭЛЕКТРОННОЙ ПОЧТЫ ОТ СПАМА

В работе исследованы методы фильтрации электронной корреспонденции, проанализированы их преимущества и недостатки, уровень реализации в почтовых программах, что позволило определить наиболее эффективные и перспективные с точки зрения практического применения. Рассматриваются такие методы фильтрации, как «черные списки», авторизация почтовых серверов, «серые списки», статистические методы фильтрации спама, фильтры, построенные с использованием технологий искусственного интеллекта. Приведена классификация методов спам-защиты, которая базируется на двух основных подходах фильтрации – по способу посылки письма (формальные методы) и по его содержанию (семантические методы фильтрации). Определено, что формальные методы требуют постоянного обновления списков доступа, создавая дополнительную нагрузку на пользователя, что позволило отнести их к малоэффективным технологиям фильтрации. Фильтры, построенные с использованием семантических методов, а именно с использованием нейросетей, требуют обучения только на начальном этапе, дообучаясь в дальнейшем самостоятельно, существенно снижая при этом нагрузку на пользователя. В результате проведенного исследования установлено, что развитие, разработка семантических методов, основанных на использовании нейронных сетей, является актуальной задачей.

Ключевые слова: фильтрация спама, «черные списки», «серые списки», формальные методы фильтрации, семантические методы фильтрации.

Стаття надійшла до редколегії 23.03.2017