

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



VŨ VĂN TUẤN

**NGHIÊN CỨU VÀ ỨNG DỤNG
HỆ THỐNG CHỐNG THẮT THOÁT DỮ LIỆU**

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN

MÃ SỐ: 60.48.01.04

LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. VŨ VĂN THỎA

HÀ NỘI - 2016

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Vũ Văn Thỏa

Phản biện 1:

Phản biện 2:

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại
Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ... giờ ngày tháng năm

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Tính cấp thiết của đề tài

Một nghiên cứu của IDC (International Data Corporation) chỉ ra rằng hầu hết các trường hợp mất mát dữ liệu đều do sơ ý, chứ không phải do mã độc gây ra. Họ đã ước tính rằng khoảng 80% những trường hợp mất mát dữ liệu như vậy là do vô tình.

Chính vì vậy, nghiên cứu, ứng dụng các giải pháp ngăn ngừa mất mát, rò rỉ thông tin là thực sự cần thiết. Do đó, học viên lựa chọn nghiên cứu đề tài: “Nghiên cứu và ứng dụng hệ thống chống thất thoát dữ liệu”.

Tổng quan về vấn đề nghiên cứu

Hiện nay, nhiều doanh nghiệp Việt Nam có thể cung cấp và triển khai các gói giải pháp ngăn ngừa thất thoát, rò rỉ dữ liệu dựa một số sản phẩm thương mại hóa của nước ngoài như giải pháp chống thất thoát, rò rỉ thông tin/dữ liệu DLP của Check Point, giải pháp Symantec Data Loss Prevention hoặc McAfee Data Loss Prevention Endpoint.

Mục đích nghiên cứu

- Nghiên cứu hệ thống chống thất thoát dữ liệu Data Loss Prevention (DLP).
- Triển khai thử nghiệm hệ thống DLP ngăn ngừa mất mát rò rỉ thông tin và đánh giá khả năng của hệ thống.

Đối tượng nghiên cứu

- Các thuật toán phân loại dữ liệu.
- Hệ thống ngăn ngừa mất mát rò rỉ thông tin DLP.

Phạm vi nghiên cứu

- Nghiên cứu các nguy cơ rò rỉ thông tin và giải pháp phòng chống.
- Triển khai thử nghiệm hệ thống thất thoát dữ liệu DLP trên môi trường giả lập và môi trường thực.

Phương pháp nghiên cứu

Phương pháp nghiên cứu lý thuyết

- Đọc tài liệu và nghiên cứu về các giải pháp chống mất mát dữ liệu - DLP (Data lost prevention) từ các hãng trên thế giới và cách thức hoạt động của chúng.

Phương pháp thực nghiệm

- Xây dựng hệ thống McAfee Data Loss Prevention Endpoint và đồng thời thử nghiệm, đánh giá kết quả.

Cấu trúc luận văn

Nội dung luận văn gồm 2 phần như sau

1. Phần mở đầu

2. Phần nội dung: bao gồm ba chương

Chương 1: Tổng quan về an toàn và bảo mật thông tin.

Chương 2: Hệ thống chống thất thoát dữ liệu DLP.

Chương 3: Triển khai hệ thống thử nghiệm.

Mỗi chương sẽ có phần kết luận riêng của từng chương đó.

Chương 1 : TỔNG QUAN VỀ AN TOÀN VÀ BẢO MẬT THÔNG TIN

1.1. Nội dung an toàn và bảo mật thông tin

An toàn thông tin bao gồm các nội dung sau:

- Tính bí mật: tính kín đáo riêng tư của thông tin.
- Tính xác thực của thông tin, bao gồm xác thực đối tác (bài toán nhận danh), xác thực thông tin trao đổi.
- Tính trách nhiệm: đảm bảo người gửi thông tin không thể thoái thác trách nhiệm về thông tin mà mình đã gửi.

Trước đây, khi nói đến an toàn thông tin người ta thường nghĩ ngay đến những phương thức bảo mật thông tin truyền thống như tường lửa hay phân quyền truy cập tập tin hay thư mục bằng ACL.

Tuy nhiên điều đó không thể ngăn ngừa được rò rỉ, thất thoát dữ liệu khi laptop hay USB bị mất cũng như được truyền qua đường thư điện tử hay thư thoại. Theo báo cáo mới nhất về an ninh toàn cầu của Microsoft thì vấn đề thất thoát thông tin trong doanh nghiệp và chính phủ ngày càng phổ biến với mức độ thiệt hại không hề thua kém so với virus hay mã độc.

Rò rỉ, thất thoát dữ liệu luôn là mối lo ngại hàng đầu của các cơ quan tổ chức. Nó có thể là thể phá hỏng quy trình kinh doanh hoặc vi phạm các chính sách bảo mật của công ty hay tổ chức.

1.2. Các nguy cơ thất thoát, rò rỉ dữ liệu

Để cung cấp một cái nhìn tổng quan hơn về an toàn và bảo mật thông tin, học viên đã phân loại các loại mất dữ liệu như sau: mất dữ liệu do tình cờ, các cuộc tấn công nội bộ và các cuộc tấn công từ bên ngoài.

1.2.1. Mất dữ liệu do tình cờ:

Một nguyên nhân điển hình cho sự mất mát dữ liệu tình cờ là nhân viên không quen thuộc với các chính sách của công ty. Nói cách khác, họ không nhận ra sự nhạy cảm của các tài liệu mà họ đang làm việc với họ hoặc đánh giá quá cao kiến thức của mình về bảo mật máy tính. Một số ví dụ phổ biến:

- Nhân viên tiết lộ thông tin nội bộ ra bên ngoài trong quá trình sử dụng : skype, yahoo, điện thoại...

