XÂY DỰNG MÔI TRƯỜNG ẢO HÓA TỰ ĐỘNG PHÂN TÍCH HÀNH VI MÃ ĐỘC

Nguyễn Văn Cường[1*], Nguyễn Hiếu Minh[1], Đỗ Thị Bắc[2]
[1]*Học viện Kỹ thuật Quân sự,*
[2]*Trường Đại học Công nghệ thông tin và Truyền thông – ĐH Thái Nguyên*

TÓM TẮT

Hiện nay cùng với sự bùng nổ của Internet cũng như các thiết bị kết nối mạng, số lượng các vụ tấn công mạng bằng cách sử dụng mã độc ngày càng phổ biến và tinh vi. Mã độc xuất hiện ngày càng nhiều với các hình thái phát triển, phương thức lây lan càng trở nên phức tạp. Việc xây dựng một hệ thống tự động phân tích hành vi mã độc, từ đó đưa ra được các phương án xử lý là điều rất cần thiết trong nghiên cứu cũng như trong thực tiễn. Trong bài báo này chúng tôi giới thiệu một phương pháp xây dựng một hệ thống tự động phân tích hành vi mã độc dựa trên công nghệ ảo hóa Virtualbox và bộ mã nguồn mở Cuckoo.

Từ khóa: *mã độc, phân tích hành vi, ảo hóa, Cuckoo*

MỞ ĐẦU

Mã độc (malware) [1] là những phần mềm được thiết kế nhằm xâm nhập trái phép vào hệ thống máy tính gây tổn hại đến tính bí mật, tính toàn vẹn và tính sẵn sàng dùng của hệ thống.

Mã độc đã xuất hiện trên thế giới từ đầu những năm 1960 và đã có sự phát triển rất mạnh mẽ từ những năm 80 trở lại đây. Có thể nói sự phát triển của mã độc máy tính có một quá trình phát triển khá dài, và nó luôn song hành cùng sự phát triển của công nghệ mạng và truyền thông. Khi mà công nghệ phần mềm cũng như phần cứng phát triển thì mã độc cũng phát triển theo. Ngày nay, cùng với sự phát triển bùng nổ các thiết bị kết nối mạng, thì mã độc không chỉ xuất hiện trong các máy tính mà còn xuất hiện trong các thiết bị điện tử, thiết bị số khác như điện thoại di động, các thiết bị điều khiển từ xa, ... Mỗi năm trên thế giới mã độc gây thiệt hại hàng chục tỉ đô la [2]. Mã độc đang có xu hướng gia tăng, đe dọa an ninh thông tin của các tổ chức và doanh nghiệp. Theo chia sẻ của các chuyên gia bảo mật, việc mất an toàn an ninh thông tin đối với các tổ chức và cá nhân không còn là nguy cơ, rủi ro nữa mà đã ở mức báo động đỏ. Đứng trước tình hình đó, việc nghiên cứu và xây dựng các hệ thống phân tích mã độc, nhằm đưa ra các công cụ phòng chống mã độc hiệu quả có tính khoa học và thực tiễn cao.

Phân tích mã độc là bài toán nghiên cứu các phần mềm độc hại bằng cách phân tích các thành phần khác nhau của nó và nghiên cứu các hành vi của nó trên hệ điều hành của một máy tính nhất định [3]. Mục đích phân tích mã độc chủ yếu để cung cấp các thông tin cần thiết để xây dựng hệ thống bảo vệ và ngăn ngừa các nguy cơ do mã độc gây ra. Nó bao gồm xác định chính xác những gì xảy ra khi mã độc thực thi, ví dụ: các tập tin nào được mã độc sinh ra, các kết nối ra mạng bên ngoài được mã độc sử dụng, những thay đổi về thanh ghi (registry) liên quan đến mã độc, các thư viện và vùng nhớ liên quan đến hoạt động của mã độc. ...Việc phát hiện ra hành vi mã độc sẽ giúp tạo ra các dấu hiệu nhận dạng mã độc để áp dụng vào phát triển các phần mềm bảo vệ mạng và máy tính.

Thông thường khi phân tích mã độc có hai kỹ thuật phân tích cơ bản: *phân tích tĩnh* (Static Analysis) và *phân tích động* (Dynamic Analysis) [4, 5, 6].

Phân tích tĩnh thường đòi hỏi người phân tích xem xét kỹ mã của mã độc đã được dịch ngược (thường được chuyển sang dạng có thể hiểu được, như assembly hay C), hiểu được luồng thực thi và các hành vi của nó thông

* *Tel: 0986 934426, Email cuongmtavietnam@gmail.com*

qua mã đã dịch ngược. Phân tích tĩnh đòi hỏi kiến thức chuyên ngành cao, nắm vững cấu trúc ngôn ngữ lập trình và hệ điều hành.

