УДК 681.3.06

А.О. Антонюк, В.С. Портяной, В.П. Шилін

ТЕХНОЛОГІЇ ЗАХИСТУ КОНФІДЕНЦІЙНОЇ ІНФОРМАЦІЇ ВІД ВНУТРІШНІХ ЗАГРОЗ

Проведено огляд методів та програмних засобів захисту конфіденційної інформації, зокрема, Information Protection and Control (IPC) – технології захисту конфіденційної інформації від внутрішніх загроз. Розглянуто методи контролю технічних каналів витоку інформації за допомогою технологій Data Loss Prevention (DLP). Детально проаналізовано методи детектування конфіденційної інформації. Подано опис одного з розповсюджених програмних продуктів сімейства InfoWatch.

Вступ

За статистикою левову долю витоків інформації складають витоки, що пов'язані з IT-інфраструктурою організації, тобто внутрішні загрози. Як правило, під внутрішньою загрозою мається на увазі навмисний інсайд, коли один із співробітників свідомо виносить корпоративний секрет назовні. Велику небезпеку представляють також ненавмисні внутрішні інциденти.

З концептуальної точки зору існує два канали витоку – по-перше, це глобальна мережа і, по-друге, мобільні пристрої. Декомпозицію можна продовжити і далі, розділивши канал Internet на підканали: електронна пошта, ftp і p2p-мережі; а канал мобільні пристрої – на підканали – ноутбуки, флеш-накопичувачі й т. д. Звичайно, кожен витік має власну причину. Поняття «причина» і «канал» взаємозв'язані, проте ототожнювати їх не можна. Як правило, причина витоку дозволяє припустити, який канал при цьому використовувався і навпаки. Причини витоків бувають зовнішніми і внутрішніми. Такий розподіл є умовним, оскільки деякі інциденти (назвемо їх змішаними) мають і ту й іншу причину. До змішаних витоків можна віднести, наприклад, хакерське вторгнення з елементами соціальної інженерії або атаку з боку колишнього співробітника [1]. Як протистояти таким порушенням безпеки?

Саме для захисту від подібних витоків інформації сьогодні широке розповсюдження отримали спеціалізовані продукти, до яких, зокрема, слід віднести Information Protection and Control (IPC) – технологію захисту конфіденційної інформації від внутрішніх загроз [2]. Рішення класу IPC призначені для захисту інформації від внутрішніх загроз, запобігання різних видів витоків інформації, корпоративного шпигунства і бізнес-розвідки.

IPC-система поєднує у собі дві основні технології:

- шифрування носіїв інформації у всіх точках мережі;
- контроль технічних каналів витоку інформації за допомогою технологій Data Loss Prevention (DLP).

Контроль доступу до мережі, програм та даних є можливою третьою технологією у системах класу IPC.

Технологія IPC є логічним продовженням технології DLP і дозволяє захищати дані не тільки від витоку технічними каналами, тобто інсайдерів, а й від несанкціонованого доступу користувачів до мережі, інформації, додатків і в тих випадках, коли безпосередній носій інформації потрапляє у руки третіх осіб. Це дозволяє не допускати витоку і в тих випадках, коли інсайдер або не має легального доступу до даних або людина отримує доступ до безпосереднього носія інформації (втрата, крадіжка або вилучення (наприклад, недобросовісними конкурентами або рейдерами)). Далі більш детально розглядається саме технологія IPC.

Завдання

Основним завданням IPC-систем є запобігання передачі конфіденційної

інформації за межі корпоративної інформаційної системи. Така передача (витік) може бути навмисною або ненавмисною. Практика показує, що більша частина (понад 75%) витоків відбувається не зі злого наміру, а через помилки, неуважність, недолугості, недбалості працівників – виявляти подібні випадки набагато простіше. Інша частина пов'язана зі злим умислом операторів і користувачів інформаційних систем підприємства, зокрема, промисловим шпигунством, конкурентної розвідкою. Очевидно, що зловмисні інсайдери, як правило, намагаються обдурити аналізатори IPC та інші системи контролю.

Додаткові завдання систем класу IPC:

- запобігання передачі зовні не тільки конфіденційної, а й іншої небажаної інформації (образливих виразів, спаму, еротики, зайвих обсягів даних і т. п.);
- запобігання передачі небажаної інформації не тільки зсередини назовні, а й зовні всередину інформаційної системи організації;
- запобігання використання працівниками Internet-ресурсів і ресурсів мережі в особистих цілях;
- захист від спаму;
- захист від вірусів;
- оптимізація завантаження каналів, зменшення нецільового трафіку;
- облік робочого часу і присутності на робочому місці;
- відстеження благонадійності співробітників, їх політичних поглядів, переконань, збір компромату;
- архівація інформації на випадок випадкового видалення або псування оригіналу;
- захист від випадкового або навмисного порушення внутрішніх нормативів;
- забезпечення відповідності стандартів у галузі інформаційної безпеки і чинного законодавства.

