**УДК 004.056.5**

**О.П. Козевич**
НДІ Національного університету “Львівська політехніка”

# КОНТЕНТНА ФІЛЬТРАЦІЯ – ТЕХНОЛОГІЯ КОМПЛЕКСНОГО КОНТРОЛЮ ІНТЕРНЕТ-РЕСУРСІВ. ОСНОВНІ ПІДХОДИ І ПРОБЛЕМИ



**Розглянуто проблеми, пов’язані з використанням і контролем інтернет-ресурсів. Проаналізовано підходи, які дають змогу вирішити ці проблеми та оцінено їх ефективність.**

**Ключові слова: фільтрація, мережа, інтернет-ресурс, фішинг.**

**The paper discusses issues related to the use and control of Internet resources. The approaches that allow solving these issues are analysed and their efficacy is appraised.**

**Key words: filtration, network, internet resources, fishing.**

Сьогодні важко переоцінити значення Інтернету в життєдіяльності людей. З кожним днем цей сервіс розширюється і стає основним джерелом інформації для мільйонів людей, а також бізнес-інструментом для багатьох компаній світу.

І хоча фактично його основна функція не змінилася з часів його створення, але інфраструктура, що використовується для вирішення цих завдань, значно змінилася. Якщо спочатку Інтернет був лише об’єднанням певної кількості мереж, то сьогодні це багатокомпонентна павутина, що охоплює весь світ. При цьому кардинально зросла швидкість передавання даних, підвищилися надійність і завадостійкість їх пересилання. І, як звичайно це трапляється, кількісні зміни переросли в якісні.

Інтернет став не лише основним джерелом інформації для багатьох людей, які не уявляють собі життя без нього, але й невід’ємною частиною багатьох бізнес-процесів.

Найбільш показовими, але не єдиними прикладами є електронна комерція (E-commerce) і Інтернет-банкінг (Internet-banking). Отже, Інтернет де-факто став робочим інструментом, без якого вже неможливо уявити наше повсякденне життя.

Однак природа Інтернету має й інший бік. Його стихійне неконтрольоване поширення забезпечило масовий характер його використання, але породило й низку глобальних проблем із серйозними наслідками. Яз зазначено в [1], головними з них є:

Перше: оскільки Інтернет є каналом у зовнішній світ, він став основним джерелом поширення шкідливого мобільного коду (вірусів, хробаків, троянських програм тощо).

Друге: глобальна мережа стали використовуватися як канал, за допомогою якого здійснюють атаки на локальні обчислювальні мережі організацій, окремі сервери і комп'ютери. Багато інтернет-ресурсів містять різний програмний код: JavaScript, Flash, ActiveX та ін. Зловмисники можуть експлуатувати цей код для організації атак на корпоративні мережі і користувацькі робочі місця.

Сучасна ІТ-структура піддається великій кількості атак, найбільш актуальними з яких є:

- Фішінг (Fishing) – способи перехоплення паролей, номерів кредитних карт і т.п. за допомогою технік соціальної інженерії;
- Spyware & Malware – засоби перехоплення даних і встановлення контролю над комп’ютером;
- Віруси та інші шкідливі коди;

- SPAM/SPIM – небажані повідомлення, що засмічують електронну пошту;
- Витік бізнес-інформації, що може нанести компанії непоправну шкоду;
- Загроза судового переслідування, пов’язана із неправомірним використанням інформації, яка захищена авторським правом.

Третє: Інтернет стали активно використовувати як засіб прихованого проникнення в корпоративні локальні обчислювальні мережі.

Четверте: сьогодні Інтернет можна розглядати як один з основних каналів витоку конфіденційної інформації.