Phân tích động là phân tích cách hoạt động của mã độc bằng cách thực thi trên một môi trường ảo hóa; quan sát, xem xét nó kết nối đến đâu, lây lan như thế nào, cài đặt những gì vào hệ thống, thay đổi thành phần nào, hoạt động ra sao? Phân tích động sẽ giúp hiểu tổng quan, đầy đủ các thức mã độc hoạt động, rút ngắn thời gian phân tích, giảm được sự phức tạp so với việc phân tích tĩnh. Vì vậy, trong bài báo này chúng tôi giới thiệu một phương pháp xây dựng một hệ thống tự động phân tích hành vi mã độc dựa trên công nghệ ảo hóa Virtualbox và bộ mã nguồn mở Cuckoo.

Phần còn lại của bài báo tổ chức như sau: phần 2, trình bày tổng quan về nguyên lý phát hiện mã độc; phần 3, trình bày một số hệ thống phân tích hành vi mã độc; phần 4, triển khai xây dựng hệ thống phân tích tự động hành vi mã độc dựa trên mã nguồn mở Cuckoo và đánh giá thử nghiệm; phần cuối cùng là kết luận các nghiên cứu đã thực hiện.

NGUYÊN LÝ PHÁT HIỆN MÃ ĐỘC

Có hai nguyên lý phát hiện mã độc: dựa vào dấu hiệu đặc trưng hoặc dựa vào đặc điểm bất thường[3, 4].

Dựa vào dấu hiệu đặc trưng là việc sử dụng một tập các mẫu nhận dạng được gọi là dấu hiệu (signature) để làm căn cứ xác định mã độc. Tập các signature sẽ được xây dựng bằng việc cập nhật các mẫu mã độc đã được kiểm chứng, với việc sử dụng những mẫu được xây dựng chuyên biệt bởi các nhà nghiên cứu.

Phát hiện mã độc dựa vào đặc điểm bất thường chia làm hai giai đoạn:1). Giai đoạn học (Learning), các bộ phân tích (detector) sẽ cố gắng học những trạng thái bình thường. Có thể học các trạng thái từ các host, PUI,… 2). Giai đoạn phát hiện (Detection), kiểm tra (Monitoring), dựa vào các trạng thái đã được học, các bộ phân tích sẽ xác định được trạng thái bất thường và đưa ra cảnh báo.

Trong thực tế, các phần mềm diệt mã độc thường dựa trên hai phương pháp trên để xây dựng nên một số kỹ thuật phát hiện mã độc như: Nhận dạng theo mã băm, quét chuỗi (scan string), giả lập mã (code emulation), phân tích tĩnh (static heuristic analysis), chặn các hành vi (behavior blocking).

Kỹ thuật nhận dạng theo mã băm, có thể sử dụng các thuật toán băm như MD5, SHA, CRC, ... để băm toàn bộ file hoặc một phần thông tin quan trong trong file từ đó so sánh với tập mẫu đã có để kiểm tra. Kỹ thuật này có độ chính xác gần như tuyệt đối tuy nhiên quá trình nhận dạng lâu nếu cơ sở dữ liệu (CSDL) mẫu lớn, quá trình xây dưng CSDL mẫu khó khăn, phức tạp. Khó có khả năng nhận dạng các biến thể của mã độc.

Kỹ thuật tìm kiếm chuỗi là một kỹ thuật sử dụng một chuỗi trích ngang (chuỗi bytes) là đặc trưng của tập tin mã độc và không tồn tại trong các tập tin sạch. Các chuỗi này sẽ được cập nhật vào CSDL mẫu dùng để nhận dạng mã độc. Ưu điểm của kỹ thuật này: nhận dạng chính xác, tốc độ nhận dạng nhanh hơn so với kỹ thuật nhận dạng theo mã băm. Tuy nhiên quá trình xây dựng và cập nhật CSDL rất phức tạp, không phát hiện được khi mã chương trình bị thay đổi.