Як було раніше сказано, IPC включає у себе рішення класу DLP, яка в IPC підтримує контроль наступних технічних каналів витоку конфіденційної інформації:

- електронна пошта;
- Веб-пошта;
- соціальні мережі й блоги;
- файлообмінні мережі;
- форуми та інші інтернет-ресурси, у тому числі виконані на AJAX-технології;
- IM (ICQ, агент Mail.ru, Skype, AOL AIM, Google Talks, Yahoo Messenger, MSN та інше);
- p2p-клієнти;
- периферійні пристрої (USB, LPT, COM, WiFi, Bluetooth й інше);
- локальні та мережеві принтери.

Технології DLP в IPC підтримують контроль, у тому числі наступних протоколів обміну даними:

- HTTP;
- HTTPS (SSL);
- FTP;
- FTP-over-HTTP;
- FTPS;
- SMTP.

Крім того, відзначимо, що обов'язковою компонентою IPC є архів, який ведеться для обраних потоків інформації (пакетів, повідомлень). Вся інформація про дії співробітників зберігається в одній або декількох пов'язаних між собою базах даних. Лідируючі IPC-системи дозволяють архівувати всі канали витоку, які вони можуть контролювати. В архіві IPC зберігаються копії закачаних в Internet документів і текстів, електронних листів, роздрукованих документів і файлів, записаних на периферійні пристрої. У будь-який момент адміністратор ІБ може отримати доступ до будь-якого документа або тексту в архіві, використовуючи лінгвістичний пошук інформації за єдиним архівом (або за всіма розподіленими архівами водночас). Будь-які повідомлення за необхідності можна подивитися або переслати, а будь-який закачаний в Internet, записаний або роздрукований на

зовнішній пристрій файл або документ, переглянути або скопіювати. Це дозволяє проводити ретроспективний аналіз можливих витоків.

Технологія IPC також включає в себе можливості щодо шифрування інформації на всіх ключових точках мережі:

- жорсткі диски серверів;
- SAN;
- NAS;
- магнітні стрічки;
- диски CD/DVD/Blue-ray;
- персональні комп'ютери;
- ноутбуки;
- зовнішні пристрої.

Технології IPC використовують різні спільні криптографічні модулі, в тому числі найбільш ефективні алгоритми DES, Triple DES, RC5, RC6, AES, XTS-AES.

IPC – системи мають агенти в усіх ключових точках мережі: сервери, сховища, шлюзи, ПК/ноутбуки, периферійні та мережеві пристрої для користувача. Технології IPC реалізовані для Windows, Linux, Sun Solaris, Novell.

Що таке DLP?

Нині немає чіткого розуміння цього терміну [3]. Дійсно, DLP-системою можна назвати навіть антивірус, оскільки він дозволяє уникнути установки троянів, які посилають інформацію з локального пристрою своїм творцям або замовникам, і програму для блокування USB-портів, оскільки вона бореться з витоками через USB-накопи-чувачі. Проте антивірус і система блокування портів не захищають організацію від витоків – вони лише закривають один з каналів або усувають одну з причин. Теоретично за допомогою подібних «клаптевих» рішень можна запобігти витокам, проте для крупних організацій такий підхід категорично неприйнятний.

Тому під DLP-системою зазвичай розуміють комплексне рішення корпоративного масштабу, яке бореться з витоками для різних каналів і причин. Це визначення є неповним, оскільки під нього цілком підходить гіпотетична система, яка фізично блокує всі порти і перекриває доступ до Internet.

Сучасним організаціям потрібні Internet і доступ до мобільних накопичувачів. Перекривати ці канали не можна – а значить, доступ до них слід контролювати. Іншими словами, DLP-система зобов'язана аналізувати трафік, який проходить по каналах і «голосно кричати» адміністратору, якщо цей трафік виявився секретним. Остання теза є ключовою, такою, що відокремлює функціонал DLP-систем від більшості других рішень з інформаційної безпеки.

У рамках окремого узятого каналу DLP-система працює як «чорний ящик», куди на вхід подається інформація, що йде по каналу, а на виході формується вердикт, чи є вхідна інформація секретною. Даний принцип не залежить від специфіки каналу, яка, втім, може бути використана в алгоритмі винесення вердикту.