- Chia sẻ file ngang hàng P2P: Nhân viên có thể dễ dàng sử dụng giao thức P2P để gửi file ra ngoài.

- Có thể do sự vô ý của nhân viên: Nhân viên đính kèm nhầm file, nhân viên chọn sai người, nhân viên bị lừa để gửi thông tin ra ngoài. Hoặc thông qua các dịch vụ cloud storage miễn phí: Nhân viên có thể upload dữ liệu nhạy cảm lên các hệ thống lưu trữ đám mây miễn phí như dropbox hay Skydriver. Nhân viên có thể upload dữ liệu lên một FTP server trên internet để gửi thông tin ra ngoài.

- In tài liệu, photocopy tự do không được quản lý tập trung. Đem tài liệu in, copy ra bên ngoài.

- Dùng điện thoại, camera chụp lại tài liệu của công ty.

- Nhân viên ra ngoài không log off tài khoản hay tắt máy tính để nhân viên khác chép dữ liệu của mình đem ra bên ngoài.

- Thiết bị USB chứa dữ liệu quan trọng bị mất hay để quên tại nơi làm việc.

1.2.2. Mất dữ liệu do tấn công nội bộ

Định nghĩa : Tấn công nội bộ là bất kỳ cuộc tấn công độc hại trên hệ thống của công ty hoặc mạng mà kẻ xâm nhập là một người đã được giao phó với truy cập được phép vào mạng, và cũng có thể có kiến thức về kiến trúc mạng.

Nói một cách khác, nếu một hacker không tìm được cách nào để tấn công vào tổ chức, sự lựa chọn tốt nhất tiếp theo để xâm nhập là thuê một nhân viên, hoặc tìm kiếm một nhân viên đang bất mãn, để làm nội gián, cung cấp các thông tin cần thiết. Đó chính là Insider Attack – tấn công nội bộ. Insider Attack có một thế mạnh rất lớn, vì những gián điệp này được phép truy cập vật lý vào hệ thống công ty, và di chuyển ra vào tự do trong công ty.

Một kiểu khác của tấn công nội bộ, chính là sự bất mãn của nhân viên. Những nhân viên làm việc với mức lương thấp kém, và anh ta muốn có mức lương cao hơn. Bằng cách xâm nhập vào CSDL nhân sự công ty, anh ta có thể thay đổi mức lương của mình. Hoặc một trường hợp khác, nhân viên muốn có nhiều tiền.

1.2.3. Mất dữ liệu do các cuộc tấn công bên ngoài

Các cuộc tấn công bên ngoài là những cuộc tấn công đánh cắp dữ liệu được thực hiện từ xa. Về cơ bản, một ai đó có quyền truy cập vào hệ thống thông qua một kết nối từ xa,

chẳng hạn như internet, và sử dụng truy cập này để ăn cắp dữ liệu, tạo ra botnet hoặc gây ra sự gián đoạn. Động cơ đằng sau các cuộc tấn công chủ yếu của bản chất tài chính.

Hiện nay có rất nhiều hình thức tấn công ăn cắp dữ liệu như :

- Tấn công trực tiếp
- Kỹ thuật đánh lừa (Social Engineering)
- Kỹ thuật tấn công vào vùng ẩn
- Tấn công vào các lỗ hổng bảo mật
- Khai thác tình trạng tràn bộ đệm
- Nghe trộm
- Kỹ thuật giả mạo địa chỉ
- Kỹ thuật chèn mã lệnh
- Tấn công vào hệ thống có cấu hình không an toàn
- Tấn công dùng Cookies
- Thay đổi dữ liệu
- Password-base Attact
- Identity Spoofing

1.3. Các giải pháp an toàn và bảo mật thông tin truyền thống

1.3.1. Các chiến lược an toàn hệ thống :

- Giới hạn quyền hạn tối thiểu (Last Privilege)
- Bảo vệ theo chiều sâu (Defence In Depth)
- Nút thắt (Choke Point)
- Điểm nối yếu nhất (Weakest Link)
- Tính toàn cục
- Tính đa dạng bảo vệ

1.3.2. Các mức bảo vệ trên mạng

Vì không thể có một giải pháp an toàn tuyệt đối nên người ta thường phải sử dụng đồng thời nhiều mức bảo vệ khác nhau tạo thành nhiều hàng rào chắn đối với các hoạt động xâm phạm. Thông thường bao gồm các mức bảo vệ sau:

- Quyền truy nhập

- Đăng ký tên /mật khẩu.
- Mã hoá dữ liệu
- Bảo vệ vật lý
- Tường lửa
- Quản trị mạng

1.3.3. An toàn thông tin bằng mật mã.

Để bảo vệ thông tin trên đường truyền người ta thường biến đổi nó từ dạng nhận thức được sang dạng không nhận thức được trước khi truyền đi trên mạng, quá trình này được gọi là mã hoá thông tin (encryption), ở trạm nhận phải thực hiện quá trình ngược lại, tức là biến đổi thông tin từ dạng không nhận thức được (dữ liệu đã được mã hoá) về dạng nhận thức được (dạng gốc), quá trình này được gọi là giải mã.

1.4. Giải pháp ngăn ngừa mất mát/rò rỉ thông tin theo hướng DLP.

Hiện nay, có ba phương pháp giúp giảm nguy cơ thất thoát thông tin, tuy nhiên cần phải hiểu rõ 3 phương pháp đó trước khi đưa ra quyết định là làm thế nào để có hiệu quả nhất, đỡ tốn kém và dễ triển khai, trong khi vẫn đảm bảo các yêu cầu về tính bảo mật. Ba phương pháp đó là:

1.4.1. Data-in-Motion:

Khái niệm: là phương pháp phát hiện và kiểm soát thông tin lưu chuyển trên mạng như qua mail, web,... phương pháp này còn được gọi là ngăn ngừa mức mạng (Network-based DLP).