Наприклад, інформаційні ресурси компаній піддаються серйозним загрозам з використанням співробітниками цих компаній безкоштовних поштових скриньок. За статистикою, в середньому 80 % співробітників різних компаній, крім внутрішніх корпоративних поштових адрес, активно використовують безкоштовні поштові скриньки, надані різними провайдерами (yandex.ru, hotmail.com, mail.ru, mail.yahoo.com і т.і.). Маючи доступ до Інтернету зі свого робочого місця і знаючи, що канал не контролюється, будь-який користувач може безперешкодно відправити за межі організації будь-яку конфіденційну інформацію. Так, згідно з [11], найпоширенішим шляхом витоку інформації є електронна пошта (див. рис. 1). Її частка найбільша – 22 %. Потім йдуть відповідно Інтернет (сайти, чати, форуми, безкоштовні поштові сервіси) – 20 %, інтернет-пейджери і мобільні накопичувачі – по 19 % та інші джерела.

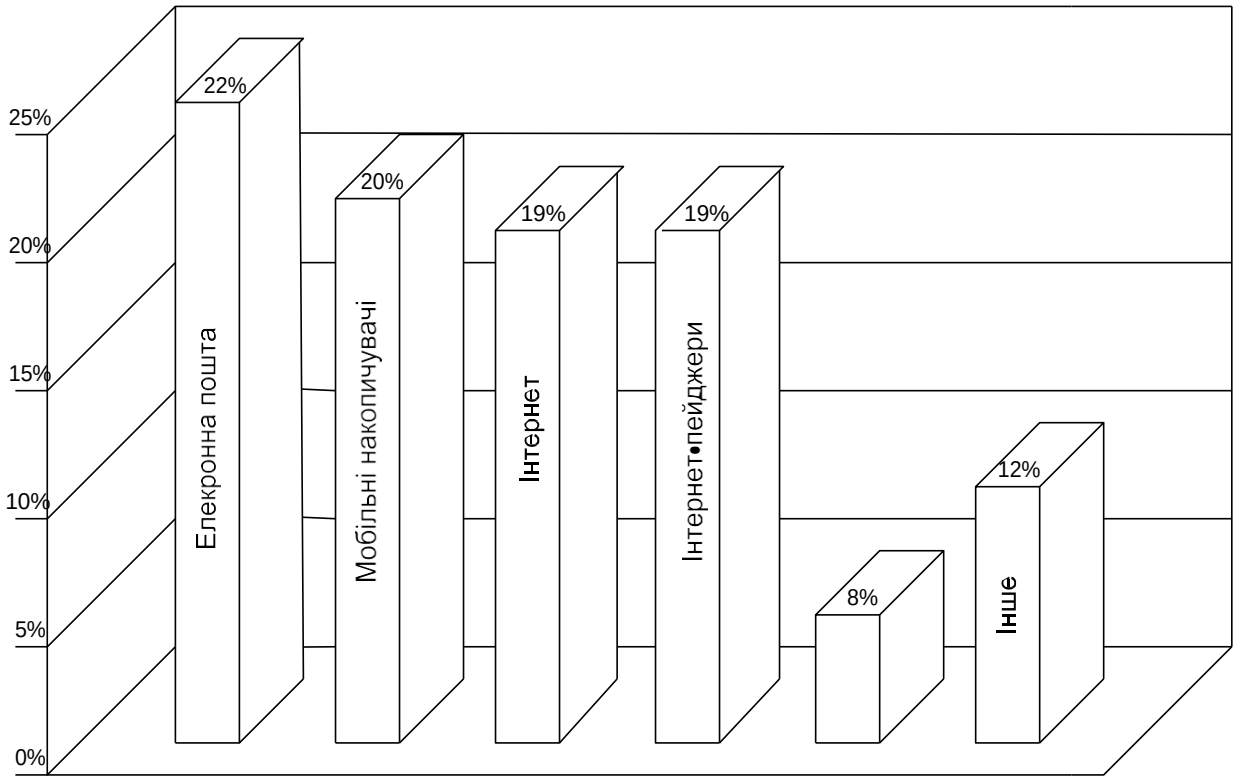


*Рис. 1. Шляхи витоку конфіденційної інформації*

Але навіть розуміючи це, лише кожна четверта компанія забороняє використовувати такі скриньки.

