HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

NGUYỄN MINH VƯƠNG

NGHIÊN CỨU, XÂY DỰNG VÀ THỬ NGHIỆM GIẢI PHÁP PHÁT HIỆN MÃ ĐỘC RANSOMWARE

Chuyên ngành : Khoa học máy tính

Mã số : 60.48.01.01

TÓM TẮT LUẬN VĂN THẠC SĨ

HÀ NỘI – 2017

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **PGS. TSKH. Hoàng Đăng Hải**

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

MỤC LỤC

MỞ ĐẦU

Trong năm 2015 và 2016 mã độc mã hóa dữ liệu (được gọi là Ransomware) quay trở lại với nhiều biến thể mới và nguy hiểm. Mã độc loại này được trang bị những thuật toán mã hóa mạnh mẽ, nhiều phương thức lây lan, nhiều biến thể khác nhau, dễ dàng tạo và sử dụng, thanh toán ẩn danh. Do vậy, tính chất nguy hiểm của Ransomware cao hơn rất nhiều cho với các trojan và virus thông thường... Một khi bị nhiễm loại mã độc này, tất cả dữ liệu gốc của nạn nhân sẽ bị mã hóa, các bản dữ liệu gốc sẽ bị xóa hoàn toàn và khả năng khôi phục dữ liệu gần như không có. Nạn nhân muốn lấy lại dữ liệu cần phải trả tiền cho kẻ tấn công để lấy key giải mã mà chúng nắm giữ. Lợi nhuận lớn từ việc phát triển mã độc để kiếm lời đã thúc đẩy sự nguy hiểm, tinh vi của mã độc lên những tầm cao mới, đặt ra nhiều thách thức đối với các biện pháp phòng vệ an ninh.

Phát hiện và xử lý ngăn chặn mã độc là một trong những biện pháp phòng vệ an ninh điển hình, trong đó chuyên gia kỹ thuật cần phân tích, phát hiện mã độc để có giải pháp phòng chống, bảo vệ an ninh cho thông tin và hệ thống thông tin, ngăn chặn và tránh bị mã độc xâm nhập.

Với sự tinh vi và đa dạng của Ransomware cách tiếp cận của những phần mềm diệt virut truyền thống dựa trên chữ ký đã không còn theo kịp sự phát triển của mã độc. Bên cạnh đó việc phân tích tĩnh đòi hỏi trình độ chuyên môn rất sâu, chi phí về thời gian, không kịp thời đáp ứng nhu cầu xử lý ngăn chặn, nhân lực tốn kém.

Với những yêu cầu thực tiễn như vậy, luận văn đặt vấn đề "**Nghiên cứu xây dựng và thử nghiệm giải pháp phát hiện mã độc Ransomware**" nhằm đưa ra một giải pháp hiệu quả, thay thế một phần kiến thức chuyên gia, dễ sử dụng, có khả năng làm chủ công nghệ trong việc phát hiện mã độc Ransomware, từ đó đưa ra biện pháp xử lý, ngăn chặn mối đe dọa này.

Chương 1 - KHÁI QUÁT MÃ ĐỘC RANSOMWARE VÀ CÁC PHƯƠNG PHÁP PHÂN TÍCH MÃ ĐỘC

1.1 Tổng quan về mã độc Ransomware

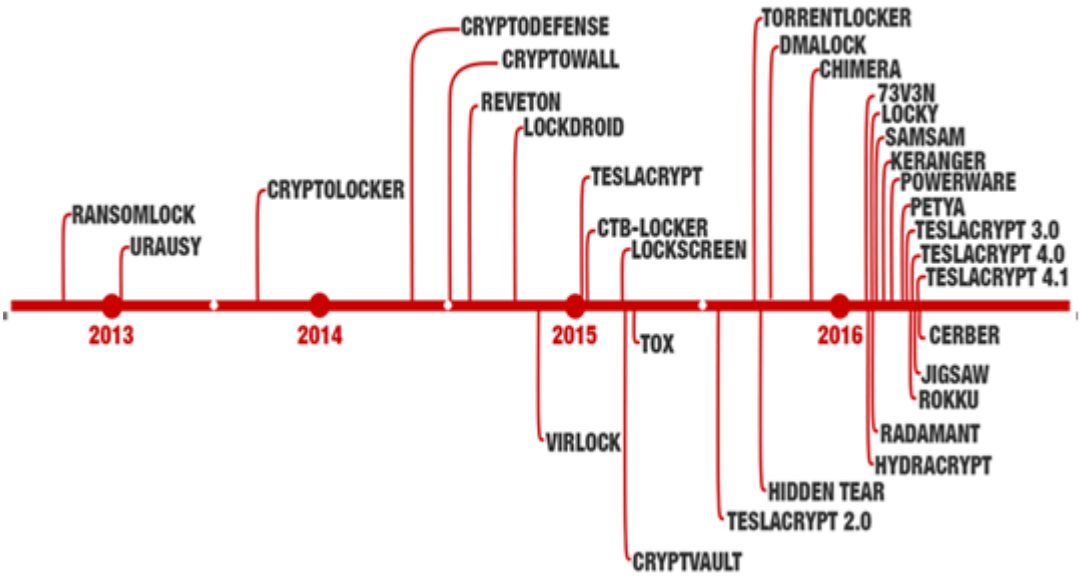
1.1.1. Khái niệm

Ransomware là một loại malware (phần mềm máy tính độc hại) ngăn chặn hoặc giới hạn người dùng sử dụng thiết bị, hệ thống hoặc dữ liệu của mình. Một số loại mã hóa tệp tin khiến nạn nhân không thể mở được tài liệu quan trọng, một số khác dùng cơ chế khóa máy để không cho nạn nhân tiếp tục sử dụng. Để có thể tiếp tục sử dụng hệ thống hoặc đọc dữ liệu cá nhân nạn nhân cần phải trả một khoản tiền cho kẻ tấn công để nhận key giải mã dữ liệu đã bị mã hóa.

1.1.2. Lịch sử phát triển, các biến thể

Mã độc tống tiền Ransomware có lịch sử hơn 20 năm hình thành và phát triển. Ransomware được phát hiện lần đầu tiên vào khoảng giữa năm 2005 - 2006 tại Nga [2]. Năm 2011, một dạng khác của Ransomware là SMS Ransomware đã được phát hiện [4]. Đến năm 2012, Ransomware Reventon sử dụng nhiều tài khoản, cách thức thanh toán khác nhau để nhận tiền của nạn nhân, thông thường là các hệ thống như UKash, PaySafeCard, hoặc MoneyPak [5]. Năm 2014 một phiên bản mã độc mới có tên gọi là CryptoWall [6]. Tháng 3/2015, sự xuất hiện của mã độc Ransomware TeslaCrypt, biến thể này thường xuyên được sử dụng trong các cuộc tấn công lớn. Trong năm 2016, có thể nói là một năm bùng nổ của mã độc Ransomware, rất nhiều cuộc tấn công lớn, sự kiện quan trọng liên quan đến Ransomware, các mẫu mới xuất hiện liên tục và tinh vi hơn rất nhiều. Trong số đó có thể kể đến những biến thể như: Ransom32 and 7ev3n, Locky, SamSam, KeRanger, Petya, Maktub, Jigsaw, CryptXXX, ZCryptor, TeslaCrypt…

Dưới đây là hình vẽ mô tả quá trình phát triển của Ransomware trong những năm gần đây (số liệu thống kê đến cuối năm 2016):



Hình 1.1: Quá trình phát triển của Ransomware.

1.1.3. Mức độ nguy hiểm, nguy cơ, hậu quả

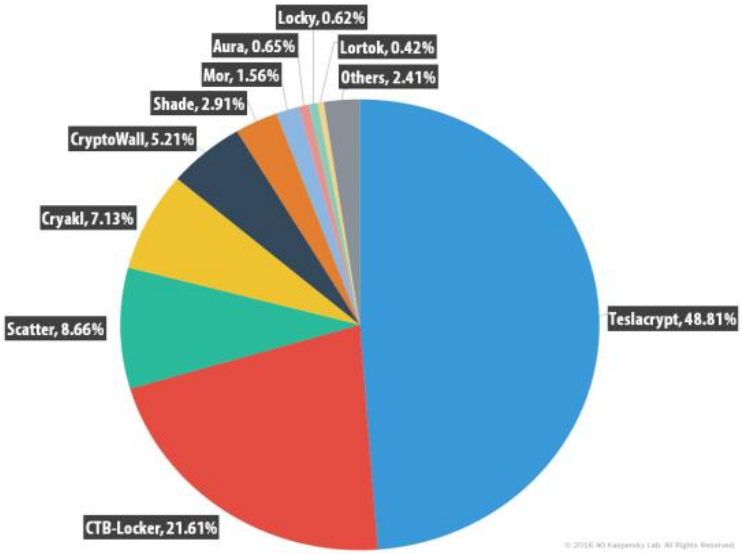
Mã độc sử dụng nhiều phương thức phát tán và lây nhiễm trong đó đặc biệt là hình thức gửi email giả mạo có chứa tài liệu văn bản có chứa mã độc Ransomware hoặc chương trình tự động tải phần mềm mã độc về máy. Các biến thể được cập nhật liên tục để vượt qua các lớp bảo mật an ninh. Các công cụ khai thác lỗ hổng bảo mật tự động được tích hợp để làm con đường lây nhiễm mã độc vào máy nạn nhân. Một khi đã nhiễm mã độc Ransomware dữ liệu sẽ bị mã hóa và tệp tin gốc sẽ bị xóa hoàn toàn và khả năng khôi phục dữ liệu gần như không có. Để lấy lại dữ liệu người dùng cần phải trả tiền chuộc và lấy key bí mật để giải mã. Theo thống kê của hãng bảo mật Kaspersky [20] từ các khách hàng sử dụng sản phẩm của Kaspersky. Việt Nam là một trong những nước bị ảnh hưởng nhiều nhất của mã độc Ransomware.