1.4.2. Data-in-Use:

Khái niệm: là phương pháp phát hiện và kiểm soát dữ liệu trên máy trạm, máy chủ như copy ra USB, ghi CD/DVD, in trên giấy,... Phương pháp này còn được gọi là ngăn ngừa tại các điểm đầu cuối (Endpoint DLP).

1.4.3 Data-at-Rest:

Khái niệm: là phương pháp phát hiện sự tồn tại của dữ liệu nhạy cảm nằm trong các máy trạm, máy chủ, ổ đĩa cứng, thiết bị đầu cuối,... việc thực thi các chính sách ngăn ngừa mất mát dữ liệu không phải là một kết quả trực tiếp của Data- at- Rest DLP. Các thông tin

thu thập được qua Data- at- Rest DLP có thể được sử dụng để đưa ra một kế hoạch hành động nhằm làm giảm nguy cơ mất dữ liệu.

1.5. Kết luận:

Hiện nay, máy tính đòi hỏi các phương pháp tự động để bảo vệ các tệp và các thông tin lưu trữ. Nhu cầu bảo mật rất lớn và rất đa dạng, có mặt khắp mọi nơi, mọi lúc. Do đó không thể không đề ra các qui trình tự động hỗ trợ bảo đảm an toàn thông tin. Data Loss Prevention là một hệ thống đáp ứng nhu cầu hỗ trợ đảm bảo an toàn thông tin đó.

Chương 2: HỆ THỐNG CHỐNG THẮT THOÁT DỮ LIỆU

Trong chương 1 học viên đã trình bày các nguy cơ và giải pháp liên quan đến phòng chống thất thoát, rò rỉ dữ liệu. Tại chương 2 này, học viên xin trình bày mô hình hệ thống chống thất thoát, rò rỉ dữ liệu DLP.

2.1. Sự xuất hiện của DLP

DLP- Data Loss Prevention là một công cụ ngăn ngừa phòng chống rò rỉ, thất thoát dữ liệu và được gọi dưới nhiều cái tên khác nhau:

- Data Loss Prevention/Protection
- Data Leak Prevention/Protection
- Information Loss Prevention/Protection
- Information Leak Prevention/Protection
- Extrusion Prevention
- Content Monitoring and Filtering
- Content Monitoring and Protection

DLP là một loại công nghệ bảo mật theo hướng bảo vệ dữ liệu nhạy cảm trong một tự động và không xâm phạm. Thông qua chính sách một hệ thống DLP tự động đảm bảo không có dữ liệu nhạy cảm được lưu trữ, gửi hoặc truy cập, trong khi vẫn cho phép người dùng sử dụng các công cụ và dịch vụ mà họ lựa chọn và cần phải hoàn thành nhiệm vụ của họ.

2.2. Mô hình, tính năng của hệ thống

Đặc điểm cơ bản của giải pháp DLP là :

- Đảm bảo nội dung nhạy cảm ở khu vực “rest” trước nguy cơ rủi ro
- Bảo vệ dữ liệu khỏi các rủi ro được biết đến trong quá trình truyền
- Thu thập những luồng thông tin đi trong mạng để xây dựng những rule và xây dựng phòng chống trước các nguy cơ rủi ro phát sinh.
- Thi hành các hành động khắc phục hậu quả như tự động mã hóa trong quá trình quét dữ liệu hoặc những hành động mà người dùng thực hiện

2.2.1. Quét nội dung

Trước khi có giải pháp DLP, các tổ chức thường xuyên phải đối mặt với một tình huống mà các dữ liệu kỹ thuật số của họ được lan truyền trên nhiều địa điểm không có

quyền kiểm soát dữ liệu nhạy cảm. DLP áp dụng chính sách quét và phát hiện nội dung được để khám phá nơi đặt tập tin nhạy cảm. Đây có thể là cơ sở dữ liệu, trên các tập tin chia sẻ, lưu trữ nội bộ trên máy tính xách tay và máy trạm... Công việc này được thực hiện giống như một máy quét chống virus, nhưng thay vì tìm kiếm virus và malware, nó sẽ tìm các tài liệu nhạy cảm và ghi lại vị trí của chúng. Từ đó, những nhà quản trị và quản lý kết quả có thể quyết định làm thế nào đối với các tập tin.

Mặc dù phát hiện nội dung thường được sử dụng khi ban đầu triển khai DLP, chạy quét thường xuyên không phải là hiếm. Tuy nhiên việc này có thể chiếm tài nguyên hệ thống và cần có thời gian để hoàn thành, vì vậy không nên được chạy trong giờ làm việc [3].

2.2.2. Endpoint Protection

DLP bảo vệ thiết bị đầu cuối được cài đặt trên các máy trạm và các thiết bị khác trong các hình thức của một đại lý. Các đại lý thực hiện các chính sách bằng cách giám sát tất cả các hoạt động dữ liệu và quét tất cả các tập tin được lưu trữ tại đó. Thông thường các đại lý cũng kiểm tra và cho phép đầu vào vật lý được kết nối. Điều này có nghĩa là một quản trị viên trung ương có thể vô hiệu hóa USB, FireWire và kiểm soát giao diện khác một cách dễ dàng. Ngoài ra, hành động ghi đĩa CD hoặc DVD có thể được ngăn chặn.