Kỹ thuật phân tích tĩnh *Heuristic*, đây là phương pháp phát hiện mã độc dựa trên cơ sở tìm kiếm các mẫu "code" được cho là giống với mã độc, thay vì tìm kiếm chính xác từng mẫu. Heuristics tĩnh được thực hiện thông qua 2 bước: thu thập dữ liệu và phân tích. Thu thập dữ liệu là quá trình xây dựng các booster và các stopper. Booster là các dấu hiệu nhỏ thường thấy ở mã độc như: vòng giải mã (decryptionloops), tự thay đổi code (Self-modifyingcode), các chuỗi lệnh bất thường... Stopper ngược lại, là các dấu hiệu ít gặp ở mã độc như lệnh hiển thị hộp thoại. Phân tích là quá trình đưa ra quyết định cuối cùng dựa trên các booster và các stopper phát hiện được. Kỹ thuật này được ứng dụng kết hợp với việc sử dụng trí tuệ nhân tạo trong việc phát hiện các

loại mã độc. Việc thu thập các đặc trưng của mã độc (feature), sau đó tính toán và đưa ra các ngưỡng lây nhiễm (threshold infected) để làm căn cứ phát hiện mã độc.

Kỹ thuật ngăn chặn hành vi là kỹ thuật cho phép ngăn chặn các hành vi, các khối lệnh bị nghi ngờ là mã độc trước khi chúng có cơ hội ảnh hưởng đến hệ thống. Các hành vi bị giám sát bao gồm: Những nỗ lực mở, xem, sửa đổi hoặc xóa bỏ các tệp tin. Những nỗ lực định dạng ổ đĩa hay vùng hoạt động không thể khôi phục khác, sửa đổi thiết lập của tệp tin thực thi và hệ thống như: thiết lập khởi động,...

Kỹ thuật giả lập mã là một kỹ thuật phát hiện mã độc bằng việc giả lập lại hệ thống CPU, hệ thống quản lý bộ nhớ, các chỉ thị máy ở cấp thấp,... giống như máy quét thực tế. Vì vậy, mã độc sẽ hoạt động trên máy ảo mà không ảnh hưởng đến bộ xử lý thật. Hiện nay, kỹ thuật này được sử dụng kết hợp rộng rãi với các kỹ thuật phân tích khác. Cùng với đó, kỹ thuật này đã trở thành một thành phần quan trọng trong công nghệ Sandbox để phát hiện mã độc. Ưu điểm của kỹ thuật này, đó là mã độc hoạt động độc lập, không ảnh hưởng đến hệ thống máy thật. Tuy nhiên việc giả lập lại các thông tin hệ thống CPU, bộ nhớ, ... là rất khó khăn; việc cập nhật và vận hành hệ thống giả lập yêu cầu tính kỹ thuật cao.

Trên thực tế, các hệ thống tự động phân tích hành vi mã độc đều được xây dựng và hoạt động dựa trên kỹ thuật giả lập mã.

MỘT SỐ HỆ THỐNG PHÂN TÍCH HÀNH VI MÃ ĐỘC

Hiện nay các hệ thống Sanbox được sử dụng chủ yếu nhằm phân tích một lượng lớn các mẫu mã độc và là một bước đầu tiên trong quá trình phân tích mã độc hoàn chỉnh. Trước khi sử dụng các kỹ thuật gỡ rối (debug) và bắt đầu phân tích mã thực thi của mã độc thì những thông tin thu được từ Sanbox sẽ rất hữu ích. Để có thể phân tích một lượng lớn mã độc, điều quan trọng là quá trình phân tích phải được tự động hoàn toàn. Đồng nghĩa với việc cần phải có những cách đơn giản để gửi các tập tin mã độc vào sandbox, xác định các tùy chọn và tính năng cho người chạy phân tích và trích xuất kết quả phân tích là một trong những điều cần quan tâm đến.

Trên thế giới đã công khai một số hệ thống hỗ trợ phân tích mã độc tự động. Mỗi hệ thống có những ưu và nhược điểm khác nhau, tuy nhiên các hệ thống này thường chỉ cho người dùng tải lần lượt mã độc lên để phân tích chứ không cho phép tùy biến để phù hợp hơn với công việc của người dùng.

Một số hệ thống phân tích mã độc tự động:

Hệ thống CWSandbox

CWSandbox [7] được phát triển bởi các thành viên tại đại học Friedrich-Alexander,Liên bang Đức. Đây là một hệ thống được thiết kế với ba tính năng: tính tự động hóa, tính hiệu quả và tính chính xác trong phân tích một tập tin win32. Tuy nhiên đây là một hệ thống thương mại hóa không phải mã nguồn mở, người dùng chỉ có thể nhập mỗi lần một mã độc và đợi trả về kết quả phân tích. Thời gian trả về kết quả phân tích lâu, đôi khi việc gửi mã độc lên bị thất bại do hệ thống quá tải[7].