Таким чином, основна цінність DLP-системи полягає в алгоритмі, на основі якого працює «чорний ящик». Архітектура і навіть список підтримуваних каналів є вторинними характеристиками даного ПЗ. Проте, з погляду кінцевого замовника, дані чинники також дуже важливі, оскільки без них додаток практично марний.

На даний момент у галузі не існує стандартного алгоритму аналізу трафіку, більш того – не існує навіть стандартної технології. Кожна, представлена на ринку компанія, сповідає власний спосіб фільтрації даних, що йдуть по каналах. У цілому можна говорити про два концептуально різних підходи, що мають як переваги, так і недоліки.

Перший підхід базується на фільтрації контенту, тобто змістовного наповнення інформації. Це означає, наприклад, що при перевірці на секретність стандартних офісних документів у форматі .doc система спочатку переведе їх в текстовий формат, а потім, використовуючи заздалегідь підготовлені дані, винесе по цьому тексту вердикт. Другий підхід (контекстна

фільтрація) використовує принципово іншу схему: замість аналізу контенту система перевіряє контекст, в якому передається інформація, а саме, вилучає метадані файлу, дивиться на його розмір або аналізує поведінку користувача.

Перші системи (їх можна назвати першим поколінням), що з'явилися на ринку, використовували саме алгоритми контентної фільтрації, до яких поступово додавалися функції по роботі з контекстом.

Проте у всіх без виключення методів є один основний недолік – низька точність визначення, оскільки жоден з наявних методів не може забезпечити захист усіх типів конфіденційної інформації. Навіть у рамках одного типу є численні обмеження. Практика показує, що навіть комбінація декілька контентних і контекстних алгоритмів рідко забезпечує прийнятну точність за умов обмеженої продуктивності системи.

За даними компанії Perimetrix ефективність фільтрації із застосуванням контентних технологій досягає лише 80 %. Це означає, що 20 % корпоративних секретів можуть безперешкодно покинути мережу компанії. Крім того, система фільтрації контента допускає помилки другого типу, визнаючи деякі легітимні відомості секретними.

Проте, не дивлячись на очевидні обмеження, контентно-контекстні системи першого покоління, як і раніше, домінують на DLP-ринку. Оскільки спочатку контентний підхід використовувався у більшості систем, більшість традиційних гравців DLP-ринку не наважуються змінити основну технологію, яка розроблялася роками.

Після того, як неефективність класичних контентних методів стала очевидною, деякі компанії почали упроваджувати принципово інший, детерміністський підхід (друге покоління DLP-систем). Такий підхід припускає, що всі конфіденційні файли мають бути спеціальним чином помічені. Розмітка файлів близька до контекстної фільтрації (а мітки близькі до метаданих), проте насправді ці підходи принципово різні. У детерміністському підході мітки спеціально вбудовуються в документ самою DLP-системою, тоді як класичний контекстний аналіз оперує незалежним від DLP-системи контекстом.

Технологічно мітка може вбудовуватися в документ по-різному. В деяких продуктах ця мітка може бути «вшита» в ім'я файлу. В цьому випадку кожне ім'я файлу має починатися, наприклад, з класу конфіденційності або рівня секретності, що приводить до створення корпоративних політик іменування файлів. Природно, на практиці це дуже незручно, не прозоро і заважає співробітникам ефективно працювати.

Тим часом є більш тонкі методи роботи з документами, які не припускають таке грубе вбудовування. Наприклад, мітка може бути інкапсульована всередину файлу, скажімо, в один з його службових заголовків. Коли такий документ покидає корпоративну мережу через мережеві канали, наприклад, електронною поштою, фільтру не слід аналізувати контент. Досить лише проаналізувати мітку та застосувати положення політики безпеки. Іншими словами, знаючи мітку, система захисту може безпомилково визначити, є файл секретним або відкритим.

Очевидно, що для впровадження детерміністської системи слід провести повну класифікацію всіх електронних документів в організації і помітити всі секретні файли відповідним чином. Також зрозуміло, що ефективність захисту помічених файлів дорівнює 100 % (звичайно, за умови захисту від несанкціонованої зміни).

Головний недолік детерміністських методів полягає у тому, що помітити всі конфіденційні документи, як правило, неможливо. Ще важче постійно підтримувати базу помічених документів у актуальному стані. Щоб вирішити цю проблему, застосовуються різні способи. Наприклад, можна переносити мітки в нові файли із старих документів і таким чином істотно спростити керування платформою.

У цілому детерміністські методи мають право на існування, проте в більшості випадків вони також неефективні.