Thiết bị đầu cuối DLP tính năng khác nhau của bảo vệ có thể được phân loại như sau:

- Hệ thống bảo vệ tập tin: Theo dõi tất cả các hoạt động tập tin để bảo vệ thời gian thực tương tự như trong các phần mềm chống virus. Điều này là để đảm bảo các file không được sao chép đến các địa điểm không được phép hoặc mã hóa cũng được áp dụng bởi các DLP khi lưu dữ liệu để biết địa điểm. Quét cũng có thể được thực hiện trên dữ liệu được lưu trữ để phát hiện hành vi vi phạm chính sách.

- Bảo vệ mạng: Theo dõi dữ liệu được truyền qua mạng khi các thiết bị đầu cuối là đi từ các mạng công ty. Nếu không, DLP mạng chịu trách nhiệm cho các chức năng này.

- Bảo vệ ứng dụng: DLP tích hợp trong hệ điều hành và các ứng dụng để ngăn chặn các hành động như việc sao chép vào clipboard, chụp ảnh chụp màn hình hoặc gõ dữ liệu nhạy cảm vào các chương trình chat.

2.2.3. Giám sát mạng

Các DLP mạng có thể hoạt động ở hai chế độ khác nhau: thụ động và chủ động. Chế độ thụ động DLP kiểm tra lưu lượng mạng và các bản ghi có bất kỳ sự vi phạm chính sách nào, trong khi ở chế độ chủ động DLP cũng có thể chặn bất kỳ gói dữ liệu liên quan đến việc vi phạm chính sách. Sử dụng chế độ nào là phụ thuộc vào yêu cầu của tổ chức.

Các vị trí của DLP thường là một vị trí trong mạng mà nó có thể chặn dữ liệu ra trong mạng cục bộ. Đây có thể là nơi dữ liệu đi đến các mạng kém an toàn khác trong cùng một công ty, các trang web từ xa kết nối với VPN, hoặc là internet.

Các kênh được kiểm tra bởi DLP mạng có thể khác nhau đối với các hãng. HTTP, FTP và e-mail dịch vụ là phổ biến nhất. Ngoài ra, các giao thức nhắn tin nhanh, HTTPS và nhiều dịch vụ chia sẻ file cũng thường theo dõi trong các sản phẩm này.

2.2.4. Quản lý trung ương

Một số chức năng chính của các máy chủ quản lý DLP bao gồm:

- Endpoint triển khai đại lý và quản lý đại lý nhiệm vụ bao gồm việc triển khai chính sách, phần mềm và đăng nhập bộ sưu tập.
- Cập nhật phần mềm, định nghĩa thông thường và thông tin cấp phép từ máy chủ của nhà cung cấp.
- Thu thập các bản ghi từ các dịch vụ khác, chẳng hạn như thu thập tập tin và DLPS mạng, và giữ cho các dịch vụ này được cập nhật với các chính sách và các bản vá lỗi phần mềm.
- Chuyển tiếp cảnh báo quan trọng đối với các quản trị viên hệ thống.
- Tạo ra các báo cáo dựa trên dữ liệu thu thập được.
- Cung cấp các công cụ để tạo và quản lý các chính sách DLP.

Nhiều điểm trong số những điểm nêu trên thường có thể được truy cập và quản lý thông qua một giao diện quản trị. Đây thường là một ứng dụng web có thể truy cập từ bất kỳ trình duyệt, mặc dù giao diện dòng lệnh cũng tồn tại.

2.3. Công nghệ cốt lõi của hệ thống

2.3.1. Policies

Trung tâm DLP là các Policy. Nếu không có các Policy sẽ không có sự phân biệt giữa dữ liệu công cộng và nhạy cảm. Policy có thể được tạo ra dựa vào các tổ chức sở hữu thông số kỹ thuật, cũng như yêu cầu bên ngoài.

Tạo các Policy là một trong số ít các công việc trong một triển khai DLP có liên quan đến toàn bộ công ty và không chỉ các bộ phận CNTT. Ở giai đoạn này, điều quan trọng là nhìn vào các chính sách Policy hiện hành và thảo luận với những người xử lý dữ liệu công ty trên làm thế nào để phân loại đúng, xác định và bảo vệ các dữ liệu này. Những chính sách này sau đó được chuyển đổi thành quy tắc mà các DLP có thể thực thi trong khi hoạt động.

Tập đoàn RSA, tập đoàn chuyên cung cấp các chiến lược, giải pháp bảo mật đề nghị để tạo nên một chính sách tốt cần trả lời những câu hỏi sau đây [6]:

- Ai là đối tượng chính sách sẽ áp dụng và làm thế nào nó ảnh hưởng đến họ?
- Những loại thông tin bạn đang cố gắng để bảo vệ?
- Tại sao bạn bảo vệ nó?
- Trường hợp bạn nên bảo vệ nó? Là các dữ liệu chuyển động hoặc trong một trung tâm dữ liệu? Là nó đang được sử dụng tại các điểm cuối?
- Khi bạn nên kích hoạt một vi phạm?
- Làm thế nào bạn nên bảo vệ các dữ liệu? Kiểm toán, mã hóa, ngăn chặn, ... Lựa chọn nên được thực hiện tùy thuộc vào loại thông tin.

2.3.2 Phân loại dữ liệu

Phương pháp để phân tích và khám phá dữ liệu nhạy cảm tồn tại. Các phương pháp phổ biến nhất là sử dụng kết hợp từ khóa, biểu thức thông dụng, so sánh dấu vân tay dữ liệu sử dụng hàm băm và sử dụng các thuật toán học máy.

2.3.2.1. Phương pháp kết hợp từ khóa

Kết hợp từ khóa (Keyword Matching) là cơ bản nhất của tất cả các phương pháp phân tích nội dung. Dựa trên một danh sách các từ khóa, máy quét sẽ đi qua các hệ thống tập tin tìm kiếm các chuỗi văn bản thô phù hợp với các từ khóa được xác định trước.