П'яте: безконтрольний доступ до Інтернету значно знижує продуктивність праці в колективі. Простота освоєння, легкість пошуку необхідної інформації та інші корисні якості Інтернету – ось причини популярності цього сервісу, зокрема з особистою метою. Не секрет, що у багатьох вже давно з'явилася звичка починати робочий день з читання новин, прогнозів погоди тощо. За даними компанії IDC, близько третини свого робочого часу працівники різних організацій і компаній

проводять в Інтернеті в цілях, що не мають безпосереднього стосунку до їхньої роботи. Це і "походи" у інтернет-магазини, і мережеві ігри, і просто пошук інформації.

Нарешті, ще один наслідок неконтрольованого використання Інтернету – це зниження пропускної здатності мережі. Згідно зі статистикою, 44 % співробітників організацій використовують корпоративні ресурси для проглядання відео, прослуховування аудіозаписів (через потокові аудіо- і відеоканали), грають в мережеві ігри, завантажують файли великого об'єму (наприклад, файли мультимедіа: графічні, музичні файли, фільми тощо), що створює значне навантаження на локальні обчислювальні мережі.

Цілком логічно після цих зауваг постає питання, що ж робити, щоб вберегтися від зазначених загроз. Тут автор цілком поділяє думку, викладену в роботі [5].

Однозначна відповідь для кінцевого користувача – робити не потрібно нічого. З нинішнім рівнем організації атак, з використанням так званих "платформ нападу", що поєднують троянські програми, бот-програми, різного роду черв'яки, DoS-атак (наприклад ШТОРМ) кінцевому користувачу справитись неможливо. Спроби окремих користувачів справитися з цією навалою низькопродуктивні, та й кваліфікація середньостатистичного користувача переважно недостатня.

Загалом проблему безпечного і продуктивного використання Інтернет-ресурсів можна вирішувати двома способами. Перший – радикальна заборона використання Інтернету без необхідності за принципом "заборонено все, що явно не дозволено".

Другий, гнучкіший варіант, що дає змогу користувачам діяти за принципом "дозволено все, що не заборонено".

Сьогодні перший спосіб ще використовують в організаціях, де циркулює секретна інформація й інформація з обмеженим доступом. І хоча він не гарантує від витоку інформації, але суттєво обмежує такі можливості.

Більшість же комерційних підприємств і компаній віддають перевагу гнучкішому способу регламентації спілкування із зовнішнім світом.

Щоб забезпечити гнучкий контроль при використанні інтернет-ресурсів, компаніям необхідно провадити відповідну політику використання цих ресурсів. Цю політику можна реалізувати ручним методом, коли відповідний штат співробітників у режимі реального часу провадять моніторинг користувачів або автоматично, за допомогою відповідних засобів.

Зрозуміло і очевидно, що ручний моніторинг вимагає значних трудозатрат для відслідковування активності користувачів і, крім того, потребує категоризації сайтів, які відвідують користувачі.

Щоб уникнути цих проблем і забезпечити гнучкий контроль використання інтернет-ресурсів, компанії повинні мати необхідний інструмент для реалізації політики використання ресурсів.

Об'єктивна потреба викликала появу на ринку великої кількості програмних продуктів, призначених для захисту від наведених вище загроз. Однак практика показала, що найефективнішими з них є засоби контролю інформаційного обміну, котрі використовують технологію контентної фільтрації від англійського "content filtering". У публікаціях зустрічаються й інші назви цієї технології – "фільтрація за вмістом", "технологія контролю вмісту інформаційного обміну" тощо, всі вони означають одне і те саме.

Не вдаючись в деталі точних визначень, зазначимо, що під контентом розуміють власне наповнення сайту, а в ширшому сенсі будь-які дані в електронному вигляді. За своєю суттю контент може бути явним і неявним, і власне неявний контент становить найбільшу загрозу.