Сьогодні сучасний ринок потребує нових рішень, що належать до класу DLP-систем третього покоління. Можливим варіантом тут є використання комплексного підходу, а саме, детерміністських методів у разі помічених документів і контентної фільтрації для решти файлів. При цьому поєднуються плюси обох підходів: точність і гнучкість фільтрації контента. За даними компанії Perimetrix ефективність такого поєднання досягає 99,6 %.

Проте проста інтеграція підходів не може бути підставою для появи чергового покоління продуктів. Розвиток галузі показав, що концептуально нові DLP-системи мають підтримувати функціонал шифрування для захисту інформації на мобільних носіях. Жоден з традиційних DLP-підходів не може забезпечити захист від крадіжки ноутбука, а цей захист украй необхідний для повноцінної DLP-системи.

Практично половина (49 %) сучасних витоків відбувається в результаті крадіжки мобільних пристроїв. Єдиним способом захисту тоді є шифрування – решта всіх підходів (паролі, фізичний захист і так далі) давно показала власну неспроможність. Проблема полягає лише в тому, що переважна більшість рішень класу DLP не може забезпечити шифрування, а рішення по шифруванню не можуть здійснювати фільтрацію витікаючого трафіку.

Підприємствам і організаціям необхідний уніфікований функціонал для того, щоб:

- отримати одну систему захисту від всіх загроз витоку інформації, а не дві;
- використовувати єдині інтегровані політики для фільтрації і шифрування документів;
- забезпечити просту і швидку відповідність різним нормативним актам і стандартам.

Технології детектування конфіденційної інформації

Діяльність наведених технологій базується на технології детектування конфіденційної інформації [4]. Розглянемо найбільш розвинуті та розповсюджені технології детектування.

Сигнатури. Найпростіший метод контролю – пошук у потоці даних певної послідовності символів. Іноді заборонену послідовність символів називають «стоп-виразом», але в більш загальному випадку вона може бути представлена не словом, а довільним набором символів, наприклад, певною міткою. Якщо система настроєна тільки на одне слово, то результат її роботи – визначення 100 % збігу, тобто метод можна віднести до детерміністського. Однак частіше пошук певної послідовності символів все ж таки застосовують при аналізі тексту. В переважній більшості випадків сигнатурні системи налаштовані на пошук декількох слів і частоту зустрічальності термінів.

До переваг цього методу можна віднести простоту поповнення словника заборонених термінів і очевидність принципу роботи, а також те, що це найбільш надійний спосіб, якщо необхідно знайти відповідність слова або виразу на 100 %.

Недоліки стають очевидними після початку промислового використання такої технології при визначенні витоків і налаштуванні правил фільтрації. Більшість виробників DLP-систем працюють для західних ринків, а англійська мова дуже «сигнатурна» – форми слів найчастіше утворюються за допомогою прийменників без зміни самого слова. В російській мові, наприклад, все набагато складніше, тому що у ній є приставки, закінчення, суфікси. Для прикладу можна взяти слово «ключ», яке може означати як «ключ шифрування», «ключ від квартири», «джерело», «ключ або PIN-код від кредитної картки», так і безліч інших значень. У російській мові з кореня «ключ» можна утворити кілька десятків різних слів. Це означає, що якщо на заході фахівця із захисту інформації від інсайдерів досить ввести одне слово, в Росії фахівцеві доведеться вводити пару десятків слів і потім ще змінювати їх в шести різ-

них кодуваннях. Реальне застосування цього методу вимагає наявності лінгвіста або команди лінгвістів як на етапі впровадження, так і в процесі експлуатації та оновлення бази. Безсумнівним недоліком є і те, що «сигнатури» нестійкі до примітивного кодування, наприклад, заміною символів на схожі за зображенням.

«Цифрові відбитки» (Digital Fingerprints або DG). Різного типу хеш-функції зразків конфіденційних документів позиціонуються західними розробниками DLP-систем як нове слово на ринку захисту від витоків, хоча сама технологія існує з 70-х років. На заході цей метод іноді називається «digital fingerprints». Суть всіх методів одна й та сама, хоча конкретні алгоритми у кожного виробника можуть відрізнятися. Деякі алгоритми навіть патентуються, що допомагає у просуванні «нової запатентованої технології DG». Загальний сценарій дії такий: набирається база зразків конфіденційних документів. Суть роботи DG досить проста і часто цим і приваблює: DLP/IPC-системі передається якийсь стандартний документ-шаблон, з нього створюється «цифровий відбиток» і записується в базу даних DF. Далі в правилах контентної фільтрації настроюється процентна відповідність шаблону з бази. Наприклад, якщо налаштувати 75 % відповідності «цифровому відбитку» договору поставки, то при контентній фільтрації DLP виявить практично всі договори цієї форми. Іноді, до цієї технології відносять і системи на зразок «антиплагіат», однак остання працює тільки з текстовою інформацією, водночас як технологія «цифрових відбитків», у залежності від реалізації, може працювати і різним медійним контентом і застосовуватися для захисту авторських прав і перешкоди випадкового або навмисного порушення законів і нормативів інформаційної безпеки.