1.1.4. Thực trạng tại Việt Nam và thế giới

Trong năm 2015 và 2016 mã độc Ransomware là vấn đề nghiêm trọng không chỉ ở Việt Nam mà cả trên phạm vi toàn thế giới. Với giá trị lớn từ đồng tiền ảo như

bitcoin được phát triển đã mang đến một phương thức thanh toán an toàn cho tin tặc. Việt nam có thời điểm đã nằm trong mục tiêu của biến thể Ransomware có tên Locky. Ngoài việc phân chia theo địa lý quốc gia đối tượng tấn công của mã độc Ransomware còn theo các nhóm người sử dụng như: nhóm người dùng thông thường, khối doanh nghiệp, người dùng công cộng.

Dưới đây là số liệu thống kê từ trang chủ của hãng Kaspersky [20] về số lượng các biến thể và mức độ ảnh hưởng của các mã độc mã hóa trong năm 2015-2016.



Hình 1.2: Thống kê số lượng người dùng bị tấn công phân loại theo nhóm mã độc tống tiền mã hóa năm 2015-2016

Table 1: Danh sách các quốc gia bị tấn công Ransomware nhiều nhất trong năm 2015-2016

Quốc gia	2014-2015	2015-2016
Liên Bang Nga	562190	867651
Ấn Độ	143973	325638
Hoa Kỳ	107755	55679
Đức	102289	138750
Việt Nam	96092	89247
Ukraine	69220	39246
Kazakhstan	62719	39179
Algeria	61623	38530
Italy	49400	59130
Brazil	43674	70078

1.1.5. Nhu cầu phân tích phát hiện mã độc Ransomware

Sự thành công liên tục của Ransomware tạo ra một mối đe dọa an ninh mạng nghiêm trọng. Việt Nam đang là một trong nhiều nạn nhân của các cuộc tấn công mã hóa dữ liệu đòi tiền chuộc. Công tác phân tích đòi hỏi trình độ chuyên gia và chuyên môn sâu. Bên cạnh đó sự phụ thuộc vào các sản phẩm nước ngoài khiến cho chúng ta luôn lệ thuộc vào các sản phẩm của nước ngoài. Chính vì vậy việc nghiên cứu các quy trình phân tích, thông tin phương thức hoạt động, và đặc biệt là giải pháp phát hiện mã độc sẽ giúp nhiều người dùng nâng cao nhận thức, cũng như hiểu biết về mã độc để làm cơ sở phát triển các công cụ phát hiện và ngăn chặn mã độc này, và làm chủ công nghệ càng trở nên cấp thiết.

1.2. Các biện pháp phòng chống

Để phòng chống mã độc Ransomware có thể sử dụng một số các giải pháp tạm thời như lưu trữ dữ liệu bằng các hệ thống lưu trữ vật lý, cách này an toàn nhưng chi phí xây dựng bảo trì vận hành cao. Sử dụng giải pháp lưu trữ đám mây nhưng giải pháp này không phù hợp với cơ quan tổ chức yêu cầu tính bảo mật cao bởi dữ liệu có thể bị lộ lọt khi đưa lên môi trường bên ngoài. Sử dụng phần mềm phòng chống mã độc chỉ có thể phát hiện những loại đã biết, không thể phát hiện được các biến thể mới. Các phương pháp này đều có những ưu nhược điểm riêng vì vậy về mặt lâu dài và hiệu quả cao cần nâng cao nhận thức của người sử dụng máy tính kết hợp với việc phát triển những giải pháp nhằm phát hiện sớm các cuộc tấn công dạng này để chủ động phòng tránh và loại bỏ các mối đe dọa.

1.3. Kết luận chương

Chương 1 luận văn đã trình bày cơ bản về thực trạng của mã độc Ransomware bao gồm các nội dung: lịch sử phát triển, mức độ nguy hiểm, thực trạng tại Việt Nam và thế giới, cách nhận biết, một số biện pháp phòng tránh tạm thời và khuyến nghị, quy trình xử lý khi nhiễm mã độc. Các nội dung trên nhằm

giúp người sử dụng có một cái nhìn tổng quát và nhận thức rõ ràng về mối nguy hại mã hóa dữ liệu. Trong chương 2 luận văn tập trung thử nghiệm các phương pháp phân tích mã độc, môi trường phân tích mã độc và những công cụ hỗ trợ để tiến hành phân tích các mẫu mã độc Ransomware để tìm những hành vi đặc trưng nhất, dấu hiệu nhận biết những hành vi này làm tiền đề cho ý tưởng xây dựng giải pháp phát hiện mã độc Ransomware dựa trên hành vi đặc trưng.

Chương 2 - PHƯƠNG PHÁP PHÂN TÍCH, PHÁT HIỆN MÃ ĐỘC RANSOMEWARE

2.1. Một số phương pháp phát hiện nhanh trong thực tiễn

2.1.1. Thông qua danh sách đen (blacklist)

Phương pháp phát hiện dựa trên dấu hiệu kết nối mạng đến các danh sách địa chỉ IP, máy chủ điều khiển C&C thuộc danh sách đen đã biết. Trong luận văn sử dụng cơ sở dữ liệu các máy chủ điều khiển C&C, địa chỉ TOR thuộc danh sách đen để phát hiện loại mã độc này. Danh sách được cập nhật từ nguồn được chia sẻ miễn phí trên trang web *https://ransomwaretracker.abuse.ch/blocklist/*.

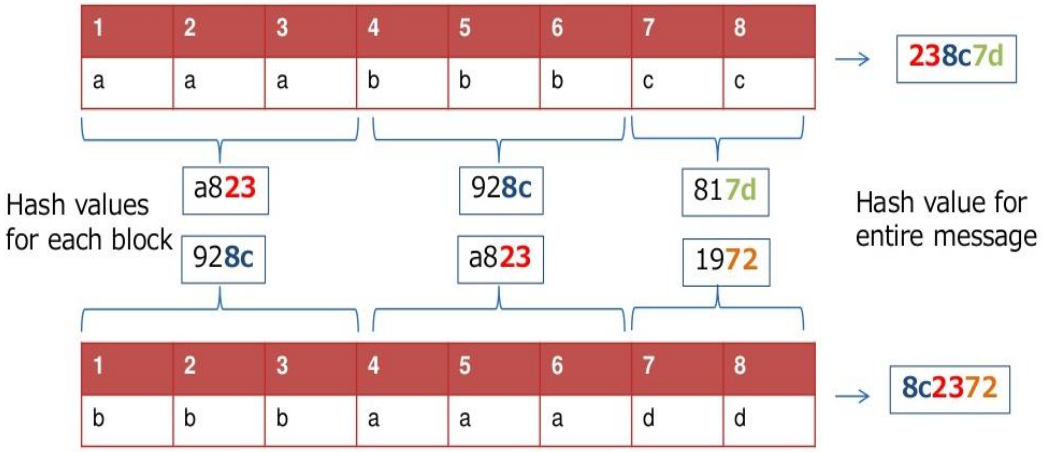
2.1.2. Hashing, dấu vân tay của malware

Hashing là một phương pháp phổ biến được sử dụng để định danh malware. Đưa file malware qua một chương trình hashing sẽ tạo ra một giá trị hash duy nhất. Với đặc tính của thuật toán hash, chỉ cần dữ liệu đầu vào sai khác 1 bit, thì giá trị hash đầu ra sẽ có những sai khác rất lớn và không thể dự đoán được nên giá trị hash đó là định danh của malware, không thể có 2 file khác nhau mà có giá trị hash giống nhau. Thuật toán Message Digest Algorithm 5 (MD5) thường được sử dụng nhiều nhất, tiếp sau là Secure Hash Algorithm 1 (SHA-1) cũng khá phổ biến. Hiện nay chuyên trang virutotal.com đang được sử dụng rất nhiều với hàng chục các hãng Antivirus uy tín.

2.1.3. Kỹ thuật Fuzzy hashing

Vẫn là nhận dạng mã độc thông qua mã hash tuy nhiên đã được bổ sung thêm các phân tích và tính toán để từ một mã hash của mã độc, có thể nhận dạng ra các hash họ hàng của của mã độc từ đó nâng cao khả năng phát hiện mã độc. Ưu điểm của kỹ thuật này là nó cao cấp hơn kỹ thuật checksum vì được cải tiến kỹ thuật phát hiện họ hàng của mã độc. Tuy nhiên nhược điểm của nó nằm ở chỗ xây

dựng các thuật toán và lựa chọn độ dài ký tự phù hợp là khó khan dẫn đến có khả năng cảnh báo giả và cảnh báo sai.