Необхідно одразу зазначити, що жодна автоматична система не здатна на 100 % гарантувати безпеку без дійової участі людини в процесі фільтрації. Зрозуміло, що кваліфікація персоналу в цьому випадку має визначальне значення.

Крім того, для подальшого коректного аналізу необхідне одне, дуже важливе припущення. Воно передбачає, що засоби контентної фільтрації застосовуються після того, як реалізовано всі

базові заходи із забезпечення безпеки, а саме: прийняття і організація відповідної політики безпеки, корекція згідно з нею корпоративної інфраструктури, встановлення відповідних оновлених операційних систем і прикладних програм, а також виконання інших вимог, що диктуються політикою безпеки.

Тільки після такого впровадження спеціалізованих засобів контентної фільтрації можна отримати бажаний результат.

ІТ-ринок пропонує різні засоби фільтрації вмісту інформаційного обміну інтернет-каналами. Їх можна умовно класифікувати за типами і методами фільтрації.

Сьогодні відомі три типи засобів, що забезпечують контроль за використанням інтернет-ресурсів. До першого типу належать міжмережеві екрани, системи виявлення вторгнень, проксі-сервери, маршрутизатори і схожі на них засоби фільтрації. Другий тип – це сучасні антивірусні програми, що мають базові можливості контентної фільтрації.

До третього типу належать спеціалізовані засоби, розроблені безпосередньо для контролю за використанням інтернет-ресурсів: системи моніторингу електронної пошти, засоби контролю Web-трафіку, антиспам-фільтри, антишпигунські програми тощо.

Кожний тип засобів контролю призначений для використання на різних рівнях мережевої ієрархії. Засоби першого і другого типів не призначені для контролю інформаційного вмісту під час передавання інформації в інтернет-мережі. Вони фільтрують трафік на мережевому і транспортному рівнях, тоді як засоби третього рівня здійснюють її на прикладному рівні. Для детальнішого ознайомлення із засобами першого і другого типу можна звернутися до [2].

На рис. 2 наведено класифікацію засобів контролю використання інтернет-ресурсів.