Hệ thống phân tích tự động ThreatExpert

ThreatExpert [8] là một hệ thống phân tích hành vi mã độc tự động được cung cấp tại địa chỉ www.ThreatExpert.com. Hệ thống có thể phân tích và báo cáo về hành vi của các loại virus, worms, trojans, adware, spyware, cũng như các mối đe dọa khác liên quan đến mã độc. Hệ thống máy chủ của ThreatExpert có thể đáp ứng được việc phân tích 1000 mã độc trên một máy chủ trong một ngày[8].

Người dùng có thể gửi các tập tin thực thi hoặc tập .DLL đến ThreatExpert tại địa chỉ http://www.threatexpert.com/submit.aspx. Thông thường thời gian phân tích một mã độc tốn từ 2-3 phút. Sau khi phân tích thành công mã độc, ThreatExpert sẽ gửi báo cáo chi tiết đến email người dùng. Một nhược điểm của hệ thống phân tích ThreatExpert đó là thời gian phân tích lâu, có khi không trả về kết quả

phân tích cho người dùng. Và vì đây là hệ thống đóng nên người dùng không thể tùy chỉnh gì thêm cho mục đích phân tích riêng.

Hệ thống phân tích tự động Joebox

Joebox [9] là một sandbox được thiết kế với sự linh hoạt và tùy biến cao (Tác giả của Joebox là Stefan Buehlmann). Người dùng có thể gửi tập tin đến Joebox thông qua giao diện web, hoặc có thể liên hệ với Joe Security để biết thêm thông tin về việc mua lại quyền sử dụng cho mục đích riêng. Một lợi thế của việc dùng Joebox là hệ thống sử dụng SSDT (system service descriptor table) và EAT (export address table) móc nối với nhân để theo dõi hành vi của mã độc, khác với các sandbox khác là móc nối với các hàm API Windows ở chế độ người dùng. Tuy nhiên chúng ta chỉ có thể gửi cùng lúc một tập tin mã độc lên Joebox mà không thể gửi cùng lúc nhiều tập tin mã độc, JoeBox không lưu bản sao trực tuyến của kết quả phân tích mã độc. Người dùng phải giữ lại kết quả phân tích nhận được bằng e-mail nếu muốn tham khảo lại trong những lần sau. Nếu không người dùng phải gửi lại tập tin để phân tích và nhận kết quả mới[9].

Hệ thống mã nguồn mở Cuckoo

Đối với các hệ thống phân tích tự động ở trên, đều là các hệ thống mã nguồn đóng. Người dùng không thể sửa đổi, tùy chỉnh thêm cho mục đích phân tích của mình. Vì vậy để thực hiện cho việc nghiên cứu, nhóm nghiên cứu sử dụng mã nguồn mở Cuckoo trong phân tích hành vi mã độc.

Cuckoo Sandbox [10] được bắt đầu vào năm 2010, là một dự án trong Google Summer of Code được thiết kế và phát triển bởi Claudio "nex" Guarnieri. Hiện tại phần mềm bao gồm hơn 50000 dòng lệnh và được lập trình bởi bốn người phát triển chính. Đây là một trong những dự án mã nguồn mở đã trở nên phổ biến trong những năm gần đây. Nó được sử dụng rộng rãi bởi các nhà nghiên cứu độc lập cũng như cho các công ty lớn và các doanh nghiệp. Đặc biệt việc dễ dàng sử dụng các hàm API và giao diện người dùng thân thiện cũng như khả năng tự động hóa và có thể tùy chỉnh là một điều cần thiết cho các chuyên viên phân tích mã độc[10].

Với ưu thế là sản phẩm mã nguồn mở, Cuckoo có thể tận dụng sự sáng tạo và đóng góp của toàn thể cộng đồng để tạo nên một hệ thống phân tích mã độc ổn định và hiệu quả hơn. Người dùng hoàn toàn có quyền điều chỉnh, tự xây dựng các mô đun riêng đối với bất kỳ giai đoạn nào của quá trình phân tích. Bên cạnh đó khả năng phát hiện các hành vi độc hại của mã độc thông qua các mô đun nhận dạng dấu hiệu và các dấu hiệu này luôn được cập nhật bởi cộng đồng mạng. Điều này trái ngược với các công cụ thương mại, tuy có nhiều đặc tính và cách tiếp cận nhưng người dùng lại không có khả năng nắm bắt và tìm hiểu sâu vào hệ thống như hệ thống mã nguồn mở Cuckoo. Mặt khác chi phí của một công cụ thương mại thường là quá cao và thường quá tầm chi tiêu cho các tổ chức, các doanh nghiệp nhỏ và vừa.

