УДК 004.056
DOI: 10.31673/2409-7292.2022.034854

Волков М.В., Бондаренко А.А.

ЗАСТОСУВАННЯ БАЙЄСІВСЬКОЇ ФІЛЬТРАЦІЇ ЕЛЕКТРОННИХ ЛИСТІВ ДЛЯ ВИЯВЛЕННЯ СПАМУ

Проведено дослідження впливу спам повідомлень та сучасних загроз щодо процесу обміну електронної пошти корпоративної інформаційної системи. Досліджено роль спаму та його вплив у корпоративних інформаційних системах. Досліджено особливості напрямів протидії та технології забезпечення захисту електронної пошти від спаму. На основі досліджень, проведених в роботі, розроблено баєсівський класифікатор для фільтрації електронних листів поштового обміну у корпоративній інформаційній системі.

Ключові слова: електронна пошта, спам, фільтрація повідомлень, класифікатор.

Вступ

Сучасна кібербезпека стає все менш означеною з більшою складністю та можливостями для обміну та передачі інформації, розвитку електронного урядування, ведення онлайн-бізнесу, надання мобільних та бездротових послуг. Інформаційне та комунікаційне середовище є відкритим для все більшого числа ризиків і загроз, які можуть мати негативні наслідки для фізичних та юридичних осіб.

Розповсюдження шкідливого програмного забезпечення є однією з найбільш небезпечних загроз, що впливає на безпеку даних у сучасному кіберсередовищі і може втручатися у весь бізнес. Дані, представлені в електронному вигляді, щодня піддаються небезпеці порушення конфіденційності, цілісності та доступності внаслідок впливу більш ніж 75 мільйонів різних примірників ШПЗ, що циркулює в інформаційно-комунікаційних системах та мережах.

Спам – це масова анонімна незапрошена розсилка поштових повідомлень користувачам, причому немає різниці, чи комерційна це реклама або просто корисна на думку посилача інформація. Слід відрізняти спам від легальних поштових розсилок, які, хоча і багато в чому повторюють багато рис спаму, є запрошеними користувачем і повинні доставлятися йому. Спам в сучасному Інтернеті є негожим зайняттям, і в законодавстві низки країн передбачені ті або інші види відповідальності за подібного роду діяльність.

Аналіз методів розпізнавання спаму

Метод "чорного", "білого" і "сірого" списків. Базою методу є аналіз зворотного IP-адресу відправника листа. Всі листи, відправлені з IP-адрес, занесених до "чорного списку", знешкоджуються ще на поштовому сервері, так і не досягаючи кінцевого користувача. Адреса вносится в "чорний список" на підставі того, що лист що прийшов з цієї адреси є спамом. З адресатами з "білого списку" дозволений обмін поштовими повідомленнями. У разі, коли IP-адреса листа не присутній ні в "чорному" ні в "білому" списку, то відправнику автоматично висилається запит на авторизацію, а IP-адреса заноситися в тимчасовий "сірий" список.

Метод листів-підтверджень. Метод базується на тому, що оскільки спам-розсилки відбуваються автоматично, по багатьом мільйонам адрес, а адреса відправника - у більшості випадків - підроблена, то підтвердження від справжнього спамера отримати не вдасться. Однак застосування даного методу різко знижує оперативність доставки листів, у багатьох випадках спам відправляється з реальних IP-адрес, а сучасне програмне забезпечення спамерів може генерувати підтвердження відправки листів.

Метод розпізнавання спаму за ключовими словами, які визначаються користувачем у вигляді деяких правил. Даний метод не отримав широкого поширення через складність і трудомісткість формування зазначених правил.

Метод створення рефератів

Основні труднощі при створенні рефератів полягає в тому, що практично однаковий зміст може бути виражений за допомогою різної кількості слів, досить великої кількості різних мовних конструкцій, словосполучень, слів синонімів. Питання дещо спрощується

через те, що реферати можуть бути створені за однаковими правилами, що враховує необхідність зменшення застосовуваних мовних конструкцій. При цьому більшість сучасних методик формування рефератів базуються на використанні семантичних мереж. У загальному випадку семантична мережа являє знання у вигляді графа, вузли якого відповідають фактам, а дуги - відносинам або асоціаціям між поняттями. Перевагою семантичних мереж є можливість визначення зв'язків між поняттями і специфічних правил виводу, визначених механізмом успадкування.