Việc sử dụng kết hợp từ khóa là chỉ nên dùng cho các văn bản đơn giản có chứa các từ khóa tĩnh phổ biến.

2.3.2.2. Phương pháp sử dụng các biểu thức thông dụng

Việc sử dụng phương pháp kết hợp từ khoá đó vẫn có thể loại bỏ các dữ liệu như số thẻ tín dụng hoặc số an sinh xã hội. Một cách hiệu quả hơn để làm điều này là với biểu thức thông dụng. Như một ví dụ, mỗi loại số thẻ sẽ có các quy định khác nhau về cách thức viết, số kí tự,...

Biểu thức thông dụng là phương pháp thích hợp cho việc phát hiện biến, dữ liệu có cấu trúc. Điều này bao gồm mã nguồn và các thẻ nhận dạng. Điều quan trọng là cho các công ty để thêm các biểu thức theo nhu cầu của họ.

2.3.2.3 Phương pháp so sánh dấu vân tay dữ liệu sử dụng hàm băm

Một phương pháp để phát hiện ra các tài liệu nhạy cảm là bằng cách so sánh nó với một nhóm các tập tin nhạy cảm. Đến kiểm tra xem hai tập tin giống hệt nhau, chúng ta có thể so sánh chúng từng bit một, nhưng một cách hiệu quả hơn để làm điều này là sử dụng hàm băm để tính toán giá trị băm tương ứng của cả tập tin và sau đó so sánh các bit của giá trị băm này.

Bởi vì không phải tất cả dữ liệu được lưu trữ đều là dạng word, hình ảnh, âm thanh và video cũng có thể nhạy cảm, nhưng đối với những định dạng này rất khó để áp dụng phương pháp khác với dấu vân tay. Các tập tin được phân tích và chia nó thành nhiều phần nhỏ hơn và cho từng đoạn một giá trị băm được tính toán. Điều này có nghĩa rằng ngay cả khi một số bộ phận của tập tin thay đổi, nó vẫn có thể được coi là nhạy cảm như các bộ phận khác vẫn được giữ nguyên và phù hợp với các giá trị băm được lưu trữ [1].

2.3.2.4. Các thuật toán học máy

Đối với nhiều tổ chức, tập đoàn lớn, thực tế là có thể có hàng Gigabyte dữ liệu nhạy cảm được tạo ra mỗi tháng. Trong tình huống như vậy thêm từ khóa, biểu thức thông dụng có thể trở nên tốn thời gian và phản tác dụng đối với việc quản lý CNTT.

Tuy nhiên, một hệ thống sử dụng các thuật toán học máy như thế này chỉ làm việc với các tài liệu văn bản, và có thể có một tỷ lệ lỗi cao nếu không được huấn luyện đúng cách. Và một vấn đề quan trọng nữa là phải xác định ngôn ngữ tài liệu nhạy cảm được viết và tạo ra hệ thống các tập tin huấn luyện cho phù hợp.

Một số thuật toán học máy phân loại tài liệu như sau:

- Thuật toán cây quyết định (Decision tree): Đây là phương pháp học xấp xỉ các hàm mục tiêu có giá trị rời rạc. Mặt khác cây quyết định còn có thể chuyển sang dạng biểu diễn tương đương dưới dạng cơ sở tri thức là các luật Nếu – Thì.

- Thuật toán K-Nearest Neighbor (KNN): Ý tưởng chính của thuật toán K-láng giềng gần nhất (K-NN) là so sánh độ phù hợp của văn bản d với từng nhóm chủ đề, dựa trên k văn bản mẫu trong tập huấn luyện mà có độ tương tự với văn bản d là lớn nhất.

- Thuật toán Naive Bayes (NB): Ý tưởng cơ bản của cách tiếp cận Naive Bayes là sử dụng xác suất có điều kiện giữa từ và chủ đề để dự đoán xác suất chủ đề của một văn bản cần phân loại. Điểm quan trọng của phương pháp này chính là ở chỗ giả định rằng sự xuất hiện của tất cả các từ trong văn bản đều độc lập với nhau. Với giả định này NB không sử dụng sự phụ thuộc của nhiều từ vào một chủ đề, không sử dụng việc kết hợp các từ để đưa ra phán đoán chủ đề và do đó việc tính toán NB chạy nhanh hơn các phương pháp khác.

- Support Vector Machine (SVM): Ý tưởng của nó là ánh xạ (tuyến tính hoặc phi tuyến) dữ liệu vào không gian các vector đặc trưng (space of feature vectors) mà ở đó một siêu phẳng tối ưu được tìm ra để tách dữ liệu thuộc hai lớp khác nhau.

- Support Vector Machines Nearest Neighbor (SVM-NN): Support Vector Machines Nearest Neighbor (SVM-NN) (Blanzieri & Melgani 2006) là một thuật toán phân lớp cải tiến gần đây nhất của phương pháp phân lớp SVM. SVM-NN là một kỹ thuật phân loại văn bản máy học sử dụng kết hợp cách tiếp cận K-láng giềng gần nhất (K-NN) với những luật ra quyết định dựa trên SVM (SVM-based decision rule).

2.4. Giải pháp McAfee

Từ bắt đầu chương 2, học viên đã trình bày những vấn đề cơ bản của một hệ thống chống rò rỉ, thất thoát dữ liệu DLP. Phần này học viên xin giới thiệu hệ thống DLP của hãng McAfee như một ví dụ về các giải pháp thương mại hiện nay.