Một lý do khác để chọn Cuckoo đó là sự thiết kế các môđun phân tích bên trong rất linh hoạt, người dùng có thể dễ dàng tùy chỉnh hoặc cập nhật khi có cả bản cập nhật mới.

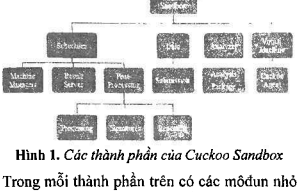
Cuckoo có khả năng thực hiện phân tích các loại tập tin như: Các tập tin thực thi trên môi trường Windows, tập tin DLL, PDF, các tài liệu Microsoft Office, địa chỉ website, các script PHP,...

TRIỂN KHAI XÂY DỰNG HỆ THỐNGCUCKOOSANDBOX

Các thành phần của Cuckoo Sandbox

Các thành phần của Cuckoo Sandbox bao gồm:

- Thiết lập tiến trình (Scheduler)
- Tiện ích (Utils)
- Phân tích (Analyzer)
- Máy ảo (Virtual Machine)



Hình 1. *Các thành phần của Cuckoo Sandbox*

Trong mỗi thành phần trên có các môđun nhỏ đảm nhiệm các chức năng nhất định. Cụ thể:

"Submission": Tải các thông tin của tập tin độc hại cần được phân tích. Các thông tin như: vị trí của tập tin độc hại cũng như các gói phân tích được chỉ định và lựa chọn thời gian chờ cho chạy tập tin. Các thông tin này được lưu trữ trong một CSDL nội bộ và ngay sau khi một máy ảo đã sẵn sàng nó sẽ được xử lý.

"Analysis Pakage": Môđun này xác định chính xác cách một tập tin được thực hiện sau khi tải và máy ảo. Nếu là tập tin PE (.exe) tập tin có thể được bắt đầu trực tiếp trong khi một tập tin dạng Word hay PDF cần phải được nạp vào một ứng dụng office nào đó. Ngoài ra Cuckoo hỗ trợ phân tích URL thông qua việc khởi động trình duyệt "internet Explorer" trên các URL được chỉ định phân tích.

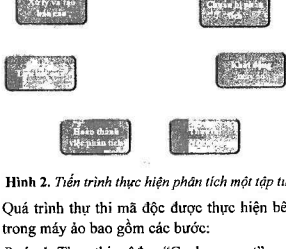
"Cuckoo agent": Môđun này được cài đặt trên các máy ảo. Có nhiệm vụ nhận các tập tin cần phân tích từ Cuckoo Sandbox và thực hiện các phân tích trên máy ảo. Thu thập kết quả phân tích và gửi trả về cho Cuckoo Sandbox

"Machine Managers": Quản lý các thông tin liên quan đến các máy ảo, quản lý các giải pháp ảo hóa, các trạng thái bắt đầu, khôi phục lại hay dừng lại của máy ảo.

"Result Server": Quản lý các cấu hình các máy chủ lưu trữ dữ liệu thu được trong quá trình phân tích.

"Processing": Môđun này được khởi động thì hoàn thành việc phân tích tập tin trên máy ảo. Các thông tin được thu thập: các tiến trình giám sát, các tài nguyên sử dụng, các hành vi được thực hiện trên máy ảo, các hàm API được gọi, ... Đầu ra của môđun này sau đó được đưa tất cả vào môđun"signature" để so sánh phát hiện hành vi đáng ngờ hoặc độc hại.

"Reporting":Trong Cuckoo lưu trữ tất cả thông tin thu thập vào CSDL và tạo ra các loại khác nhau của tập tin đầu ra - chẳng hạn như kết xuất ra một tập tin HTML hoặc một báo cáo dưới dạng các chuỗi JSON.



Hình 2. *Tiến trình thực hiện phân tích một tập tin*

Quá trình thự thi mã độc được thực hiện bên trong máy ảo bao gồm các bước:

Bước 1: Thực thi môđun "Cuckoo agent"

+ Lắng nghe kết nối từ máy chủ.

+ Khi có kết nối từ máy chủ, chấp nhận kết nối.

+ Nhận mẫu mã độc cần phân tích và các thông tin cấu hình yêu cầu.

+ Chạy môđun Analyzer.

Bước 2: Môđun Analyzer được kích hoạt

+ Khởi tạo một số chức năngcơ bản: Mở một kênh IPC (inter-process communication), Dump cấu hình đầu tiên của tiến trình (tạo giao thức Named pipe, địa chỉ IP, và cổng của môđun nhận kết quả "Result Server").