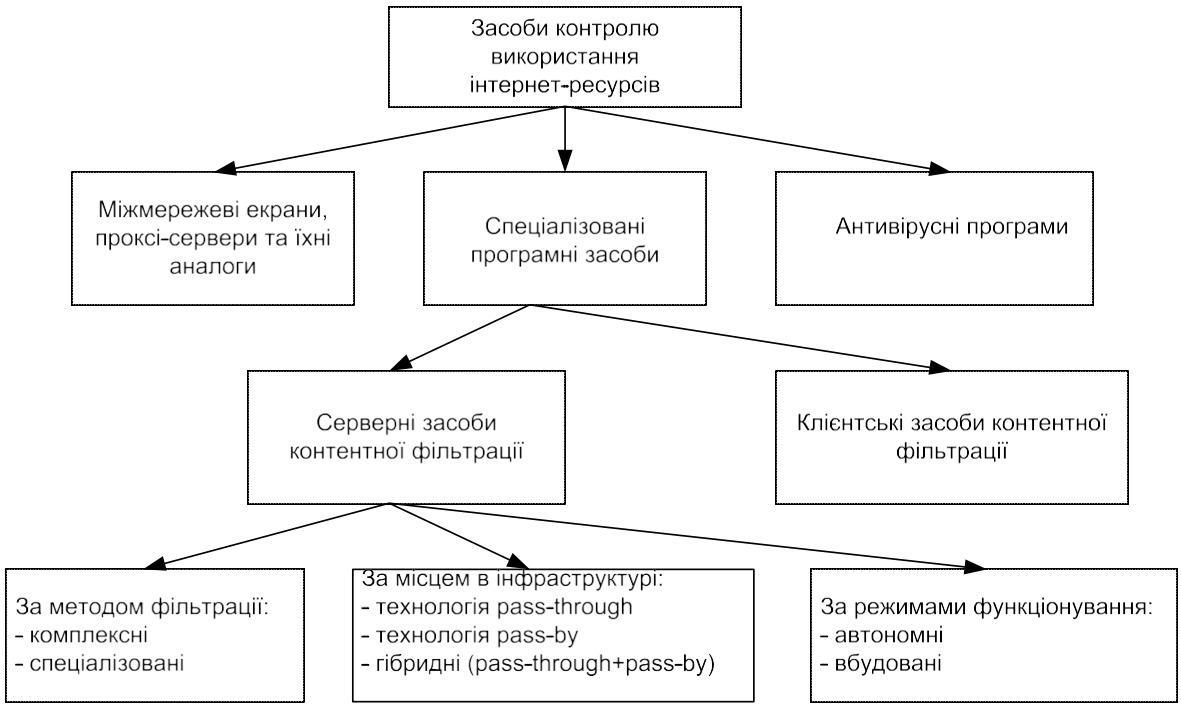


*Рис. 2 Класифікація засобів контролю за використанням Інтернет-ресурсів*

На підприємствах, де активно використовують Інтернет, актуальними є завдання контролю доступу користувачів до інтернет-ресурсів і запобігання витоку конфіденційної інформації, використання засобів першого і другого типів недостатньо.

Це обумовлено недоліками, які детально проаналізовано в [1]. Ми зосередимо свою увагу на засобах третього типу, спеціалізованих програмних засобах, що дають змогу ефективно контролювати інформаційний обмін.

Як ми вже зазначали вище, сучасним загрозам інформаційній безпеці неможливо запобігти без використання контентної фільтрації, в основу якої покладено детальний аналіз поштового і Web-трафіків. Як слушно зазначено в [5], постійне прагнення до латентності комп'ютерних злодіянь, технологічний рівень атак, а також інтелектуальна складова шкідливих програм розвиваються настільки швидко, що для своєї безпеки і безпеки свого бізнесу необхідно як мінімум не програвати, а бажано випереджати рівень розвитку цих технологій. Це означає, що вимагати відповідних знань у цій галузі від диверсифікованих за рівнем кваліфікації користувачів комп'ютерів – це як мінімум невідповідально. Тому клієнтські засоби контентної фільтрації хоч і мають право на життя, але ефективного захисту забезпечити не можуть.

Найбільш адекватним сучасному рівню загроз вирішенням проблеми очищення інформації від шкідливого навантаження є шлюзові (серверні) засоби контентної фільтрації.

Як було зазначено вище, основними джерелами загрози для корпоративних комп'ютерів є Інтернет і електронна пошта. Комплексний підхід до безпеки, що забезпечується засобами контентної фільтрації, передбачає рішення, однаково ефективні для аналізу даних, що надходять як через поштові протоколи, так і через Web.

Зрозумілим також є те, що всі такі засоби повинні мати високу продуктивність, щоби вносити мінімальні затримки в роботу користувачів. Це доволі важливо при виборі обладнання, адже швидкість роботи безпосередньо залежить від технології реалізації і режиму роботи. Наприклад, з використанням контентного фільтра, що працює як проксі-сервер, уповільнення в роботі буде відчутним, адже для аналізу контенту він відзеркалюється на проксі-сервер, який переглядає кожен файл і лише потім доставляє запитуваний контент на робочі місця. Такий підхід до роботи спричиняє дратівливі для користувачів паузи і безпосередньо залежить від пропускної здатності мережі.

Альтернативою проксі-методам є технології, що дають змогу "на льоту" аналізувати файли, передані за допомогою різних протоколів, на наявність шкідливого навантаження, причому на реальній швидкості або ж із незначним обмеженням щодо пропускної здатності мережі. Ці методи мають інші недоліки, які детальніше проаналізуємо нижче.