Застосування семантичних мереж дозволяє абстрагуватися від малоінформативних елементів формально-синтаксичної структури тексту (порядку слів, застави і т.п.) і представляє його пропозиціональну структуру в термінах описуваних ситуацій (предикатів) і їх учасників (аргументів) в певних семантичних ролях [4, 5, 6]. Однак, в задачі розпізнавання спаму, повне уявлення змісту тексту в формі семантичної мережі є надмірною і непродуктивним. Таке уявлення має великий обсяг (перевищує обсяг документа), а його обробка вимагає розвинених нетривіальних засобів для пошуку та порівняння структур на графах, що в свою чергу вимагає використання значних обчислювальних ресурсів [5].

Функція смислового пошуку дозволяє отримати відповідь на запит, сформований у вигляді фрази природної мови, словосполучень або просто набору ключових слів. При цьому яку видобувають у відповідь на запит інформація може мати іншу граматичну форму або взагалі не згадуватися явно в тексті запиту, проте мати логічний зв'язок з текстом запиту.

У той же час, можливості смислового пошуку і рубрикації тексту не відповідають потребам системи захисту від спаму. Так розумовий пошук слова будови в листі, присвяченому рекламі житла закінчився безрезультатно. Результати [4, 5, 6] показують, що якісно вирішити питання рубрикації і смислового пошуку можливо за рахунок порівняння рефератів і / або тематичної структури текстів з використанням граматичних словників. Для проведення такого порівняння, можливо, використовувати рекурентні семантичні нейронні мережі або імовірнісні нейронні мережі [6, 7]. Порівняння зазначених типів нейронних мереж показує, що рекурентні семантичні нейронні мережі мають більшу продуктивність і потужністю в задачах класифікації та кластеризації зразків текстів. Однак їх реалізація в системах розпізнавання спаму вимагає проведення додаткових досліджень. Крім цього практична реалізація зворотних семантичних мереж не завжди можлива через використання значних обчислювальних ресурсів.

Мета статті – побудова ефективного фільтру для захисту електронної пошти корпоративної інформаційної системи від спаму.

Метод байєсівської фільтрації

Кожному слову, яке зустрічається в електронному листуванні присвоюється два значення: ймовірність його наявності в спам (z) і ймовірність його присутності в листах, дозволених для проходження ($1\text{-}z$). Кожному новому листу за допомогою формули Байеса виставляється оцінка (Z):

$$Z = A/(A+B), \tag{1}$$

де

$$A = z_1 \times z_2 \times \ldots \times z_i \times \ldots \times z_n, \tag{2}$$

$$B = (1\text{-}z_1) \times (1\text{-}z_2) \times \ldots \times (1\text{-}z_i) \times \ldots \times (1\text{-}z_n), \tag{3}$$

zi – спам-оцінка кожного слова, що входить в лист.

Якщо отримана оцінка менше деякого заздалегідь визначеного граничного значення, то лист трактується як спам.

Очевидно, що ефективність даного методу багато в чому залежить від правильності розрахунку спам-оцінок слів, які входять до листа. Для цього необхідно провести статистичний аналіз як спаму, так і звичайних листів одержуваних кожним користувачем.

Таким чином, метод байєсівської фільтрації передбачає деяке запізнювання, пов'язане з накопиченням кожним користувачем достатнього обсягу статистичного матеріалу (архіву листів). Ще одним недоліком методу є пропуск спаму, якщо в листі щодо мало слів з високою спам-оцінкою.

Відзначимо, що ця обставина використовується спамерами як для обходу, так і для компрометації фільтрів. Наприклад, лист, що складається з набору нейтральних слів не розпізнає як спам.

У більшості сучасних антиспамових систем реалізовані комплексні методи захисту, які по завіренням їх розробників можуть фільтрувати до 98% спаму. Однак час реакції на новий вид спам-листів найбільших поштових служб інтернету становить не більше 20-30 хв.

Концепція фільтрації електронних листів

Сам факт існування досить дорогих масових розсилок електронних листів рекламного характеру свідчить про те, що для багатьох користувачів Інтернету спам представляє великий інтерес. Очевидно, що цей інтерес обумовлений вмістом спам-листів. У той же час користувачі, які не цікавляться запропонованою тематикою, відносяться до спаму негативно. З цих причин основним критерієм фільтрації електронних листів може бути відповідність змісту електронного листа і інтересів користувачів:

$$\begin{cases} \forall P, T \in \{I\} \to C \\ \forall P, T \notin \{I\} \to S \end{cases}, \tag{4}$$

де P – електронний лист, T – тематика електронного листа, $\{I\}$ – множина (область) інтересів користувачів, C – цільовий лист, S – спам.