Đặc điểm

- Đảm bảo nội dung nhạy cảm ở khu vực “rest” trước nguy cơ rủi ro.
- Bảo vệ dữ liệu khỏi các rủi ro được biết đến trong quá trình truyền.
- Thu thập những luồng thông tin đi trong mạng để xây dựng những rule và xây dựng phòng chống trước các nguy cơ rủi ro phát sinh.
- Thi hành các hành động khắc phục hậu quả như tự động mã hóa trong quá trình quét dữ liệu hoặc những hành động mà người dùng thực hiện.

Các thành phần:

McAfee ePolicy Orchestrator (EPO) - giao diện điều khiển quản lý tập trung cho các giải pháp DLP của McAfee cũng như tất cả các công nghệ bảo mật thiết bị đầu cuối của McAfee.

McAfee DLP Manager Appliances: các sự kiện quản lý tập trung, khả năng quản lý sự cố, và quản lý của NDLP

McAfee Host DLP Solution: bao gồm Host DLP và endpoint

McAfee Network DLP Monitor: Phân tích tất cả thông tin liên lạc mạng và phát hiện ra mối đe dọa cho dữ liệu của bạn. Thiết bị này sẽ nắm bắt tất cả lưu lượng truy cập mạng của bạn để lại, chỉ số này để điều tra trong tương lai và cũng có thể đánh giá nó so với quy định thời gian thực.

McAfee Network DLP Discover: Đánh giá trung tâm dữ liệu của bạn và tất cả các nguồn dữ liệu khác trên mạng của bạn để phát hiện và phân loại dữ liệu của bạn và đặt nền tảng cho việc bảo vệ tự động

McAfee Network DLP Prevent: Block/Encrypt cả chiều inbound và outbound dựa trên chính sách được định nghĩa.

2.5. Kết luận:

Ở chương này, học viên đã trình bày mô hình cơ bản của hệ thống chống rò rỉ, thất thoát dữ liệu DLP. Cùng với đó, học viên cũng đã giới thiệu một giải pháp thương mại do McAfee cung cấp. Mục đích cuối cùng của hệ thống DLP là việc phân loại và bảo vệ dữ liệu nhạy cảm bằng các chính sách do cơ quan, tổ chức sử dụng tự tạo ra.

CHƯƠNG 3: TRIỂN KHAI HỆ THỐNG THỬ NGHIỆM

Mục tiêu của chương này là triển khai thử nghiệm hệ thống chống rò rỉ, thất thoát dữ liệu dựa trên giải pháp của McAfee.

3.1. Hệ thống thông tin thử nghiệm và các yêu cầu đặt ra

Yêu cầu triển khai:

- Cài đặt và triển khai hệ thống chống thất thoát dữ liệu DLP.
- Kiểm soát thiết bị ngoại vi. Chỉ kết nối với thiết bị ngoại vi đã được đăng kí.
- Block việc upload dữ liệu.
- Triển khai hệ thống giám sát mạng cũng như chính sách theo dõi, ghi log sự kiện toàn bộ hệ thống.
- Phân loại dữ liệu nhạy cảm và tạo khu vực bảo vệ dữ liệu nhạy cảm.
- Việc triển khai không ảnh hưởng người dùng cuối.

3.2. Triển khai hệ thống

- Triển khai McAfee ePolicy Orchestrator (EPO) tại máy chủ x3850 M2.
- Triển khai McAfee EPS xuống tất cả các máy trạm một cách tự động thông qua một Agent duy nhất, McAfee Agent. Thực hiện phân hoạch, gộp nhóm tài nguyên (máy chủ, máy trạm) để dễ dàng xây dựng chính sách bảo mật cũng như quản lý các máy trạm, máy chủ.

3.2.1. Cấu hình yêu cầu

Bảng cấu hình yêu cầu cho việc cài đặt, triển khai hệ thống chống rò rỉ, thất thoát dữ liệu DLP.

3.2.2. *Cấu hình tính năng Block USB và CD-Rom*

Đây là tính năng ngăn chặn kết nối của thiết bị ngoại vi, tuy nhiên có thể linh hoạt tùy chỉnh cho phép sự kết nối của một thiết bị cụ thể.

3.2.3. *Tạo Tag cho File Server cần bảo vệ dữ liệu*

3.2.4 *Cấu hình tính năng Block Upload*

3.2.5. *Thiết lập các chính sách*

3.2.5.1. Phân loại thông tin

- Thông tin bình thường: là những thông tin ngoài công việc, không liên quan đến công ty, được trao đổi hàng ngày giữa các nhân viên trong công ty với nhau (chuyện gia đình, tình cảm, đời sống hằng ngày).

- Thông tin nhạy cảm: là những thông tin liên quan đến công việc trong nội bộ công ty giữa nhân viên với nhân viên hoặc nhân viên với khách hàng (doanh thu, lợi nhuận, tiền lương, PR, chăm sóc khách hàng).

- Thông tin mật: là những thông tin quan trọng của công ty, chỉ những người có quyền hạn thuộc công ty mới được biết (thông tin cá nhân của nhân viên, khách hàng, username, password, hợp đồng, thông tin đối tác).

- Thông tin tuyệt mật: là những thông tin mang tính chiến lược kinh doanh, định hướng của công ty (bản thiết kế, kế hoạch).

3.2.5.2. Chính sách quản lý thông tin

Đối với thông tin bình thường: nhân viên được tùy ý sử dụng, trao đổi ngoài giờ làm việc.

Đối với thông tin nhạy cảm:

- Các nhân viên phải đảm bảo rằng tất cả các thông tin nhạy cảm ở dạng bản cứng hoặc tài liệu điện tử phải an toàn trong khu vực làm việc của mình.

- Máy tính của mỗi nhân viên phải được khóa lại khi không làm việc và được tắt hoàn toàn khi hết giờ làm việc.

- Những tài liệu lưu hành nội bộ không được để trên bàn làm việc mà phải được cất trong một ngăn kéo và được khóa cẩn thận khi nhân viên đi ra ngoài hoặc khi hết giờ làm việc.