**Аналіз пошти**

Під час використання електронної пошти виникає необхідність захищати як вхідний, так і вихідний трафік. Але завдання, які вирішуються для кожного з цих напрямів, доволі сильно відрізняються. Для вхідного трафіку необхідно забезпечити контроль за шкідливими кодами (вірусами), фішингом та небажаною поштою (спамом), тоді як у вихідної пошти контролюється вміст, передавання якого може призвести до витоку важливої інформації, поширенню компрометуючих матеріалів тощо.

Більшість продуктів, наявних на ринку, надають можливість контролю лише вхідного трафіку. Це здійснюється за рахунок інтеграції з антивірусними системами, реалізації різних механізмів захисту від небажаної пошти та фішинга. Багато з цих функцій вже вбудовано до поштових клієнтів, але повністю вирішити завдання вони не можуть.

Сьогодні більшість рішень, призначених для захисту вхідного трафіку від небажаної кореспонденції, використовують одну з двох найпоширеніших технологій серед продуктів класу "email security". Одна з них – це використання репутаційного механізму відстеження несанкціонованих інтернет-розсилок, що містять підозрілий контент. Інші технології використовують метод сканування контенту кожного листа, що надходить на робочі станції мережі.

І той, і інший методи мають право на самостійне існування, але загальний вектор рішень для забезпечення безпеки поштового трафіку спрямований у бік комплексного підходу.

Репутаційний механізм відстеження передбачає наявність певної бази даних, яка б забезпечувала перевірку репутації сайту, що надає інформацію.

Системи контролю, що використовують цей метод фільтрації, використовують спеціальний сервіс категоризації, що надається провайдерами інтернет-послуг або спеціалізованими компаніями. Основу сервісу становить база даних, яка постійно оновлюється і поповнюється.

Переважно лише на спеціалізованих заходах розробники таких рішень афішують, як саме вони привласнюють сайту ту чи іншу репутацію і які категорії, критерії та система оцінок для цього застосовуються. Для цього можуть використовуватися і вік сайту (чим він "старший", тим краща в нього репутація), і контент (лінгвістичний аналіз), і наявність адреси в black-листах провайдерів, і багато інших параметрів. Детальнішу інформацію про категоріювання сайтів і розміщених на них даних можна отримати в [6].

Слід зауважити, що цей механізм застосовують для захисту вхідного трафіку від фішингу. Найчастіше цей захист здійснюється порівнянням отриманих поштових повідомлень з наявною базою даних адрес сайтів і повідомлень, про яку йшлося вище. Але не лише для захисту систем від фішингу використовують репутаційний механізм відстеження. Пізніше ми ще повернемося до цього методу захисту.

Переходячи до другого методу аналізу вхідного трафіку і аналізуючи вимоги, які повинні задовольняти системи сканування електронної пошти засобами контентної фільтрації, необхідно зазначити, що аналізувати контент необхідно за всіма компонентами електронної пошти: атрибутами конверту, заголовками повідомлення, MIME-заголовками, тілом самого повідомлення, приєднаними файлами. Системи контролю електронної пошти повинні поділяти поштове повідомлення на складові, причому це повинно бути повний поділ, незалежно від складності листа і рівнів вкладення.

Система повинна також надійно визначати типи файлів-вкладень. Під надійністю ми розуміємо визначення, що ґрунтується не на імені файла чи на інформації, вписаній поштовим клієнтом при додаванні файла (MIME-тип) – така інформація може бути недостовірною в результаті – або свідомих спроб обманути систему контролю, або неправильних налаштувань поштової програми відправника.

Важливе значення для системи контролю має повнота проведених перевірок, тобто кількість і різноманітність критеріїв аналізу. Система повинна мати можливість здійснювати фільтрацію за різними атрибутами, за обсягом повідомлень і вкладених файлів, за кількістю і типом вкладень, а також вміти аналізувати вміст прикріплених файлів незалежно від того, чи є ті файли стиснутими чи архівованими.