Виходячи з можливостей потенційних користувачів системи захисту, формування області їх інтересів необхідно реалізувати за допомогою одного або декількох фрагментів тексту на природній мові. Як зазначених фрагментів можуть використовуватися спеціальним чином оброблені цільові листи, а також безпосередньо введений текст. Можливою проблемою реалізації залежності (4) є визначення користувачами системи захисту, всієї області інтересів користувачів електронної пошти. На практиці може виявитися, що навіть кінцевому користувачеві чітко визначити межі цієї області досить важко. При цьому межі області інтересів можуть змінюватися в часі. Тому багато потенційно цікаві листи можуть бути розцінені як спам. Для вирішення даної проблеми розділимо всі електронні листи на три групи: цільові листи, спам і нейтральні листи. До групи підозрілих потраплятимуть ті листи, тематика яких не належить ні інтересам користувачів, ні тематиці спаму. З огляду на запропоновану класифікацію, модифікуємо критерій фільтрації (4):

$$\begin{cases} \forall P, T \in \{I\} \to C \\ \forall P, T \notin \{I\} \wedge T \notin \{Q\} \to F \\ \forall P, T \notin \{I\} \wedge T \in \{Q\} \to S \end{cases}, \tag{5}$$

де F – нейтральний лист, $\{Q\}$ – множина тем спаму.

Тематика спаму

Практичний досвід, а також результати [1,2,3] показують, що спам часто є в основному у текстових листах, які іноді мають графічні файли-вкладення. При цьому основними тематичними напрямками спаму є [1,2,3]:

- Реклама споживчих товарів (R_t). Рекламується реальний товар, і вказуються джерела (посилання на сайт або номер телефону) більш докладної інформації.
- Реклама товарів і послуг "для дорослих" (R_p).
- Реклама програмного забезпечення і комп'ютерів (R_k).

- Реклама туристичних компаній, що пропонують різні види відпочинку і подорожей (R_o).

- Запрошення на семінари і тренінги (R_{st}).

- Послуги по електронній рекламі (R_{er}).

- Платні дзвінки. Рекламується товар і / або послуга і вказується номер телефону, дзвінки на який є платними (R_z).

- Розкрутка сайту. Лист містить інформацію з метою заохотити користувачів відвідати певний сайт (R_w).

- Фінансовий спам. До цього виду спаму відносяться листи з рекламою різного виду фінансових пірамід, пропозиції зробити певну інвестицію або реклама покупки акцій (R_f).

- Збір інформації. Одержувачу пропонують заповнити анкету і відіслати дані за вказаною адресою (R_i).

- Політичні або PR-акції. Цей вид спаму характерний в періоди загострення політичної обстановки (R_{pr}).

- Поширення троянських програм. При відкритті листа активізується програма типу троянський кінь, яка виконує деякі несанкціоновані дії, наприклад, збирає і відсилає зловмисникові необхідну інформацію з комп'ютера (W_t).

- Фішинг. Це поширення підроблених повідомлень від імені банків / фінансових компаній. Метою такого повідомлення є несанкціонований збір ідентифікаційних даних (паролів, пін-кодів, логінів) користувачів. Зазвичай такий спам змушує користувача ввести свої ідентифікаційні дані (W_f).

Тестові розсилки. Найчастіше є порожні листи (T_p), листи з декількома словами (T_s) або з певним набором символів (T_b). Такі розсилки переслідують відразу кілька цілей. З одного боку, це звичайне тестування нового або модифікованого спамерського програмного забезпечення. З іншого боку, листи таких розсилок досить часто проходять антиспам-фільтри (не містять спамерського контенту), викликаючи у користувачів недовіру до захисту від спаму. Модифікуємо (5), з урахуванням поширених тем спаму:

$$\begin{cases} \forall P,T \in \{I\} \to C \\ \forall P,T \notin \{I\} \wedge T \notin \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \to F \\ \forall P,T \notin \{I\} \wedge T \in \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \to S \\ \{R_t,R_p,R_k,R_o,R_{st},R_{er},R_z,R_w,R_f,R_i,R_{pr},W_t,W_f,T_p,T_s,T_b,N\} \in Q \end{cases}, \quad (6)$$

де N – спам-листи з тематикою, яка не належить ні до однієї з вище перерахованих поширених тем спаму.

Завдання – визначення відповідності змісту електронних листів з інтересами користувачів або з тематикою спаму.