До переваг технології «цифрових відбитків» (Digital Fingerprints) можна віднести простоту додавання нових шаблонів, досить високий ступінь детектування і прозорість алгоритму технології для співробітників підрозділів по захисту інформації. Спеціалістам СБ і ІБ не треба думати про «стоп-вирази» та іншу лінгвістику, витрачати багато часу на аналіз потенційно небезпечних словоформ і вбивати їх в базу, витрачати ресурси на впровадження та підтримку лінгвістичної бази.

Основним недоліком, який на перший погляд неочевидний і схований за «патентованими технологіями», є те, що, незважаючи на всю простоту і фактичну відсутність лінгвістичних методів, необхідно постійно оновлювати базу даних «цифрових відбитків». І якщо у випадку з «сигнатурами», такий метод не вимагає постійного оновлення бази словами, то він вимагає оновлення бази «цифрових відбитків». До недоліків «цифрових відбитків» можна віднести те, що фактично від «розширення бази словами» підтримка DLP в ефективному стані переходить на «пошук та індексування нових і змінених файлів», що є більш складним завданням, навіть якщо це робиться DLP-системою напівавтоматично. Великі компанії, в яких з'являється до десятка тисяч нових і оновлених документів кожен робочий день тільки на серверних сховищах часто просто не в змозі відслідковувати все це в режимі реального часу, не кажучи вже про персональні комп'ютери і ноутбуки. У такому разі застосування DG малоефективне, тому «цифрові відбитки» в більшості DLP розраховані на компанії SMB-сектора (менш 500 користувачів). На додаток до цього цифрові відбитки займають приблизно 10–15 % від розміру конфіденційних документів, і база постійно розростається, що вимагає додаткових інвестицій у збільшення систем зберігання інформації і продуктивність DLP-серверів. Крім того, низькорівневі хеш-функції (у тому числі й DG) нестійкі до примітивного кодування, що розглядалося щодо «сигнатур».

«Мітки». Суть цього методу полягає у призначенні спеціальних «міток» всередині файлів, що містять конфіденційну інформацію. З одного боку, такий метод дає стабільні та максимально точні відомості для DLP-системи, з іншого – потрібно досить сильні зміни в інфраструктурі ме-

режі. У лідерів DLP/IPC-ринку реалізація даного методу не зустрічається, тому розглядати її докладно не має особливого сенсу. Можна лише зауважити, що, незважаючи на явне достоїнство «міток» – якість детектування, є багато суттєвих недоліків: від необхідності значної перебудови інфраструктури всередині мережі до введення безлічі нових правил і форматів файлів для користувачів. Фактично впровадження такої технології перетворюється у впровадження спрощеної системи документообігу.

Регулярні вирази. Пошук за регулярними виразами («масками») є також давно відомим способом детектування необхідного вмісту, однак в DLP він став застосовуватися відносно недавно. Часто цей метод називають «текстовими ідентифікаторами». Регулярні вирази дозволяють знаходити збіги за формою даних, у ньому не можна точно зазначити точне значення даних, на відміну від «сигнатур». Такий метод детектування ефективний для пошуку:

- ІПН;
- КПП;
- номерів рахунків, кредитних карт, телефонів, паспортів, клієнтських номерів.

До переваг технології регулярних виразів у першу чергу варто віднести те, що вони дозволяють детектувати специфічний для кожної організації тип контенту, починаючи від кредитних карток і закінчуючи назвами схем обладнання, специфічних для кожної компанії. Крім того, форми основних конфіденційних даних змінюються вкрай рідко, тому їх підтримка практично не вимагатиме часових ресурсів. До недоліків регулярних виразів можна віднести їх обмежену сферу застосування в рамках DLP/IPC-систем, так як знайти за допомогою них можна тільки конфіденційну інформацію лише певної форми. Регулярні вирази не можуть застосовуватися незалежно від інших технологій, однак можуть ефективно доповнювати їх можливості.