Hình 2.1: thuật toán Fuzzy Hashing

2.1.4. Kỹ thuật Scan String

Kỹ thuật này sử dụng một chuỗi trích ngang (chuỗi byte) là đặc trưng của tập tin mã độc và không tồn tại trong các tệp tin sạch để làm cơ sở dữ liệu mẫu dùng để nhận dạng mã độc. Với ưu điểm nhận dạng chính xác, tốc độ nhận dạng nhanh hơn so với kỹ thuật checksum, tuy nhiên quá trình xây dựng và cập nhật cơ sơ dữ liệu phức tạp, nhận dạng bị động và không phát hiện được mã chương trình bị thay đổi.

2.1.5. Kỹ thuật Code Emulation

Là một kỹ thuật phát hiện mã độc dựa trên việc mô phỏng lại hệ thống CPU, hệ thống quản lý bộ nhớ, các mã máy ở cấp thấp. Ưu điểm mã độc hoạt động độc lập không ảnh hưởng đến hệ thống máy thật. Nhược điểm quá trình mô phỏng đòi hỏi kỹ thuật cao và khó khăn.

2.2. Thiết lập môi trường hỗ trợ phân tích, phát hiện mã độc

Môi trường cần phải đầy đủ các phần mềm, bộ công cụ cần thiết nhằm đảm bảo mã độc có thể hoạt động và thể hiện hết được các hành vi, hàm chức năng được thiết kế. Có hai cách xây dựng môi trường phân tích là xây dựng trực tiếp trên phần

cứng hoặc xây dựng hệ thống phân tích trên phần mềm. Khi xây dựng môi trường phân tích trên phần cứng (môi trường thực) khả năng phân tích hiệu quả, lượng thông tin thu dược chính xác và đầy đủ, tuy nhiên việc xây dựng trên môi trường thật cần rất cẩn thận và yêu cầu phải kiểm soát được hệ thống này sẽ không làm ảnh hưởng đến các hệ thống dịch vụ khác trong và ngoài mạng. Bên cạnh đó để vận hành hệ thống thật cần có trình độ chuyên môn cao kết hợp với các quy trình kỹ thuật để tránh xảy ra các thả họa lây nhiễm trên toàn mạng.

2.3. Phân tích đánh giá các phương pháp

2.3.1. Phương pháp phân tích tĩnh

Được chia thành hai cấp độ là cơ bản và nâng cao. Phân tích tĩnh cơ bản có thể xác định một tập tin là độc hại, cung cấp các thông tin chức năng của tập tin độc hại đó, thông tin này sẽ cho phép người dùng tạo ra các chữ ký mạng đơn giản. Mức nâng cao của phân tích tĩnh là kỹ thuật dịch ngược (Reverse Engineering) nội dung bên trong của mã độc bằng cách dịch ngược các hàm chức năng được gọi và thực thi. Các chỉ thị được thực thi bởi CPU, phân tích tĩnh nâng cao sẽ cho biết tiến trình nào là đáng ngờ.

Ưu điểm: Có thể hiểu và biết chính xác các kỹ thuật viết mã độc, hoạt động của mã độc.

Nhược điểm: Phân tích tĩnh nâng cao đòi hỏi nhiều kiến thức chuyên môn về lập trình, cấu trúc mã lệnh, và các khái niệm về hệ điều hành và thời gian phân tích lâu, tốn nhiều công sức. Không phù hợp với thực tế cần phân tích nhanh và số lượng lớn.

2.3.2. Phương pháp phân tích động

Phân tích động là kiểm tra bất kỳ quá trình nào chạy khi thực thi mã độc. Phân tích động thường được tiến hành sau khi phân tích tĩnh đã không khả năng phân tích được mã độc, khi mà mã độc được sử dụng kỹ thuật làm rối obfuscation,

pack hoặc khi đã sử dụng hết các kỹ thuật phân tích tĩnh sẵn có. Phân tích động còn có thể liên quan đến việc giám sát hoặc kiểm tra hệ thống sau khi mã độc được thực thi. Không giống phân tích tĩnh, phân tích động cho phép quan sát được chức năng của mã độc. Ví dụ nếu mã độc là một keylogger, phân tích động cho phép bạn xác định được tệp tin nhật ký trên hệ thống, tìm được nơi sẽ gửi thông tin đến... Những thông tin rõ ràng như vậy rất khó để thu được nếu chỉ sử dụng các kỹ thuật phân tích tĩnh.

Ưu điểm: Quá trình phân tích diễn ra nhanh hơn, dễ dàng hơn. Các hành vi được ghi lại một cách rõ ràng, người phân tích không quá quan trọng về kiến thực chuyên gia trong lĩnh vực dịch ngược. Các công nghệ mới và hiện đại được tích hợp như giả lập CPU, kích hoạt thời gian chạy chống ngủ đông, tự động đáp ứng các yêu cầu đầu vào...

Nhược điểm: Quá trình phân tích động vẫn có khả năng bỏ sót một số hành vi khi mã độc sử dụng các kỹ thuật có yêu cầu đầu vào cụ thể mà một môi trường phân tích không cung cấp tự động được. Mặt khác việc phân tích động đòi hỏi sử dụng nhiều công cụ kết hợp và quan trọng nhất là môi trường để thực hiện phân tích. Nếu môi trường thực hiện phân tích không đạt chuẩn sẽ dẫn đến những sai lệch hành vi của mã độc.

2.4. Tiến hành phân tích thu thập hành vi đặc trưng

2.4.1. Ý tưởng

Thu thập các mẫu mã độc đã biết, phân loại các mẫu mã độc theo các phiên bản biến thể. Sử dụng phương pháp phân tĩnh cơ bản để tìm hiểu đặc tính của mã độc. Sử dụng phương pháp phân tích động bằng bộ công cụ đã giưới thiệu để xem xét các hành vi và quá trình thay đổi các giá trị registry, thêm sửa xóa các tệp tin, lời gọi hàm chức năng trên hệ thống. Sử dụng kết quả phân tích của các Sandbox đối chiếu với quá trình phân tích bằng tay các biến thể của họ mã độc TeslaCrypt và

CryptoWall để tìm kiếm những hành vi đặc trưng. Phân loại các hành vi đặc trưng theo các giai đoạn lây nhiễm và thực thi trong máy nạn nhân.

2.4.2 Công cụ hỗ trợ

Công cụ Process Hacker: Công cụ cho phép xem danh sách các tiến trình dưới dạng cây của từng tiến trình và dịch vụ đang chạy dưới tiến trình nào. Hoặc các tiến trình được sinh ra từ tiến trình gốc. Ngoài ra công cụ Process Hacker cũng thu được kết quả tương tự để kiểm chứng.

Công cụ Process Moniter: cho phép theo dõi các tiến trình sinh ra sẽ hoạt động như thế nào và có những tác động gì với hệ thống

Công cụ SysTracer: kiểm tra các thay đổi giá trị của Registry bằng cách so sánh hai trạng thái trước và sau khi chạy mã độc để được kết quả những giá trị registry nào bị thay đổi và thay đổi do chương trình nào yêu cầu.

Công cụ Moniter API: Để phát hiện những hàm và thư viện mã độc gọi trong quá trình chạy các chương trình thực thi trên hệ thống.

Công cụ WireShark: Wireshark là một công cụ kiểm tra, theo dõi và phân tích thông tin mạng WireShark hỗ trợ các rất nhiều giao thức, từ những loại phổ biến như TCP, IP đến những loại đặc biệt như là AppleTalk và Bit Torren.

Ngoài ra luận văn còn sử dụng bộ công cụ phân tích tổng hợp do hãng Microsoft cung cấp có tên "Sysinternal Suite" và rất nhiều phần mềm hỗ trợ khác trong quá trình phân tích mã độc.

2.4.3. Thực hiện phân tích mẫu Ransomware

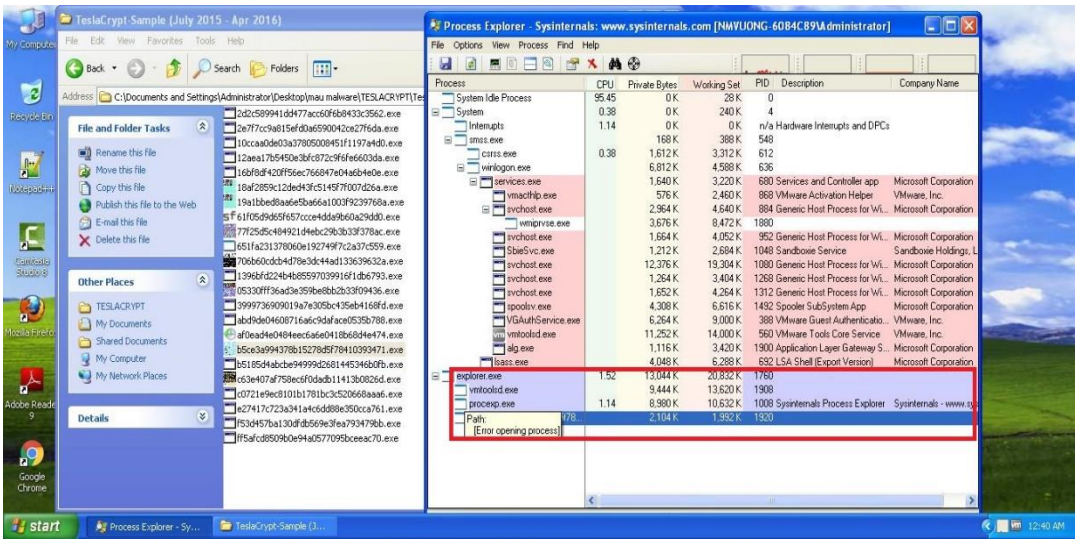
a. Tạo môi trường ảo

Sử dụng phiên bản HĐH Windows XP SP3 vì Windows là hệ điều hành phổ biến của tất cả mọi loại mã độc (trừ mã độc dành cho mobile) có thể hoạt động được. Thứ hai, hệ điều hành này gọn nhẹ và đơn giản, chỉ cần 512MB RAM là đủ cho việc phân tích các mẫu malware nhỏ đến vừa. Các dịch vụ trên XP cũng ít và đơn giản, không rắc rối và nhiều như các phiên bản sau của họ Windows. Và SP3 là