На наш погляд зазначене завдання може бути віднесена до класу задач спілкування людини з обчислювальною системою на природній мові. В даний час, незважаючи на значні успіхи ця проблема далека від вирішення. Тому пошук рішення слід обмежити, порівнявши з існуючими можливостями методик розуміння тексту і потребами системи захисту від спаму. Слід враховувати, що система розпізнавання не обов'язково повинна зрозуміти зміст тексту електронного листа, інтереси користувача і тематики спаму. Завдання полягає в тому, що б порівняти формальний опис змісту зазначених текстів і віднести електронного листа до одного з заздалегідь відомих класів.

Загальна структура домена антиспамової обробки

Компоненти цієї основи включають в себе об'єкт антиспамової обробки, підоб'єкти антиспамової обробки, сервери електронної пошти та клієнтів електронної пошти. Ці компоненти можуть зв'язуватися один з одним за допомогою загальнодоступних протоколів передачі повідомлень (табл. 1).

Об'єкт антиспамової обробки отримує повідомлення від під об'єктів антиспамової обробки і приносить їм нові правила. Під об'єкти антиспамової обробки повинні перевіряти період дії правил, що надходять від об'єкта антиспамової обробки, і вносити в них поліпшення. Клієнт електронної пошти є об'єктом, з яким безпосередньо взаємодіють користувачі. Сервер електронної пошти здійснює доставку електронної пошти в мережі електрозв'язку на базі IP.

Таблиця 1

Компоненти домена аниспамової обробки

DNS	Domain Name Server	Система найменувань доменів
E-mail	Electronic mail	Електронна пошта
ESMTP	Extended Simple Mail Transfer Protocol	Розширений простий протокол передачі електронної пошти
FTP	File Transfer Protocol	Протокол передачі файлів
HTTP	Hypertext Transfer Protocol	Протокол передачі гіпертексту
IMAP4	Internet Message Access Protocol v4	Протокол доступу до повідомлень інтернету, версія 4
IP	Internet Protocol	протокол Інтернет
POP3	Post Office Protocol v3	Поштовий протокол, версія 3
RBL	The term is commonly used to describe Real-time Blacklist	Цей термін зазвичай використовується для опису "чорного списку" в реальному масштабі часу
SASL	Simple Authentication and Security Layer	Рівень простий аутентифікації і безпеки
SMTP	Simple Mail Transfer Protocol	Простий протокол передачі електронної пошти
URL	Uniform Resource Locator	Уніфікований покажчик ресурсу

Клієнт електронної пошти направляє скарги подоб'екти антиспамової обробки. У конкретних ситуаціях клієнт електронної пошти може подати скаргу безпосередньо об'єкту антиспамової обробки верхнього рівня.

Функції домену антиспамової обробки

Функції клієнта електронної пошти

• Крім виконання загальних функцій передачі електронних повідомлень, клієнт електронної пошти забезпечує механізм, що допомагає користувачам надсилати скарги про спамові інформації об'єкту антиспамової обробки. Одержувачам електронної пошти потрібно визначити, чи є та чи інша електронне повідомлення спамом, виходячи з його змісту, назви або адреси.

• Клієнт електронної пошти може завантажувати правила, фільтруючі спам, автоматично з об'єкта антиспамової обробки. Правила фільтрації встановлюються згідно з повідомленнями про скарги, що надходять від клієнтів електронної пошти. Вони включають граничний розмір одного електронного повідомлення, кількість електронних повідомлень, що направляються за певний період часу, ключові слова в основному тексті електронних повідомлень і т. Д.

• Клієнт електронної пошти може переслати спам, що розсилається по електронній пошті, об'єкту антиспамової обробки для подальшої обробки або видалення деяких правил фільтрації, що викликають помилкове спрацьовування. Об'єкт антиспамової обробки може негайно оновити правила фільтрації відповідно до вимог або скаргами від клієнта електронної пошти.

• Клієнт електронної пошти може безпосередньо відфільтрувати спам, що розсилається по електронній пошті. Зазвичай одержувачі повинні знати про результати фільтрації, з тим щоб не допустити виникнення проблеми помилкового спрацьовування.

Функції сервера електронної пошти.

• Здійснюючи загальні функції передачі електронної пошти, сервер електронної пошти виконує свої звичайні дії з обміну електронною поштою з іншим сервером електронної пошти або з відправлення та одержання електронної пошти між клієнтами електронної пошти; в той же час сервер електронної пошти повинен заборонити функцію відкритої ретрансляції, з тим щоб спамери не змогли змусити його передати спам-повідомлення іншого сервера електронної пошти.

• Будь-який абонент повинен пройти перевірку, перш ніж він направить електронне повідомлення через сервер електронної пошти. Різні системи електронної пошти можуть використовувати різні механізми перевірки. Перевірка проводиться між сервером електронної пошти і клієнтом електронної пошти.