- Nhân viên không được viết mật khẩu cá nhân của mình lên giấy dán, notebook, hay những vị trí dễ tiếp cận khác.

- Các nhân viên không được phép tiết lộ mật khẩu tài khoản của mình cho người khác hoặc cho phép những người khác sử dụng tài khoản của mình, bao gồm cả gia đình khi đang thực hiện công việc tại nhà.

- Nhân viên không được tự ý tiết lộ các thông tin liên quan đến công ty trên các trang blog, trên các mạng xã hội như Facebook, Twitter, Google Plus,...

Đối với thông tin mật:

- Bao gồm tất cả các chính sách trên.

- Thông tin cá nhân, username, password được lưu trữ trong các server phải được đặt trong những phòng đặc biệt, được khóa chắc chắn và được giám sát liên tục qua camera, chỉ có nhân viên IT phụ trách mới được phép tiếp cận.

- Các văn bản, giấy tờ quan trọng phải được lấy ra khỏi máy in ngay lập tức sau khi in xong.

- Tất cả dữ liệu mật được lưu trữ trong các thiết bị ngoại vi như CD-ROM, DVD hay USB đều phải được mã hóa và đặt password.

Đối với thông tin tuyệt mật:

- Bao gồm tất cả các chính sách trên.

- Tất cả các tài liệu, giấy tờ sau khi không còn được sử dụng phải được băm nhỏ trong máy cắt giấy và thùng xử lý dữ liệu bí mật phải được khoá cẩn thận.

- Bảng trắng được sử dụng trong các cuộc hội họp cần phải được xóa sạch ngay sau khi cuộc họp kết thúc.

- Tất cả máy in và máy fax phải được xóa hết dữ liệu, giấy tờ ngay sau khi chúng được in.

- Tất cả thông tin tuyệt mật của công ty được lưu trữ phân tán trên hai file server, được phân quyền, gán nhãn tự động bằng Windows Server 2012 Dynamic Access Control kết hợp Right Management Services (cho phép người gửi phân quyền tương tác với nội dung cho người nhận như: cấm in tài liệu, cấm chuyển email cho người khác, thiết lập thời gian hết hạn của tài liệu) và được tự động mã hóa. Không có nhân viên nào được phép truy xuất những thông tin này trừ ban lãnh đạo của công ty.

3.2.5.3. Quản lý truy cập

- Các nhân viên có thể truy cập, sử dụng hoặc chia sẻ thông tin độc quyền chỉ trong phạm vi được ủy quyền và nó thực sự cần thiết để thực hiện nhiệm vụ công việc được giao.

- Bộ phận IT có trách nhiệm tạo các hướng dẫn liên quan đến sử dụng của các cá nhân trên các hệ thống mạng Internet/Intranet/Extranet. Trong trường hợp không có các chính sách như vậy, nhân viên phải được hướng dẫn bởi về việc sử dụng và nếu có bất kỳ vấn đề gì, nhân viên cần tham khảo ý kiến trưởng phòng của phòng ban mà họ đang làm việc.

- Mọi hoạt động của người dùng trong hệ thống đều sẽ được ghi log lại. - Tất cả các thiết bị di động và máy tính có kết nối với mạng nội bộ phải tuân thủ các chính sách quyền truy cập tối thiểu.

- Tất cả các thông tin bí mật của công ty được lưu trữ phân tán trên hai file server, được phân quyền và gán nhãn tự động bằng Windows Server 2012 Dynamic Access Control kết hợp windows Right Management Services và được tự động mã hóa. Không có nhân viên nào được phép truy xuất những thông tin này trừ ban lãnh đạo của công ty.

- Mức độ sử dụng mật khẩu phải tuân thủ các chính sách mật khẩu.

- Tất cả các nhân viên phải cực kỳ cẩn trọng khi mở các file đính kèm trong email nhận được từ người gửi không rõ, vì có thể chứa phần mềm độc hại.

- Cấm sao chép trái phép các tài liệu có bản cứng.

- Các nhân viên không được phép tiết lộ mật khẩu tài khoản của mình cho người khác hoặc cho phép những người khác sử dụng tài khoản của mình, bao gồm cả gia đình khi đang thực hiện công việc tại nhà.

- Cấm mọi hành vi vi phạm an ninh hay làm gián đoạn truyền thông mạng bao gồm: cài đặt các malicious code, tấn công từ chối dịch vụ, giả mạo hay ăn cắp các thông tin của các nhân viên khác.

3.3. Đánh giá hệ thống thử nghiệm:

3.3.1. Kết quả đạt được

Hệ thống DLP thực hiện rất tốt việc block thiết bị ngoại vi và các hoạt động upload file. Việc bảo vệ dữ liệu nhạy cảm cũng được thực hiện tốt. Đặc biệt, với đặc thù là cơ quan nhà nước, việc các dữ liệu nhạy cảm hầu hết là các văn bản biểu mẫu nên hệ thống DLP hoạt động rất tốt, việc quét phát hiện file nhạy cảm đã bao gồm được toàn bộ số dữ liệu

nhạy cảm nhưng việc phân biệt dữ liệu nhạy cảm với những biểu mẫu không quan trọng là chưa hoàn toàn đúng.

Việc triển khai hệ thống DLP đã ngăn chặn sự rò rỉ một cách tình cờ do việc sử dụng các thiết bị ngoại vi hay việc upload file. Đồng thời, hệ thống DLP giúp người dùng xử lý dữ liệu nhạy cảm một cách chính xác bằng các chính sách đã được thiết lập. DLP đảm bảo rằng người dùng sẽ không xử lý các dữ liệu nhạy cảm một cách vô trách nhiệm như gửi kèm nó lên email hay tải lên các trang tin công cộng.