Якщо репутаційний метод фільтрації захищає користувача від фішингу, то методи аналізу контенту вхідної кореспонденції повинні захищати користувачів від спаму. Переважно використовують три основні методики визначення належності листа до спаму.

Перша методика виявляє наявність в листі певних ключових слів або словосполучень, характер написання листа, а також специфічної адресної інформації.

Друга методика пов'язана із визначенням адреси відправника та його приналежності до так званих "чорних списків" поштових серверів. До цих списків заносяться сервери або комп'ютери, що зафіксовані в масових розсилках спаму, тобто фактично використовується репутаційний механізм відстеження.

Третя методика ґрунтується на теорії ймовірності і використовує для фільтрації спаму статистичний алгоритм Байєса. Сьогодні поштові сервери для фільтрації спаму використовують фільтри DSPAM, SpamAssassin, SpamBayes, SpamProbe, Bogofilter, CRM114, які застосовують байєсівські методи. За наявними оцінками, цей метод є доволі ефективним. Досліджені 8 тисяч

листів, половина з яких була спамом, дала можливість виявити 99,5 % спам-повідомлень, а кількість помилкових спрацювань фільтра виявилась нульовою [6].

Ефективне виявлення шкідливих кодів (вірусів), фішингу та небажаної пошти (спаму) – це дуже важливо, але не менш важливе значення в сучасних системах контролю електронної пошти має гнучкість реагування системи на результат аналізу. Система повинна вміти блокувати (цілковито або тимчасово) доставку кореспонденції, переміщувати підозрілі листи в карантинну зону для подальшого аналізу, надсилати повідомлення адміністратору мережі тощо.

Як зазначено вище, важливо контролювати вихідну пошту, передавання якого може призвести до витоку важливої інформації, поширення компрометуючих матеріалів тощо.

Для боротьби з витоками інформації використовують різні способи, що ґрунтуються на перехопленні і глибинному аналізі повідомлень відповідно до прийнятої політики фільтрації. У цьому випадку виникає необхідність коректно визначити тип файлів, мову і методи кодування тексту, здійснити семантичний аналіз вихідних повідомлень.

Одним із способів такої фільтрації може бути обмеження за типом переданих повідомлень. Наприклад, можна встановити заборону на відправку за межі організації файлів Microsoft Word або EXEL, які можуть містити важливу інформацію. При цьому тип файла не повинен визначатися іменем файла чи основуватись на інформації, вписаній поштовим клієнтом під час додавання файла (MIME-тип). Для точної обробки даних система фільтрації повинна визначати тип файла за його сигнатурою.

Останнім часом для забезпечення безпеки інформаційних систем важливим став такий фактор, як наявність у компанії архіву поштових повідомлень. Нові розробки у галузі контентної фільтрації передбачають наявність спеціальних модулів архівування. Створення архіву, це не просто автоматичний запис поштових повідомлень, але здатність системи аналізувати їх зміст протягом всього життєвого циклу повідомлень, можливість отримання необхідних статистичних даних з архіву, створених з використанням різних критеріїв.

**Аналіз Web-трафіка**

Що стосується фільтрації Web-трафіка, то для забезпечення інформаційної безпеки можуть бути використані різні підходи.

Так, наприклад, у роботах [4, 5] викладені два постулати, які, на думку авторів, повинні підвищити рівень інформаційної безпеки.

Перший з них – це максимальне зниження ролі людського фактора на вплив політики безпеки роботи в Інтернеті. Небезпека тут полягає в тому, що непідготовлені користувачі можуть приймати рішення з питань, в яких вони не компетентні.

Другий – найбільшу увагу під час фільтрації Web-трафіка необхідно звернути на боротьбу із шпигунським (spyware) програмним забезпеченням.