+ Chạy một số gói đặc biệt liên quan tới các ứng dụng mà mã độc sẽ sử dụng (PDF, DOC, DLL, EXE...). Các gói này do môđun "Analysic packages" quản lý.

+ Chèn bộ thư viện CuckooMon có trong môđun "Analysic packages". Bộ thư viện động CuckooMon là thành phần quan trọng của môđun "Analysic packages" chứa các bản

ghi về các hành động của một tiến trình được tiến hành trong môi trường ảo hóa. Đây là thành phần đo đạc thực tế của Cuckoo Sandbox để xác định các hành vi của mã độc thông qua các kỹ thuật Hooks inline.

Bước 3: Thực thi mã độc

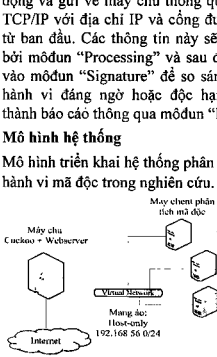
+ CuckooMon được chèn và thực thi đầu tiên thông qua hàm khởi tạo APC(adenomatous polyposis coli).

+ Khởi tạo và chạy các Hooks API có liên quan đến mã độc và thông báo cho môđun "Analyzer " thông qua giao thức "Named Pipe".

+ Mã độc bắt đầu thực thi. Trong quá trình mã độc thực thi, CuckooMon ghi lại các hành động và gửi về máy chủ thông qua giao thức TCP/IP với địa chỉ IP và cổng được khởi tạo từ ban đầu. Các thông tin này sẽ được xử lý bởi môđun "Processing" và sau đó được đưa vào môđun "Signature" để so sánh phát hiện hành vi đáng ngờ hoặc độc hại và đưa ra thành báo cáo thông qua môđun "Reporting"

Mô hình hệ thống

Mô hình triển khai hệ thống phân tích tự động hành vi mã độc trong nghiên cứu.



Hình 3. *Mô hình thử nghiệm hệ thống phân tích tự động hành vi mã độc*

Cuckoo Sandbox bao gồm một phần mềm quản lý trung tâm chạy trong máy chủ Ubuntu để thực hiện xử lý mẫu cũng như phân tích mã độc và một số máy khách (máy ảo để chạy mã độc).

Máy chủ sử dụng CPU64bit, RAM 4GB, ổ cứng 500GB.

Máy chủ: Chạy thành phần cốt lõi của Cuckoo Sandbox bao gồm nhiều đoạn kịch bản viết bằng Python để quản lý toàn bộ các phân tích và quá trình thực hiện cũng như xuất kết quả phân tích. Đồng thời máy chủ chứa hệ thống trang web do nhóm tự xây dựng có thể kết nối và truy xuất tới Cuckoo Sandbox.

Phần mềm ảo hóa VirtualBox: Là một sản phẩm ảo hóa từ Oracle. Lợi thế của VirtualBox là nó có thể cài đặt và chạy trên nhiều nền tảng như Windows, Linux và Mac. VirtualBox là phần mềm nguồn mở và hoàn toàn miễn phí

Máy khách: Là những máy ảo riêng biệt chạy trên nền VirtualBox nơi để thực thi mã độc một cách an toàn riêng biệt cách ly với bên ngoài. Các máy khách được điều khiển bởi Cuckoo thông qua một trình quản lý máy ảo mã nguồn mở. Máy khách bao gồm ba máy chạy Windows XP SP1, Windows XP SP3, Windows 7. Mỗi máy ảo có dung lượng 15 GB, Ram 512Mb

Hệ thống trang Web được cài đặt trên máy chủ Ubuntu: Cho phép kết nối chung CSDLvới phần mềm Cuckoo, cho phép thao tác lên Cuckoo thông qua giao diện Web, hiển thị các báo cáo chi tiết về quá trình phân tích, liên kết một số website phân tích mã độc online.

Kết quả thử nghiệm

Thử nghiệm 1: Phân tích một mã độc (kiểu keyloger) tự xây dựng

Nhóm xây dựng một mã độc (keyloger) có khả năng ghi lại nhật ký bàn phím và gửi thông tin này ra ngoài.

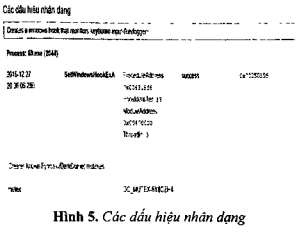
Bước 1: Kiểm tra trên website https://www.virustotal.com. Kết quả có 10/53 phần mềm diệt Anti-virus không phát hiện ra mã độc trong đó có phần mềm Bkav.