Hệ thống DLP ghi nhận tập trung toàn bộ hoạt động của hệ thống mạng. Có thể tra cứu một cách rõ ràng từng hoạt động trong toàn hệ thống.

3.3.2. Hạn chế của hệ thống

Hệ thống DLP sẽ ghi nhận toàn bộ hoạt động của hệ thống mạng, điều này sẽ tạo cơ sở pháp lý cho những hành động phá hoại cũng như đánh cắp thông tin trong chính nội bộ cơ quan. Tuy nhiên, DLP không ngăn chặn việc người dùng xem các file chứa dữ liệu nhạy cảm, người dùng có thể chụp ảnh màn hình, ghi nhớ thông tin hoặc ăn cắp các ổ đĩa cứng. Để ngăn chặn điều này, cần có biện pháp theo dõi, bảo vệ riêng. Mặt khác, với một nhân viên IT có các kiến thức kỹ thuật có thể vô hiệu hóa hệ thống này, những hành động tấn công, đánh cắp sẽ không được ghi lại. Tuy nhiên, có hệ thống DLP sẽ vẫn tốt hơn rất nhiều, bởi vì không có nhiều người có khả năng này. Và việc ngăn chặn các cuộc tấn công nội bộ không hẳn là cần DLP, nó cần được đề phòng bằng chính các chính sách của các công ty, tổ chức.

Đối với các cuộc tấn công từ bên ngoài, hiệu quả của việc tấn công này phụ thuộc vào cách thức tấn công. Nếu chỉ là việc cài đặt các malware hoặc truy cập các tập tin theo dõi, hệ thống DLP có thể đưa ra cảnh báo và bị phát hiện. Các hoạt động trong mạng có thể được ghi nhận lại toàn bộ bởi hệ thống DLP. Nhưng ngay cả khi có các bằng chứng đó, bạn cũng khó có khả năng giải quyết việc này.

3.3.3. Các biện pháp khắc phục hạn chế

Như đã trình bày ở phần 1.2.2.2, những điều sau sẽ giải quyết phần lớn các cuộc tấn công nội bộ:

- Các chính sách rõ ràng: Nêu rõ chính sách công ty một cách súc tích và dễ hiểu sẽ làm tăng khả năng một nhân viên sẽ thực sự đọc và áp dụng nó khi làm việc. Các chính sách cần hướng dẫn nhân viên về những gì các hành vi, hoạt động xác định được làm, bị cấm.

- Đào tạo tốt: Đào tạo nhân viên về nhận thức chính sách an ninh cũng như giải thích ý nghĩa đằng sau các chính sách khác nhau của công ty sẽ làm tăng hiểu biết của nhân viên về toàn bộ quá trình làm việc, và làm thế nào họ có thể giúp cải thiện nó.

- Kiểm tra lý lịch: Thực hiện kiểm tra nền tảng của nhân viên có thể hỗ trợ trong việc ngăn chặn cá nhân không đáng tin cậy ở giai đoạn đầu.

- Bảo mật vật lý: Hãy chắc chắn rằng cơ sở hạ tầng công nghệ thông tin quan trọng và lưu trữ các thông tin nhạy cảm được bảo vệ. Trộm cắp có thể xảy ra ngay khi có cơ hội tiếp cận cơ sở hạ tầng và nơi lưu trữ thông tin nhạy cảm, hạn chế cơ hội này sẽ đi một chặng đường dài trong việc bảo vệ tài sản kinh doanh.

- Xây dựng niềm tin: Đối xử lao động công bằng và sự tin tưởng là một trong những công cụ đơn giản nhất trong việc chống lại tinh thần thấp và cũng đi một chặng đường dài trong việc xây dựng một lực lượng lao động trung thành. Làm thế nào bạn có thể tin tưởng một người mà không tin tưởng bạn?

Đối với việc phát hiện và ngăn chặn các cuộc tấn công từ xa, việc bổ sung DLP có thể có một số tác dụng. Tuy nhiên, việc hệ thống có một bức tường lửa, IDS, phần mềm chống virus, tiến hành đào tạo nâng cao nhận thức an ninh của nhân viên an toàn hơn nhiều trong việc phòng chống mối đe dọa này, hơn là chỉ cài đặt một hệ thống DLP.

3.4. Kết luận

DLP chỉ là một trong những công nghệ bảo mật thông tin, nó không thể thay thế cho những công nghệ khác. Hệ thống DLP cũng vượt trội trong việc ngăn chặn rò rỉ ngẫu nhiên và các cuộc tấn công từ những người thiếu hiểu biết về công nghệ của hệ thống. Vì vậy, việc triển khai DLP cùng với các công nghệ bảo mật khác là đáng giá.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Hart, Michael, Manadhata, Pratyusa and Johnson, Rob. *Text Classification For Data Loss Prevention*. s.l. : Springer Berlin / Heidelberg, 2011.
- [2] Michael Sonntag (JKU), Vladimir A. Oleshchuk (UiA). *Data Loss Prevention Systems and Their Weaknesses*.
- [3] Mogull, Rich. *Best Practices for Endpoint Data Loss Prevention*. s.l. : Securosis, L.L.C., 2009.
- [4] Mogull, Rich. *Implementing DLP: Deploying Network DLP*. Securosis. [Online] February 13, 2012. [Cited: May 9, 2012.]
- [5] Park, Y., et al. s.l. : *System for automatic estimation of data sensitivity with applications to access control and other applications*, 2011.
- [6] Quellet, Eric. s.l. : *Data Loss Prevention*, 2009.
- [7] Verizon. *2015 Data Breach Investigations Report*.