• Будь-який постачальник послуг електронної пошти може вести "чорний список" спамерів, в якому міститься деяка інформація про спамерів (наприклад, найменування хоста, найменування домену або адресу електронної пошти). Сервер електронної пошти відмовляється отримувати електронні повідомлення, які виходять від цих спамерів.

• Сервер електронної пошти може повернути команду перевірки джерела, який вказаний в інформації про відправника електронного повідомлення.

• Деякі команди SMTP можуть використовуватися спамерами, для того щоб вгадати дійсну обліковий запис сервера електронної пошти. Сервер електронної пошти забороняє ці команди, наприклад EXPN і VRFY.

• Деякі види електронних повідомлень рекламного та пропагандистського характеру направляються без надання будь б то не було інформації про відправника. Сервер електронної пошти повинен автоматично додати посилання HTTP в текст електронного повідомлення.

• Сервери електронної пошти виявляють спам-повідомлення за допомогою антиспамових технологій, повідомляють про спам подоб'єкти антиспамової обробки і завантажують з нього правила фільтрації.

• У разі виявлення спаму сервер електронної пошти повинен здійснити резервне копіювання вихідного спаму, що включає щонайменше заголовок електронної пошти джерела, і представити його в фільтр.

• Сервер електронної пошти повинен надавати інформацію системного журналу і свої статистичні дані, які періодично копіюються, і передавати їх подоб'єкти антиспамової обробки.

• Сервер електронної пошти повертає інший номер стану відповідно до іншими правилами.

• Сервер електронної пошти може обмежити обсяг трафіку, що направляється конкретним абонентом електронної пошти.

Функції об'єкта антиспамової обробки.

• Обмін правилами фільтрації з іншими об'єктами антиспамової обробки. Для передачі інформації можуть використовуватися різні протоколи, наприклад FTP і HTTP.

• Зберігання вихідної інформації про спам-повідомленнях, отриманих від абонентів, і об'єкта антиспамової обробки.

• Широкомовна передача правил фільтрації подоб'єкти антиспамової обробки і попередження їх про небезпечні електронних повідомленнях.

• Об'єкт антиспамової обробки повинен керувати правилами фільтрації і підтримувати їх. Ці правила можуть бути отримані через веб-сайт для:

• Отримання повідомлень від абонентів і подоб'єктов антиспамової обробки;

• Широкомовної передачі достовірної інформації, в тому числі інформації, що стосується контролю і управління.

Функції підоб'єкту антиспамової обробки.

• Отримання повідомлень про скарги від абонентів і правил фільтрації від об'єкта антиспамової обробки.

• Зберігання вихідної інформації про спам (або заголовків спаму), отриманої від абонентів і різних організацій.

• Широкомовна передача правил фільтрації серверів електронної пошти або клієнтам електронної пошти та попередження про небезпечні електронні повідомлення будь-якого користувача.

• Відстеження розповсюдження спаму і збір відповідної інформації.

• Повідомлення про стан поширення спаму і передача відповідної інформації об'єктів на більш високих рівнях.

• Створення нових правил фільтрації з резервних копій сумнівних електронних повідомлень, перевірка і зміна діючих правил фільтрації. Ці правила можуть бути отримані через веб-сайт для:

- створення повідомлень про спам від абонентів і серверів електронної пошти;
- створення нових правил фільтрації.

Висновки

Провівши аналіз ми прийшли до висновку що для захисту електронної пошти від спаму і фішингу звичайних, теоретичних правил безпеки (перевірка листів навчання працівників) недостатньо. Для максимального захисту даних необхідно звертатися за допомогою спеціалізованого програмного забезпечення, платного чи безкоштовного, наприклад, антивірусів. Ефективність підходу щодо побудови баєсівського класифікатора полягає в тому, що його можливо використовувати у якості концепції для побудови систем захисту у корпоративній інформаційній системі від спам повідомлень та сучасних загроз.

Перелік посилань

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

5. Спам // Знаймо разом. – Режим доступу: http://znaimo.com.ua/Спам.

6. Text Classification using Naive Bayes [Електронний ресурс] – 2015. – Режим доступу: http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2blearnnote07-2up.pdf

7. The importance of neutral examples for learning sentiment [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Rita McCue, Jonathan Schler – 21.10.2005 – http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.9735 77

8. Too much information [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Режим доступу: https://www.economist.com/node/18895468

Надійшла: 02.09.2022 Рецензент: д.т.н., професор Вишнівський В.В.