Лінгвістичні методи (морфологія, стеммінг). Найбільш поширеним на сьогоднішній день методом аналізу в DLP/IPC-системах є лінгвістичний аналіз тексту. Він настільки популярний, що часто саме він у просторіччі іменується «тематичною фільтрацією», тобто несе на собі характеристику всього класу методів аналізу вмісту. Мовознавство як наука складається з багатьох дисциплін – від морфології до семантики, і лінгвістичні методи аналізу різняться між собою. Є технології, які використовують лише «стоп-вирази», що вводять тільки на рівні коріння, а сама система вже становить повний словник; є що базуються на розставленні ваг на терміни, що найчастіше зустрічаються в тексті. Є у лінгвістичних методах і свої відбитки, що базуються на статистиці; наприклад, береться документ, рахується п'ятдесят найбільш уживаних слів, потім вибирається з 10 найуживаніших з них у кожному абзаці. Такий «словник» є практично унікальною характеристикою тексту і дозволяє знаходити в «клони» значущі цитати. Аналіз всіх тонкощів лінгвістичного аналізу не входить до рамок цієї статті, однак необхідно зауважити ширину можливостей цієї технології у рамках IPC-систем.

До достоїнств лінгвістичних методів у DLP можна віднести те, що в морфології та інших лінгвістичних методах високий ступінь ефективності, порівняно з сигнатурами, при набагато менших трудовитратах на впровадження і підтримку (зниження трудовитрат на 95 % за відношенням до «сигнатур»). При цьому у випадку з використанням лінгвістичних методів детектування немає необхідності відстежувати появу нових документів і направляти їх на аналіз у IPC-систему, так як ефективність лінгвістичних методів визначення конфіденційної інформації не залежить від кількості конфіденційних документів, частоти їх появи і продуктивності системи фільтрації вмісту. Недоліки лінгвістичних методів також досить очевидні, перший з них – залежність від мови – якщо організація представлена в декількох країнах, бази конфіденційних слів і виразів доведеться створювати окремо для кожної мови і країни, з огляду на всю специфіку. При цьому

звичайна ефективність такого методу складе в середньому 85 %. Якщо залучати професійних лінгвістів, то ефективність може зрости до 95 % – більше може забезпечити лише ручна перевірка або «сигнатури», проте щодо ефективності і трудовитрат рівних лінгвістичним методам поки не знайшли.

Ручне детектування («Карантин»). Ручна перевірка конфіденційної інформації іноді називається «Карантин». Будь-яка інформація, яка потрапляє під правила ручної перевірки, наприклад, у ній зустрічається слово «ключ», потрапляє у консоль фахівця інформаційної безпеки. Останній по черзі вручну переглядає таку інформацію та приймає рішення про пропуск, блокування або затримку даних. Якщо дані блокуються або затримуються, відправнику надсилається відповідне повідомлення. Безперечною перевагою такого методу можна вважати найбільшу ефективність. Проте, такий метод у реальному бізнесі можна застосовувати лише для обмеженого обсягу даних, тому що потрібно велика кількість людських ресурсів, так як для якісного аналізу всієї інформації, що виходить за межі компанії, кількість співробітників інформаційної безпеки має приблизно збігатися з кількістю інших офісних співробітників. А це неможливо навіть у силових і військових структурах. Реальне застосування для такого методу – аналіз даних обраних співробітників, де потрібна більш тонка робота, ніж автоматичний пошук за шаблонами, «цифрових відбитків» або збігів зі словами з бази.

Забезпечення захисту конфіденційної інформації від витоків за допомогою програмних продуктів компанії InfoWatch

Як приклад розглянемо основні властивості сімейства програмних продуктів компанії InfoWatch – одного з поширених продуктів забезпечення захисту конфіденційної інформації третього покоління. Більш детальну інформацію про них можна знайти в [5–6].

Сімейство продуктів InfoWatch являє собою досить збалансований засіб для ефективного керування інформаційною безпекою компанії, без якого немислимо безперервне ведення сучасного бізнесу.

InfoWatch допомагає зберегти репутацію замовника, мінімізувати фінансові ризики, пов'язані із втратою конфіденційності даних і привести інформаційну систему у відповідність із національними й міжнародними законами та стандартами.

Рішення компанії InfoWatch забезпечують контроль над найпоширенішими шляхами витоку, моніторинг доступу співробітників до корпоративних інформаційних ресурсів і зберігання докладного архіву операцій з документами.

InfoWatch у масштабі реального часу фільтрує поштовий і Веб-трафік і контролює операції з документами на робочих станціях, запобігаючи виводу конфіденційних даних за межі інформаційної системи. У випадку виявлення фактів порушення корпоративної політики IT безпеки система оперативно повідомляє про інцидент компетентним особам і поміщає підозрілі об'єкти в область карантину.

InfoWatch Traffic Monitor (далі Traffic Monitor або система) – це розподілена багатокомпонентна система, призначена для контролю за трафіком вихідних листів, переданих за протоколами SMTP і HTTP.