phiên bản ổn định nhất của hệ điều hành này. So với SP2, SP3 được phép cài đặt nhiều gói phần mềm từ Microsoft hơn. Tắt các dịch vụ ảnh hưởng đến mạng của Windows như Windows Update, Firewall… để tránh cản trở mã độc hoặc các gói tin bị lẫn vào dữ liệu mạng giám sát. Cài thêm các gói phần mềm bổ trợ: Net Framework (tất cả các bản từ 2.0 đến mới nhất) và Java Runtime Environment các bản cũ 32 bit vì tính tương thích với hệ điều hành, Adobe Flash Player, Office 2003, python 2.7 và một số phần mềm khác.

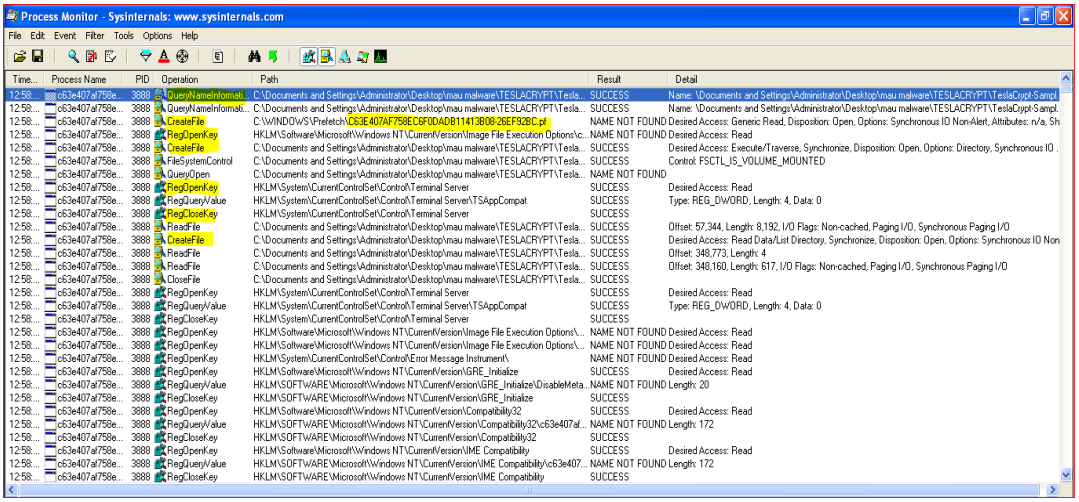
b. Thu thập mẫu và phân tích

Sử dung công cụ **Process Explorer** thu thập các tệp tin mà mã độc lây nhiễm tạo thêm vào hệ thống.



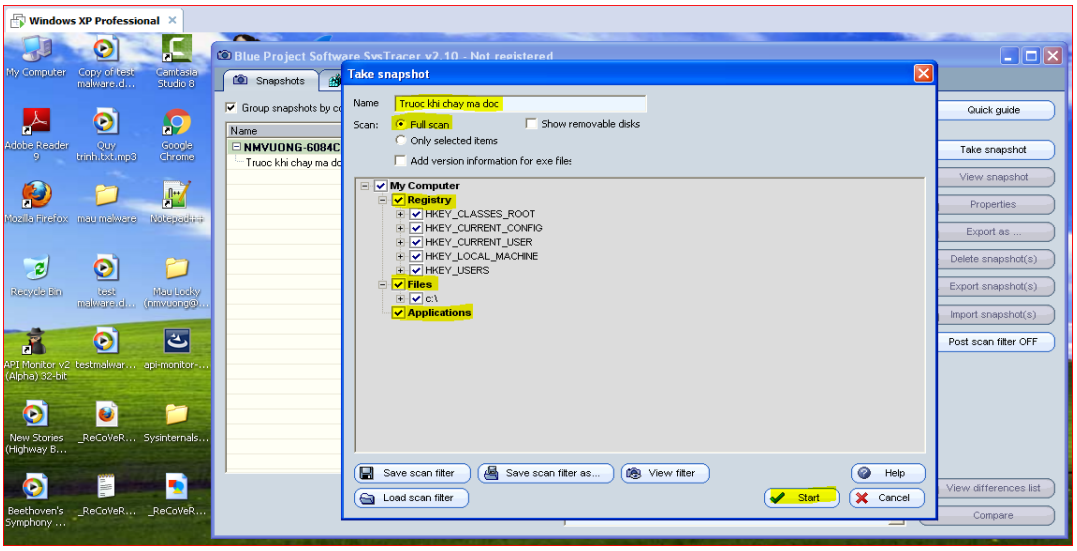
Hình 2.2: Phân tích mã độc bằng công cụ Process Moniter

Sử dụng công cụ Process Moniter cho phép theo dõi các tiến trình sinh ra sẽ hoạt động như thế nào và có những tác động gì với hệ thống. Đây là công cụ rất mạnh với bộ lọc cơ động giúp chuyên viên phân tích dễ dàng xem xét việc thêm sửa xóa các tệp tin trên hệ thống.



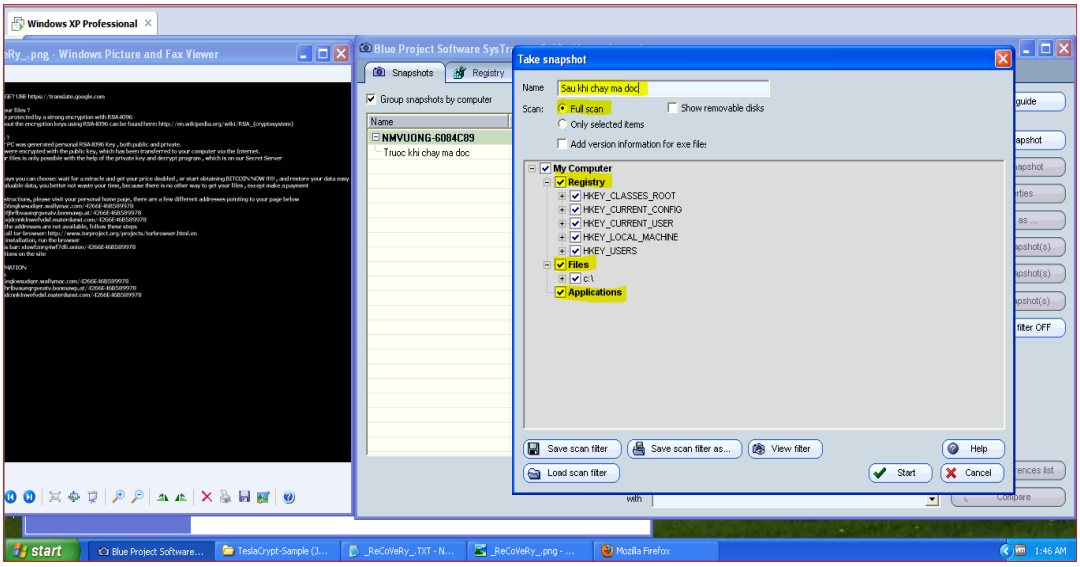
Hình 2.3: Sử dụng bộ lọc trong công cụ Process Moniter

Sử dụng công cụ **SysTracer** để thực hiện kiểm tra các thay đổi giá trị registry trên máy nạn nhân. Đầu tiên ta sử dụng một bản lưu trạng thái registry sạch (trước khi chạy tệp tin mã độc).