Bước 2: Kết quả Phân tích mã độc bằng Cuckoo

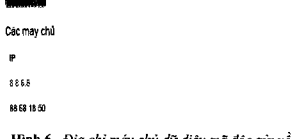
Thông tin chi tiết mã độc:

Kích thước tập tin	6737920 bytes
MD5	a0a162047 2ac4898d75dab81d70c309ca
SHA1	4419dc76f7f7112087fc3026f73800b451b10e5f5a
SHA256	02c5ca2a1e7fdb6380e9331ccc971c52720f2a225dfac7604571 6ed84abfcc7af1
SHA512	7ed9c644b46439831833091c58920cf86858fa0e36ed497e722045ebf575f85a11a8 a7a1f00ef1946b5cc11767dcfe1dc6c9f72 1ab b472e7e914e6fcc432b6ee5585f32b
CRC32	073C1002
Sodeep	12288:r9HFPrbRXlnVVpuoW5xpmx0O1FVBzO+RZTFRZHvVkNCCLr6Ek5C/p6c0U 1mrVVjfvqm2vFbURR6vkH100Z58

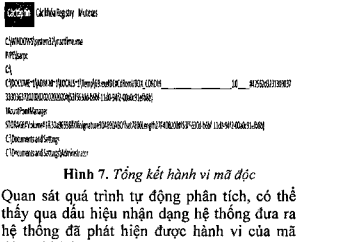
Hình 4. *Thông tin chi tiết về mã độc*



Hình 5. *Các dấu hiệu nhận dạng*



Hình 6. *Địa chỉ máy chủ dữ điệu mã độc gửi về*



Hình 7. *Tổng kết hành vi mã độc*

Quan sát quá trình tự động phân tích, có thể thấy qua dấu hiệu nhận dạng hệ thống đưa ra hệ thống đã phát hiện được hành vi của mã độc: ghi lại phát ký bàn phím, địa chỉ dữ liệu gửi về, các Registry mã độc tham vấn tới, …

Trường hợp 2: Thử nghiệm hệ thống với một số mã độc tự sinh

Chuẩn bị bộ mã độc bằng cách sử dụng hai công cụ sinh mã độc tự động: phần mềm sinh mã độc JPS (Virus Maker 3.0) và TeraBIT Virus Maker 3.2.

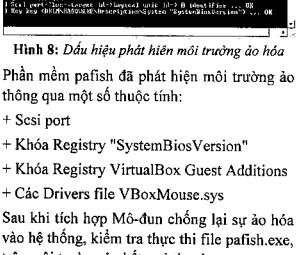
Kết quả thu được:

Tên Mã độc	Khả năng của mã độc	Khả năng phát hiện trên Cuckoo	Tỷ lệ phát hiện trên virustotal.com
01.exe	Chặn mở Notepad Tắt Task Manager Mở cácthư mục	Có	45/55
02.exe	Chặn việc Copy, Move trong window	Có	47/53
03.exe	Chặn mở MsConfig	Có	49/55
04.exe	Thêm 30 tàikhoản vào Window	Có	51/55
05.exe	Luôn luôn Log off	Có	46/53
06.exe	Đóng Internet Explorer sau 10s	Có	47/53
07.exe	Xóa các file trên Desktop	Có	48/55
08.exe	Xóa Windows Fonts Xóa tất cả các file window Document	Có	48/55
09.exe	Tắt chức năng Command Prompt Tắt chức năng Printer	Có	49/54
10.exe	Tắt Regedit Tắt Task Manager Tắt Firewall	Có	48/55
11.exe	Format tất cả các ổ cứng	Có	37/42
12.exe	Tạo bàn phím gõ sai Tạo con chuột di chuyển linh tinh	Có	47/54
13.exe	Ẩn Desktop Icons Ẩn Forder Option Menu Ẩn Taskbar	Có	47/54
14.exe	Khóa các Driver, Forders	Có	48/54
15.exe	Mute, open/close CD, Play beep Sound	Có	47/52
16.exe	Xóa ứng dụng Run từ Start Menu Xóa Start Button	Có	47/54
17.exe	Dừng SQL Server	Không	47/54
18.exe	Tắt máy tính sau 5 phút	Có	47/54

Trường hợp 3: Tích hợp mô-đun khả năng chống lại sự phát hiện môi trường ảo của mã độc do nhóm tác giả xây dựng vào hệ thống

Đối với một số mã độc, trước khi chạy sẽ kiểm tra phát xem môi trường thực thi có phải là máy ảo hay không. Nếu là máy ảo, mã độc sẽ ngừng thực thi.