У роботі [1] сповідується ширший підхід до засобів фільтрації Web-трафіка. Основним критерієм вибору засобів фільтрації Web-трафіка, на думку авторів, є функціональність, тобто залежність від завдань, які стоять перед тією чи іншою компанією. В роботі детально проаналізовано функції, які повинні виконувати засоби фільтрації Web-трафіка, а також способи реалізації цих функцій.

На наш погляд, при фільтрації як пошти, так і Web-трафіка завдання інформаційної безпеки певною мірою переплітаються. Так, при фільтрації Web-трафіка актуальними є запобігання витоку конфіденційної інформації, моніторинг підозрілої і забороненої активності користувачів, захист від атак з використанням засобів соціальної інженерії (фішинг/фармінг), захист від вірусів і спаму, а також захист від дії шпигунського програмного забезпечення.

Як бачимо, завдання значною мірою є спільними для обох видів фільтрації, і в тому немає нічого дивного, оскільки цілі фільтрації збігаються.

Оскільки ми вже приділили багато уваги завданням, які виконуються під час фільтрації пошти, розглянемо детальніше способи захисту від дії шпигунського програмного забезпечення. Класичним методом боротьби із spyware є сигнатурний аналіз. Цей метод непоганий, але ефективність його застосування при сучасному розвитку загроз є недостатньо високою з таких причин [4]:

1. Для визначення spyware за сигнатурою необхідно контролювати 100 % Web-трафіка за протоколами HTTP, FTP і дуже бажано – за протоколом HTTPS. Це достатньо ресурсомістка задача, і, як ми вже зазначали, розв'язати її без затримок у роботі користувачів можуть небагато компаній.
2. База сигнатур повинна бути актуальною, інакше нові штами spyware виявлені не будуть.
3. Найефективніше spyware виявляється при тотальному скануванні файлової структури на актуальній базі сигнатур. Однак таке сканування займає неприпустимо багато часу.

Як слушно зауважено в [5], класичний сигнатурний аналіз, що ґрунтується на постійно поновлювальній базі, має один значний недолік: інтервал часу, протягом якого сигнатура потрапляє в базу, може бути доволі великим. Оскільки швидкість поширення шкідливого ПЗ постійно зростає, то база сигнатур відстає за часом, що приводить до можливості проникнення шкідливих програм на комп'ютери користувачів.

Ефективно боротися з такого роду атаками необхідно комплексними методами, які розглянемо в наступних публікаціях.

Звичайно, в короткій статті неможливо розглянути всі проблеми, що існують в сфері безпеки IT. За межами статті залишилися такі важливі питання, як фільтрація трафіка в середовищі Web 2.0, що застосовують технологію AJAX, використання якої серйозно ускладнило контентну фільтрацію Web-трафіка. Також не розглянуто вельми актуальне питання фільтрації VoIP (Voice over IP)-трафіка, засобів для передавання звукової і відеоінформації між комп'ютерами.

Це надзвичайно важливі і актуальні завдання, і в наступних публікаціях ми детально проаналізуємо способи захисту інформаційного середовища із використанням зазначених технологій.

*1. Слепов О. Контроль использования интернет-ресурсов / О. Слепов // Информационный бюллетень "JET INFO"/ – 2005. – № 2(141). – С.2–20. 2. Слепов О. Контентная фильтрация / О. Слепов // Информационный бюллетень "JET INFO"/ – 2005. – № 10(149). – С.3–35. 3. Отт А. О контентной фильтрации / А. Отт // Информационный бюллетень "JET INFO" / – 2005. – № 10(161). – С.3–23. 4. Бычек В. "Шпиону" вход заказан / В. Бычек // "Information Security". – 2008. – № 2. – С.50–52. 5. Комарова Н. Технология комплексной безопасности интернет-контента / Н. Комарова // "Information Security". – 2008. – № 4. – С.32–33. 6. Отт А. Современные тенденции в области контентной фильтрации / А. Отт // Информационный бюллетень "JET INFO"/ – 2012. – С.3–23.*