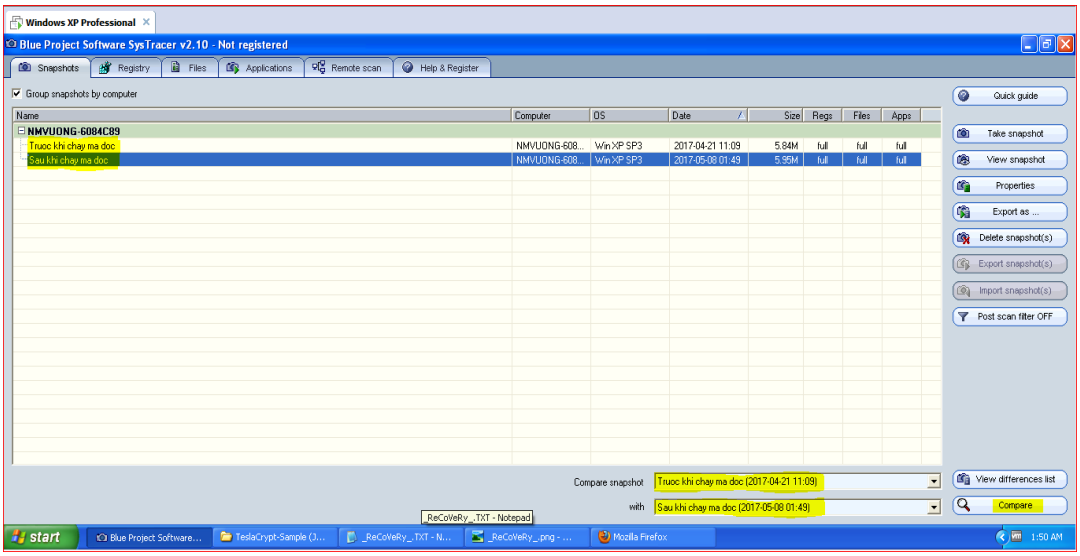
Hình 2.4: Tạo bản Snapshot trạng thái hệ thống trước khi chạy

Tiếp theo tiến hành thực hiện chạy mã độc và thực hiện chụp trạng thái sau khi chạy mã độc.



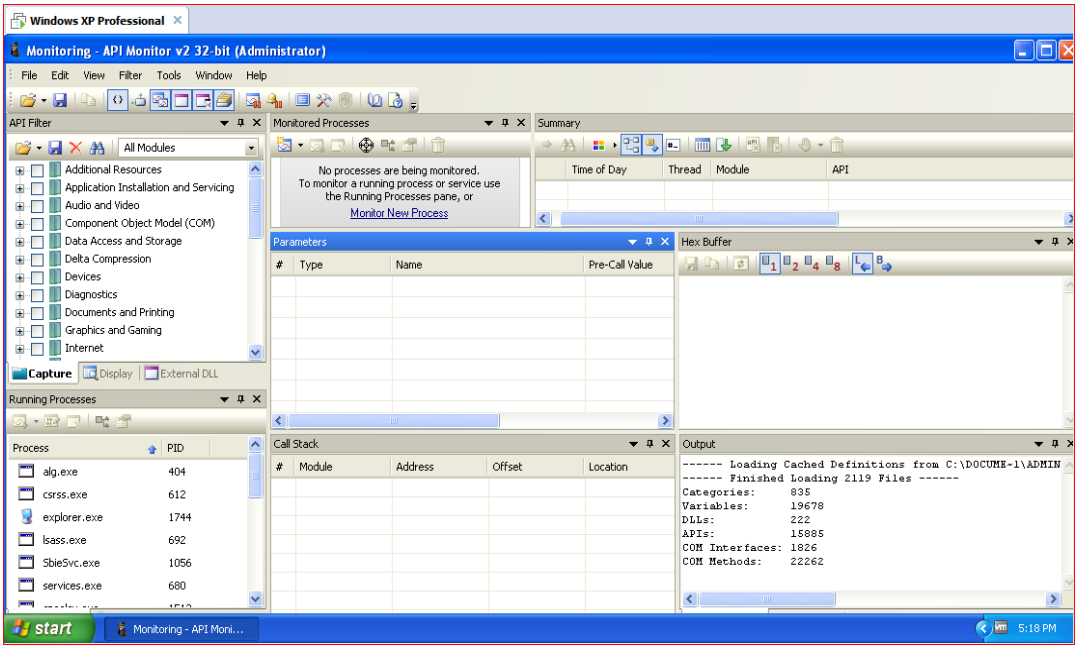
Hình 2.5: SysTracer Sau khi chạy mã độc

So sánh hai trạng thái này sẽ thu thập được các giá trị registry đã bị thêm sửa xóa để kiểm tra hành vi của mã độc.



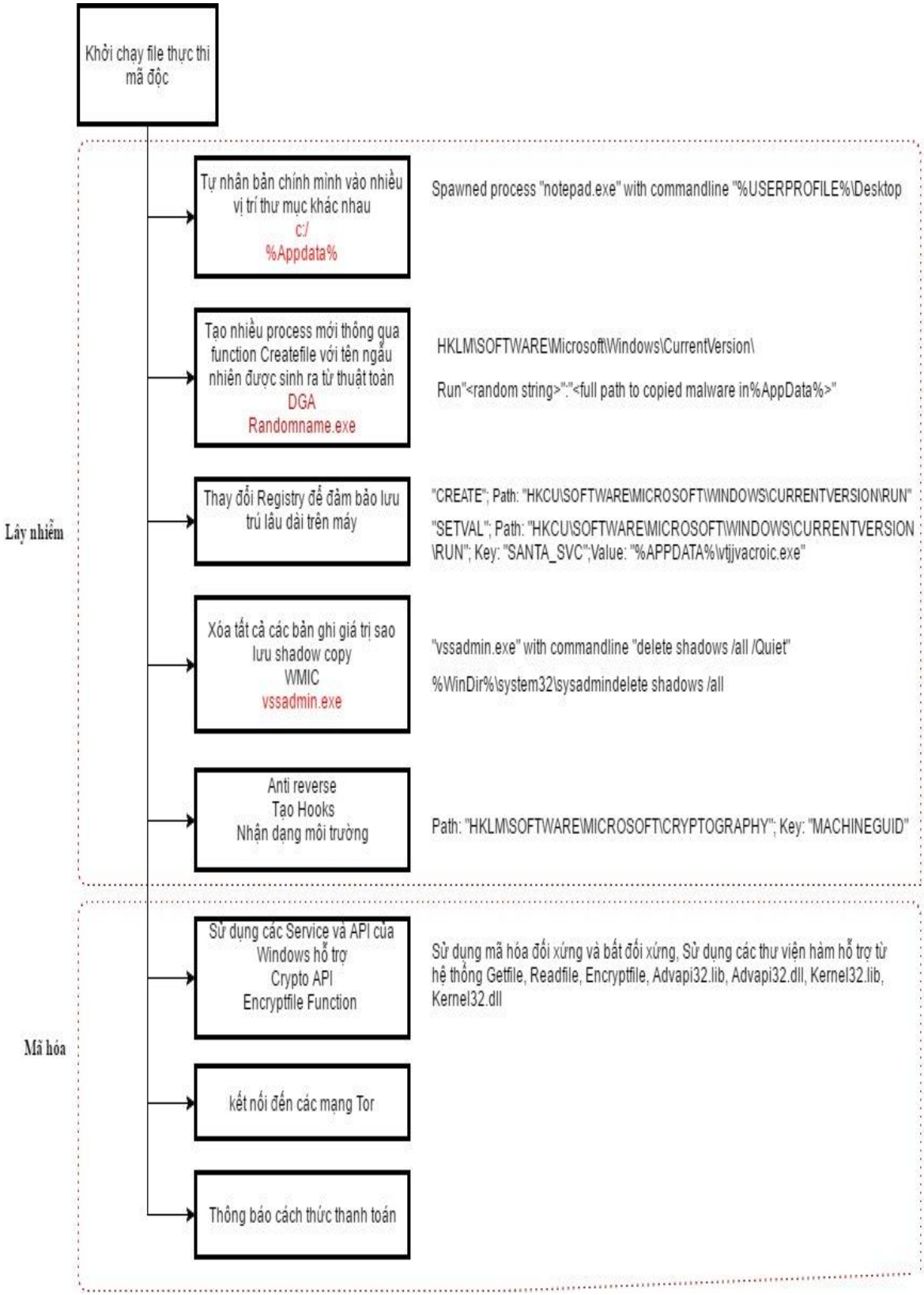
Hình 2.6: So sánh 2 trạng thái trước và sau để thấy sự thay đổi giá trị hê thống

Sử dụng công cụ **Moniter API** để phát hiện những hàm và thư viện mà mã độc gọi trong quá trình chạy



Hình 2.7: Công cụ Moniter các API

Qua quá phân tích có thể chia quá trình hoạt động của mã độc thành hai giai đoạn là lây nhiễm và mã hóa tương ứng với các hành vi và được mô tả cụ thể trong mô hình:



Hình 2.8: Mô hình hành vi

2.5. Kết luận chương

Môi trường phân tích là yếu tố rất quan trọng, ảnh hưởng trực tiếp đến kết quả phân tích chính vì vậy cần lựa chọn và thiết lập môi trường ổn định và phù hợp với mục tiêu phân tích. Ngoài ra việc lựa chọn kỹ thuật phân tích và công cụ cũng cần được phân tích kỹ để thấy những điểm mạnh và điểm yếu của từng công cụ. Việc lựa chọn những công cụ đơn giản dễ sử dụng được ưu tiên để giải quyết nhu cầu cần phân tích nhanh mà thu thập được các hành vi, các lệnh gọi hàm hệ thống. Sau khi phân tích điểm mạnh yếu các phương pháp phân tích luận văn đã thực nghiệm phân tích bằng phương pháp phân tích động để lựa chọn ra một số các hành vi đặc trưng. Chương tiếp theo luận văn sẽ tiến hành xây dựng một chương trình có khả năng phát hiện mã độc mã hóa dữ liệu dựa trên hành vi đã thu thập được.