Для виконання даного завдання в системі передбачені два види перехоплювачів:

- Mail Monitor – призначений для перехоплення SMTP-трафіку. Перехоплення SMTP-трафіку здійснюється за допомогою інтеграції з поштовим сервером Postfix;
- Web Monitor – призначений для перехоплення HTTP-трафіка. Поставляється у вигляді окремого компонента – Traffic Monitor ISA Server Plugin. Перехоплення HTTP-трафіка здійснюється за допомогою інтеграції із Proxy-сервером MS ISA Server.

У процесі роботи системи Traffic Monitor накопичується великий обсяг даних. Завдання зберігання отриманих даних вирішується інтеграцією системи Traffic Monitor із СУБД Oracle.

До складу системи Traffic Monitor входять такі компоненти:

- Traffic Monitor Server (сервер поштової фільтрації);
- Traffic Monitor ISA Server Plugin (модуль фільтрації POST-запитів);
- компоненти користувальницького інтерфейсу:
  - Traffic Monitor Security Administrator;
  - Traffic Monitor Security Officer;
  - Traffic Monitor DB Administrator;
  - Traffic Monitor Analyser;
  - Traffic Monitor Eraser.

Traffic Monitor Server (далі Traffic Monitor Server або сервер поштової фільтрації) призначений для виконання функцій:

- здійснення тематичної класифікації листів;
- виділення з потоку електронної кореспонденції листів, що містять ознаки конфіденційної інформації;
- збереження копій листів, що відповідають корпоративній політиці безпеки в архів;
- модифікація заголовків листів, списків одержувачів, відповідно до вимог корпоративної політики безпеки;
- збереження затриманих листів, що порушують корпоративну політику безпеки в карантин (з можливістю відкладеного прийняття Офіцером безпеки рішення про відправлення листа одержувачам);
- створення архіву листів шляхом експорту листів з бази даних у зазначене місце (з можливістю імпорту створеного архіву листів назад у базу даних).

Кожний електронний лист, що проходить через сервер поштової фільтрації, перевіряється на його відповідність корпоративній політиці безпеки, причому виконується перевірка атрибутів листа на відповідність або невідповідність певним умовам, а також аналізується зміст самого листа та вкладених файлів.

Traffic Monitor ISA Server Plugin (далі Traffic Monitor ISA Server Plugin або модуль фільтрації POST-запитів) здійснює перехоплення таких POST-запитів, структура яких дає можливість перетворювати їх в SMTP-листи. До таких POST-запитів належать листи, що відправляються на сайти Web-пошти, форуми й т. ін. З перехоплених POST-запитів формуються SMTP-листи, які перевіряються **сервером контентної фільтрації**, що є частиною **сервера поштової фільтрації**.

Після проходження процедури фільтрації запити, у яких не знайдено ознак порушення корпоративної політики безпеки, пропускаються на зовнішній сервер для відправлення зазначеним адресатам. Відправлення запитів з виявленими ознаками порушень блокуються.

Всі листи, сформовані з перехоплених POST-запитів, перенаправляються в чергу повідомлень (локальне сховище повідомлень сервера поштової фільтрації), що очікують запису в базу даних (архів/карантин).

Фільтрація листів сервером контентної фільтрації здійснюється за умови, що конфігурація зазначеного сервера настроєна належним чином.

Завдання забезпечення необхідної конфігурації сервера контентної фільтрації вирішуються за допомогою програми **Traffic Monitor Security Administrator.** Право на запуск даної програми має користувач, якому призначена роль Адміністратора інформаційної безпеки.

Програма призначена для керування настроюваннями параметрів фільтрації й завантаження даних параметрів на сервер контентної фільтрації.

У процесі роботи Адміністратор інформаційної безпеки вирішує завдання створення бази профілів і бази контентного аналізу, підтримки зазначених баз в актуальному стані, завантаження бази профі-

лів і бази контентного аналізу на сервер контентної фільтрації.

Листи, у яких сервером поштової фільтрації були знайдені ознаки порушення корпоративної політики безпеки, відображаються в програмі **Traffic Monitor Security Officer.** За допомогою даної програми Офіцер безпеки може вирішувати наступні завдання:

- аналіз листів, затриманих сервером контентної фільтрації;
- ухвалення рішення про доставку затриманих листів.

За результатами аналізу Офіцер безпеки робить висновок, чи дійсно лист містить ознаки конфіденційної інформації або мало місце помилкове спрацьовування сервера контентної фільтрації. Залежно від винесеного вердикту Офіцер безпеки може приймати відповідні рішення.