Để kiểm tra khả năng chống ảo hóa, sử dụng phần mềm pafish.exe. Đây là phần mềm để kiểm tra các đặc tính riêng để phát hiện máy ảo. Khi chạy phần mềm trên hệ thống khi chưa tích hợp mô-đun chống ảo hóa. Pafish đã phát hiện được môi trường ảo (phần bôi đỏ)



Hình 8: *Dấu hiệu phát hiện môi trường ảo hóa*

Phần mềm pafish đã phát hiện môi trường ảo thông qua một số thuộc tính:

+ Scsi port

+ Khóa Registry "SystemBiosVersion"

+ Khóa Registry VirtualBox Guest Additions

+ Các Drivers file VBoxMouse.sys

Sau khi tích hợp Mô-đun chống lại sự ảo hóa vào hệ thống, kiểm tra thực thi file pafish.exe, trên môi trường ảo kết quả thu được:

Hình 9: *Dấu hiệu ngăn chặn sự phát hiện ảo hóa của hệ thống Cuckoo*

Như vậy ngoài khả năng phân tích hành vi mã độc, hệ thống còn có khả năng chống lại sự phát hiện môi trường ảo hóa của mã độc.

Khi một mã độc cố gắng tìm các đặc tính liên quan đến môi trường ảo hóa, Cuckoo sẽ giả các phản hồi để mã độc không phát hiện ra và thực thi.

KẾT LUẬN

Bài báo đã trình bày các kết quả nghiên cứu và xây dựng một hệ thống phân tích tự động hành vi mã độc dựa trên phần nguồn mở Cuckoo, tích hợp thêm mô-đun chống lại sự phát hiện môi trường ảo.

Kết quả thử nghiệm cho thấy, hệ thống có khả năng nhận biết tốt, đưa ra được các phân tích chi tiết, rõ ràng các hành vi của mã độc. Có khả năng ngăn chặn sự phát hiện ảo hóa của mã độc.

Bên cạnh đó với sự hỗ trợ, phát triển mạnh mẽ từ cộng đồng, hiện nay hệ thống Cuckoo không chỉ có khả năng phân tích được các file thực thi trên Windows, mà còn có khả năng phân tích các file nhiễm mã độc khác như DLL, PDF, Microsoft Office Documents, URLs, PHP scripts, ...

Ngoài ra với khả năng tùy biến cao, hệ thống Cuckoo đang ngày càng được hoàn thiện, phát triển và là lựa chọn hàng đầu cho các cá nhân, tổ chức nghiên cứu về lĩnh vực mã độc.

TÀI LIỆU THAM KHẢO

1. Peter Mell, Karen Kent, Joseph NusBaum, "Guide to Malware Incident Prevention and Handling". NIST Special Publication 800-83, Novemver 2005.
2. http://www.mcafee.com/us/resources/reports/rp-quarterly-threats-aug-2015.pdf.
3. Michael Sikorski, Andrew Honig, "Practical Malware Analysis". San Francisco, 2012.
4. Peter Szor, "The Art of Computer Virus Research and Defense", 2005.
5.https://www.solutionary.com/resource-center/blog/2014/12/basic-malware-analysis/
6. Michael Hale Ligh, Steven Adair, Blake Hartstein, Matthew Richard, "Malware Analyst's Cookbook and DVD", 2011.
7. http://cwsandbox.org/
8. www.ThreatExpert.com
9. http://www.joesecurity.org/joe-sandbox-cloud
10. http://Cuckoosandbox.org/
https://www.virustotal.com/

SUMMARY

BUILDINGA VIRTUALIZED ENVIRONMENT FOR AUTOMATIC ANALYSIS OF MALWARE BEHAVIOR

Van Cuong Nguyen[1*], Hieu Minh Nguyen[1*], Thi Bac Do[2]
[1]Le Quy Don University, [2]College of Information and Communication technology - TNU

Along with the explosion of the Internet and networking devices, the number of network attacks using malicious code is increasingly common and sophisticated. The construction of a system that automatically analyzes the behavior of malicious code is essential in the study as well as in practice. We needto build a malware analysis system that would allow us to analyze malware easily without compromising our system. In this paper, we introduce a method to build a system that automatically analyzes the behavior of malicious code based on virtualization technology and the open source Cuckoo Sandbox.

Keywords: *Malware; behavior analysis, virtualization; Cuckoo*

Ngày nhận bài:06/01/2016; Ngày phản biện 28/02/2016; Ngày duyệt đăng. 15/3/2016
Phản biện khoa học: *TS. Tống Minh Đức – Học viện Kỹ thuật Quân sự*

* *Tel: 0986 934426, Email: cuongmtavietnam@gmail.com*