Chương 3 - XÂY DỰNG VÀ THỬ NGHIỆM GIẢI PHÁP PHÁT HIỆN RANSOMWARE

3.1. Kiến trúc và các thành phần của giải pháp

3.1.1. Ý tưởng đề xuất

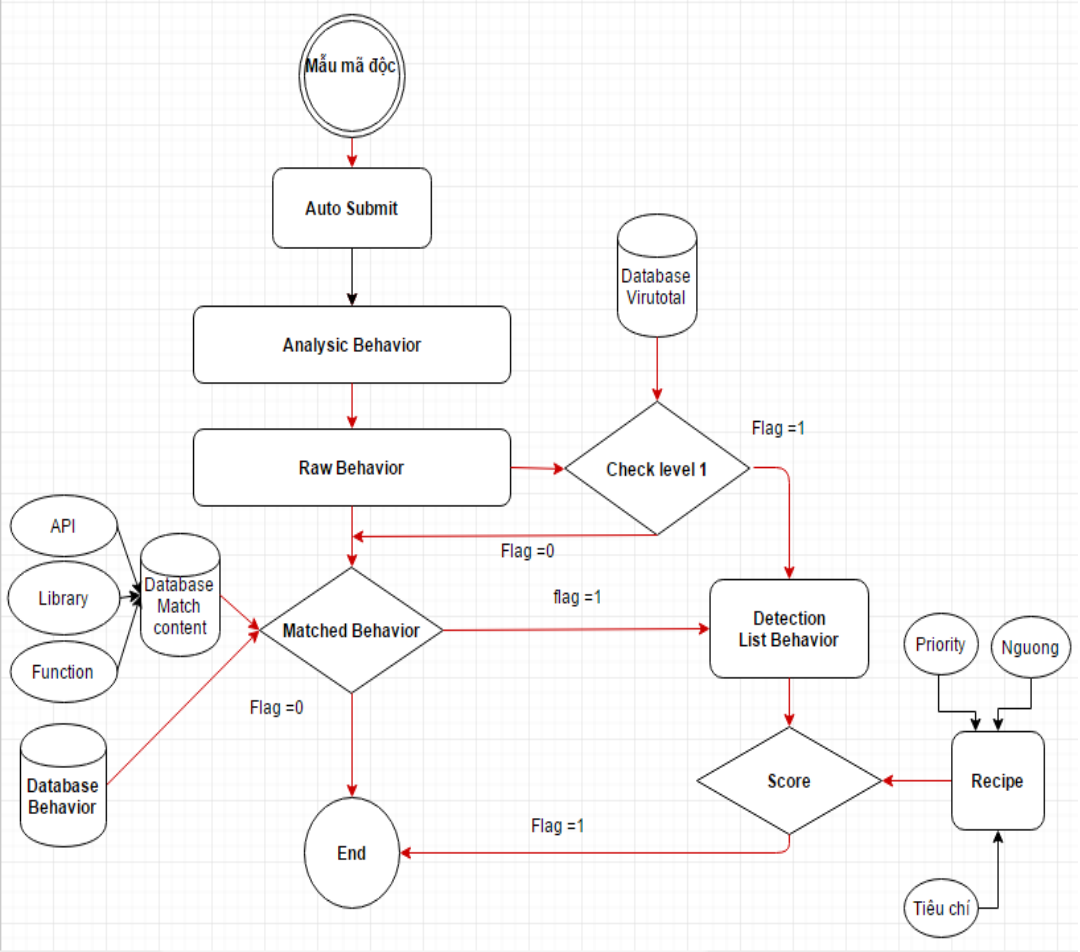
Qua việc phân tích một số mẫu mã độc điển hình trong Chương 2 của luận văn cho thấy mã độc Ransomware được thiết kế rất tinh vi và có nhiều kỹ thuật nhằm vượt qua các phần mềm phát hiện và phòng chống mã độc. Điển hình như kỹ thuật làm rối mã, kỹ thuật chống phân tích, kỹ thuật khai thác lỗ hổng trên hệ điều hành… Trong khi đó các giải pháp phát hiện đều đang có những điểm yêu nhất định. Kỹ thuật heuristic có thể giúp phát hiện ra được các hàm và thư viện có khả năng được sử dụng bởi mã độc. Tuy nhiên điểm bất cập của nó là trong các trường hợp khác nhau thì việc gọi hàm cũng có mục đích khác nhau, không phải tất cả đều là mục đích phá hoại. Kỹ thuật phát hiện dựa trên hành vi có phần chính xác hơn tuy nhiên không phải hành vi nào của mã độc cũng được thể hiện rõ ràng khi hoạt động. Kỹ thuật sử dụng chữ ký tuy không phát hiện được biến thể mới nhưng chúng vẫn nên được sử dụng để phát hiện sớm trong các cuộc tấn công sau và làm giàu thông tin cơ sở dữ liệu. Với những ưu nhược điểm của các kỹ thuật phát hiện mã độc cùng với những kỹ thuật che dấu sẵn có dẫn đến việc phát hiện chính xác mã độc và không bị cảnh báo sai là nhu cầu cần thiết. Xuất phát từ quá trình nghiên cứu tìm hiểu này luận văn đề xuất xây dựng một giải pháp kết hợp giữa các kỹ thuật phát hiện nêu trên nhằm mục đích phát hiện sớm mã độc mã hóa dữ liệu Ransomware.

3.1.2. Kiến trúc và các thành phần chương trình

Chương trình gồm 2 module, module 1 và module 2. Module 1 có chức năng tải (upload) mã độc vào môi trường giả lập để tiến hành phân tích và lấy dữ liệu về hành vi của mã độc, sau đây gọi là module chuyển đổi tập dữ liệu mẫu (Module tiền xử lý). Module 2: ngoài chức năng phân tích các hành vi và sử dụng CSDL hành vi

mẫu để phát hiện mã độc, module này còn tích hợp chức năng tính điểm đánh giá mức độ nguy hiểm và thống kê.

Sơ đồ kiến trúc tổng thể của chương trình như sau:



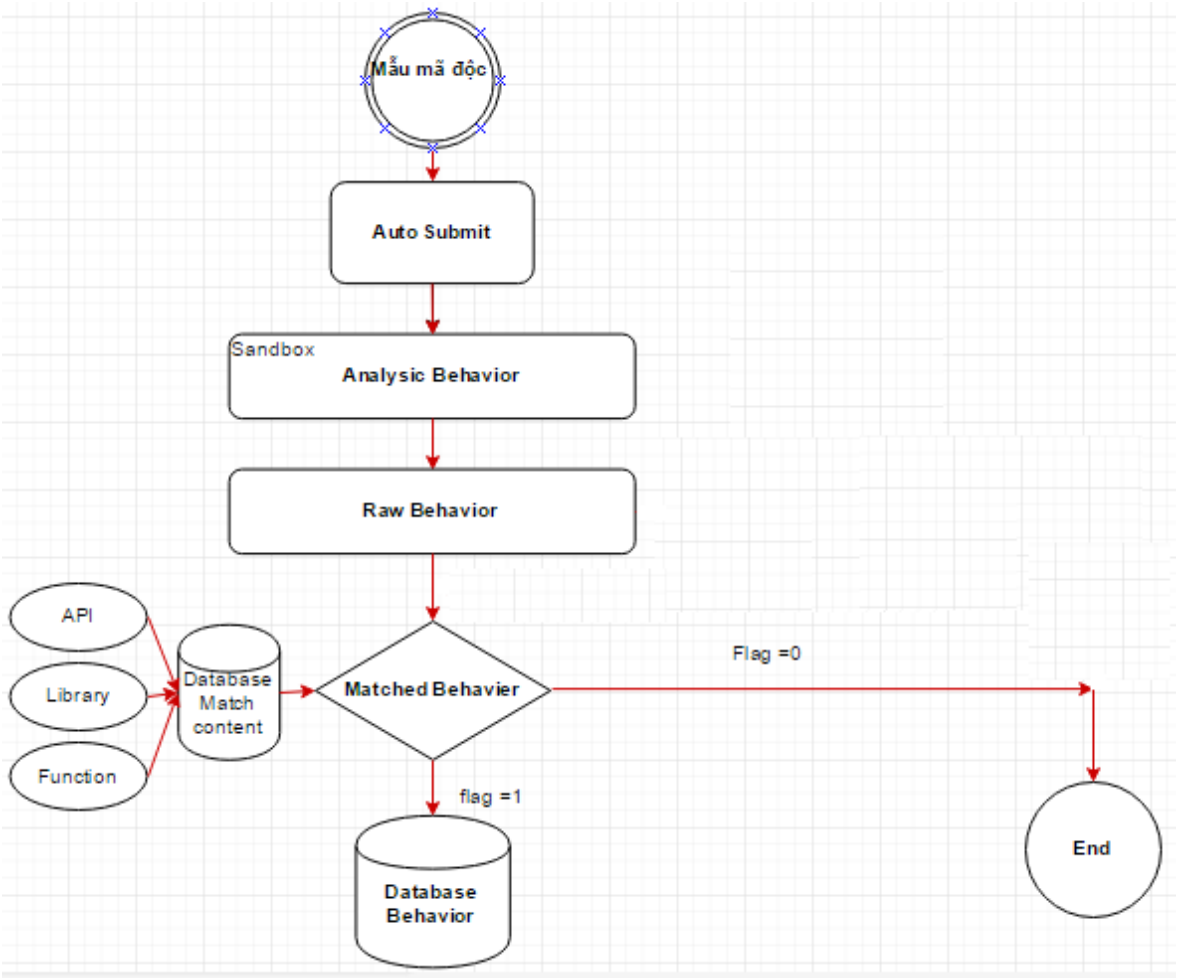
Hình 3.1: Kiến trúc chương trình

Kiến trúc chương trình được chia làm hai thành phần chính là: Thành phần xử lý dữ liệu và thành phần phát hiện mã độc.