Програма **Traffic Monitor DB Administrator** призначена для виконання ряду завдань адміністрування системи. Право на роботу з даною програмою має користувач, якому призначена роль **Адміністратора сховища.** Основні функції Адміністратора сховища:

- розподіл ролей користувачів;
- настроювання ротації сегментів.

Усі облікові записи створюються, редагуються та видаляються Адміністратором сховища, який контролює стан облікових записів і при необхідності може заблокувати обліковий запис користувача. Крім того, якщо в СУБД Oracle включений аудит входу користувачів у систему, то Адміністратор сховища може переглядати дані журналу аудита через інтерфейс програми Traffic Monitor DB Administrator.

Настроювання ротації сегментів здійснюється Адміністратором сховища з метою побудови безперервного циклу прийому й обробки листів. Усі листи надходять у сегмент даних, де потім обробляються для подальшого використання. Сегмент даних – це логічна частина БД. У кожний момент часу над сегментом даних можна робити тільки одну операцію: або прийом нових листів, або переіндексацію отриманих листів. Адміністратор сховища задає періодичність, з якої виконується ротація сегментів.

Програма **Traffic Monitor Analyzer** призначена для аналізу листів, що зберігаються в базі даних. Установка даної програми виконується опціонально. Право на роботу з даною програмою має користувач, якому призначена роль **Аналітика.** Основні завдання Аналітика:

- пошук листів по реквізитах і контексту;
- перегляд і збереження листів;
- видалення листів з бази даних.

Для пошуку листів у програмі передбачені функції роботи з пошуковими запитами: створення нового пошукового запиту, редагування та видалення існуючих пошукових запитів.

Також завданням програми Traffic Monitor Analyzer є забезпечення можливості перегляду та збереження знайдених листів.

Існує можливість звільнення місця на жорсткому диску шляхом видалення тих листів, подальше зберігання яких у базі даних не потрібно. Для виконання даного завдання призначена програма **Traffic Monitor Eraser**.

Працювати із програмою Traffic Monitor Eraser може користувач, якому призначена роль **Чистильника**, основною функцією якого є видалення листів з бази даних.

Рішення InfoWatch забезпечують ефективний контроль і аудит стану інфраструктури внутрішньої IT-безпеки організації. Завдяки багаторівневому моніторингу дій користувачів InfoWatch дозволяє створити комплексний захист конфіденційної інформації проти навмисних і необережних дій персоналу. Реалізація такої стратегії допомагає протистояти промисловому шпигунству та внутрішньому саботажу, мінімізувати операційні ризики, пов'язані із втратою конфіденційності даних.

Висновки

Наведений аналіз сучасних технологій захисту конфіденційної інформації

від витоків дозволяє зробити наступні висновки:

- на сьогодні не існує надійних і ефективних методів захисту конфіденційної інформації від витоків;
- жоден з наведених методів детектування конфіденційної інформації не може вважатися цілком надійним;
- розвиток технологій захисту інформації від витоків має здійснюватися в напрямку більш детальної спеціалізації;
- розробка засобів захисту конфіденційної інформації має здійснюватися комплексно;
- найважливішою складовою системи захисту конфіденційної інформації від витоків слід вважати наявність якісної політики безпеки.

1. 2006 CSI/FBI Computer Crime and Security Survey http://www.infowatch.ru/threats?chapter=147151396&id=2926721
2. *Внутренние* ИТ-угрозы в России 2005 http://www.infowatch.ru/downloads/docs/iw2005.pdf
3. *Cyber* Cop Scanner http://www.nss.co.uk/grouptests/va/edition2/nai_cybercop_scanner/nai_cybercop_scanner.htm
4. *World* Wide Digital Security http://www.pcworld.com/article/id,143371-c,privacysecurity/article.html
5. *Сімейство* продуктів InfoWatch http://www.infowatch.ru/solution
6. *Traffic* Monitor http://www.infowatch.ru/solution?chapter=204274952

Отримано 30.03.2010

Про авторів:

Антонюк Анатолій Олександрович,
кандидат фізико-математичних наук,
доцент кафедри інтелектуальних систем прийняття рішень,
Портяной Володимир Семенович,
головний конструктор,
Шилін Володимир Петрович,
старший науковий спеціаліст.

Місце роботи авторів:

Національний університет державної податкової служби України,
08201, м. Ірпінь, Київської області
вул. К. Маркса, 31.
Тел.: (04597) 53220
e-mail: tolik__@ukr.net

Інститут програмних систем
НАН України,
03187, Київ-187,
Проспект Академіка Глушкова, 40.
Тел.: (044) 526 4579
e-mail: v.portyanoy@isofts.kiev.ua
e-mail: shilin@isofts.kiev.ua