3.1.3. Các Module chương trình

a. Xây dựng module xử lý dữ liệu

Module chuyển đổi tập dữ liệu (tiền xử lý) có chức năng thực hiện thu thập các hành vi của mã độc sau khi chạy phân tích động. Đầu ra của module là bảng CSDL các hành vi của các mẫu và biến thể.

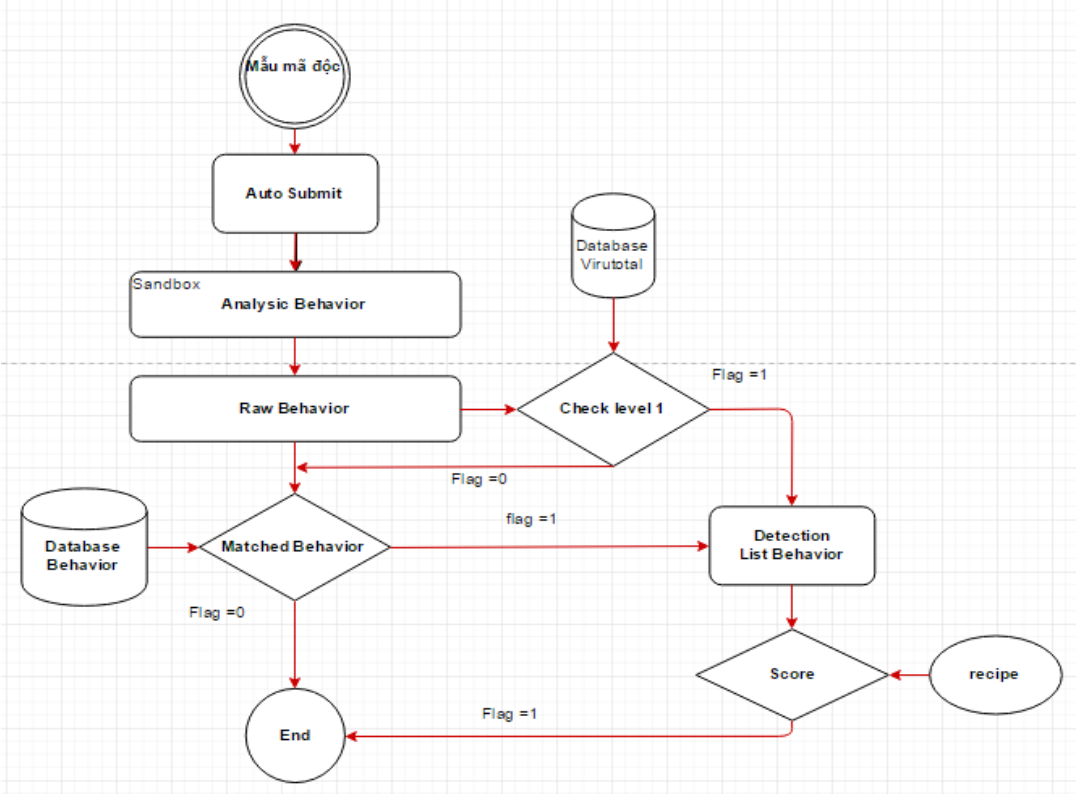


Hình 3.2: Xử lý dữ liệu

b. Xây dựng module phát hiện mã độc

Module phát hiện mã độc với khả năng chính là phát hiện mã độc nó còn có chức năng cập nhật hành vi vào database, tính điểm để đưa ra kết luận và thống kê số liệu khi cần.

Công thức tính điểm được mô tả như sau:



Hình 3.3: Module phát hiện mã độc

Tính điểm đánh giá mức độ nguy hiểm

Chương trình tham khảo cách tính điểm theo nghiên cứu của Robert J. Bagnall và Geoffrey French: "The Malware Rating System (MRS)TM" [15] theo các nội dung sau.

- Tiêu chí đánh giá phần mềm độc hại.
- Các ngưỡng xếp hạng Payload.
- Xếp hạng cho phần mềm độc hại.

Các căn cứ để tính điểm đánh giá gồm:

- Căn cứ vào hành vi của mã độc được đánh giá là nguy hiểm theo mức độ ảnh hưởng vào hệ thống.
- Căn cứ vào các kỹ thuật được sử dụng để xác định mức độ ưu tiên của hành vi (priority).

Table 1: Tiêu chí đánh giá phần mềm độc hại

Bảng 3: Tiêu chí đánh giá phần mềm độc hại

Tiêu chí	Mô tả
Tải trọng tiềm ẩn	Tải trọng tiềm ẩn của module có thể làm suy giảm hoặc làm hỏng mục tiêu
Tiềm năng phát triển	Sự nhanh chóng hoặc dễ dàng để các mã có thể chạy trên hệ thống
Mức độ nguy hại	Mục tiêu ẩn chứa trong payload

Table 2: Các ngưỡng xếp loại Payload

Bảng 4: Đề nghị các ngưỡng xác định xếp loại Payload	
Rating	Mô tả
10	Đa hình, chưa xác định trước
9	Xóa các tập tin cần thiết, lây nhiễm mạng; có thể sụp đổ một mạng do làm tràn băng thông
8	Xóa, sửa đổi hoặc ghi đè lên các tệp tin thiết yếu và lây nhiễm mạng
7	Sửa đổi hoặc ghi đè lên các tập tin không cần thiết, lây nhiễm mạng; Tràn dung lượng băng thông
6	Xóa các tập tin không cần thiết, lây nhiễm mạng
5	Xóa các tệp tin không cần thiết
4	Làm tràn bang thông mạng
3	Lây nhiễm các tệp tin cần thiết
2	Lây nhiễm các tệp tin không cần thiết
1	Không có ảnh hưởng lâu dài

Table 3: Phân loại mức độ nguy hiểm theo điểm

Bảng 5: Phân loại điểm theo mức độ nguy hiểm		
Điểm	Phân loại	Mô tả mức độ nguy hiểm
0-20	1	Tối thiểu
21-40	2	Thấp
41-60	3	Nguy hiểm
61-80	4	Rất nguy hiểm
81-100	5	Thảm họa

Tính mức ưu tiên (Priority) khi đánh giá các hành vi

Luận văn đưa ra cách sắp xếp các hành vi theo mức độ ưu tiên và cách tính điểm đánh giá như sau. Bảng 6 có 8 hành vi đặc trưng của mã độc Ransomware và có tổng điểm ưu tiên là 36 điểm.

Table 4: Tính mức ưu tiên (Priority) khi đánh giá các hành vi

Bảng 6: Mức độ ưu tiên theo hành vi và điểm ưu tiên			
Số TT	**Hành Vi**	**priority**	**score**
Behavior 1	Create new process	1	1/36
Behavior 2	Modifie Regedit	2	2/36
Behavior 3	Creates mutants	4	4/36
Behavior 4	Delete Shadow	5	5/36
Behavior 5	Anti Environment	3	3/36
Behavior 6	Call API Crypt	7	7/36
Behavior 7	Alert	8	8/36
Behavior 8	Connect TOR or BL Network	6	6/36

Công thức tính điểm đánh giá hành vi

Điểm được tính bằng trung bình điểm của các hành vi tương ứng với mức độ ảnh hưởng và được cụ thể hóa bằng điểm ưu tiên (priority). Số lượng các hành vi và tổng các giá trị ưu tiên (priority) sẽ được tính làm trung bình cho điểm với mỗi mẫu mã độc như sau.

$$Score = \sum_{i=1}^{n} (W_i) * 100 \quad (1)$$

Với

$$W_i = \frac{P_{HV_i}}{\sum_{i=1}^{n} P_{HV_i}} \quad (2)$$

Trong đó:

P_{HV_i}: Mức độ ưu tiên tính theo bảng priority

W_i : Điểm trung bình tính theo tổng số các hành vi

Ví dụ cách tính điểm:

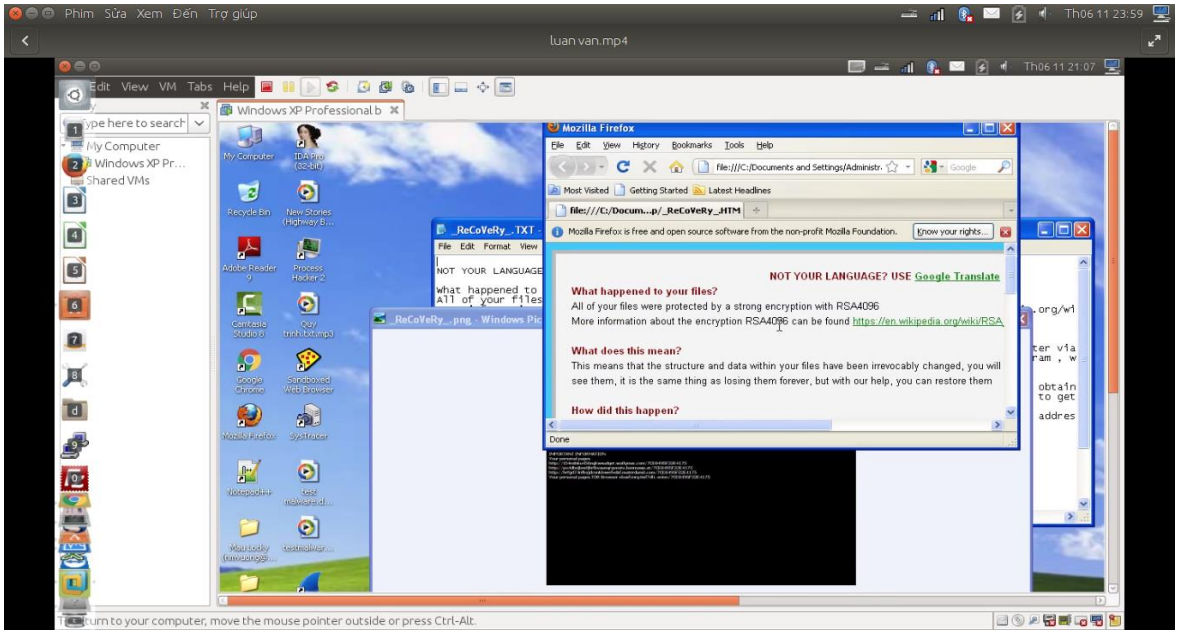
Giả sử mẫu mã độc "A" có các hành vi 1, 3, 6 và 7 tương ứng với Create new process (Score = 1/36), Creates mutants (Score = 4/36), Call API Crypt (Score = 7/36) và Alert (Score = 8/36). Vậy tổng điểm hành vi của mã độc "A" là 20/36 tương ứng với số điểm thực tế là 55.5 điểm.

3.2. Thử nghiệm giải pháp

3.2.1. Kịch bản thử nghiệm 1

Thực hiện phân tích 1 mẫu mã độc cùng họ TestlaCrypt đã biết

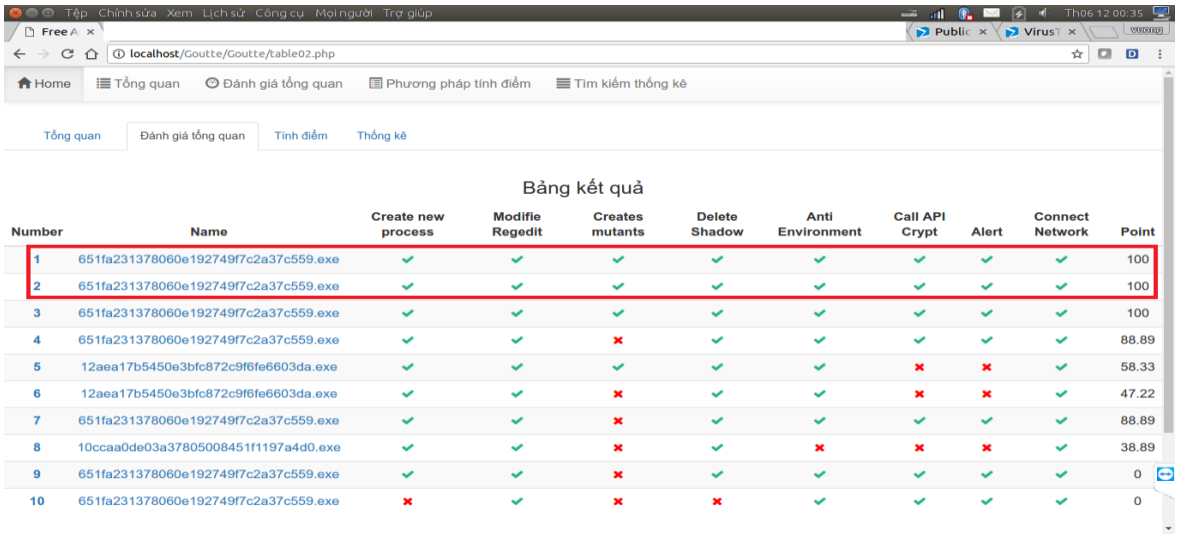
Pha 1: Thực hiện chạy 1 mẫu trên máy ảo để xem quá trình mã hóa file trên máy ảo và quan sát kết quả. Máy ảo được cài đặt trên hệ điều hành Vmware phiên bản 12.0, trên máy ảo được cài đặt một số phần mềm bổ trợ cần thiết gồm: phần mềm adobe, firefox ver39.x, python2.7 library, office 2003, netframework 2.x…



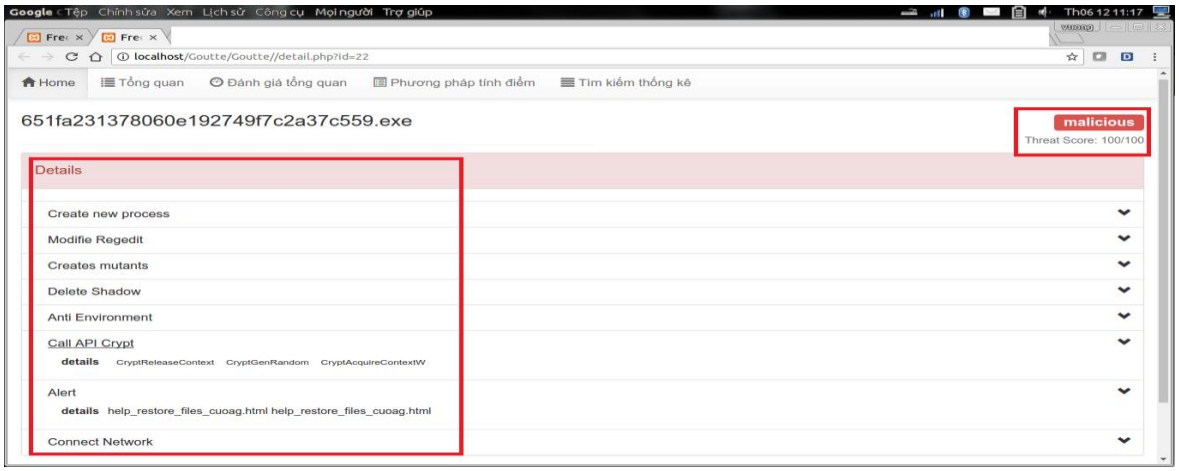
Hình 3.4: Chạy mã độc TeslaCrypt

Pha 2: Thực hiện phân tích hành vi của mẫu mã độc thông qua sandbox, dữ liệu đầu ra sẽ được module phát hiện mã độc sẽ tiến hành phân tích hành vi và đưa ra điểm

số tương ứng với các hành vi đã thu thập được là cơ sở để kết luận tệp tin thực thi là mã độc Ransomware.



Hình 3.5: Chương trình phát hiện mã độc

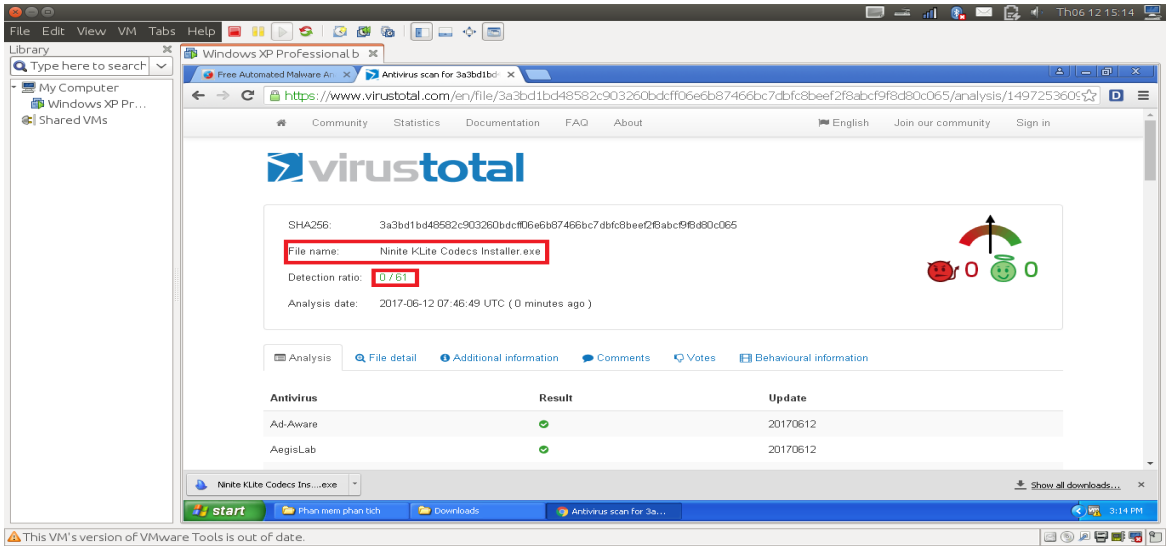


Hình 3.6: Liệt kê các hành vi nguy hiểm

3.2.2. Kịch bản thử nghiệm 2

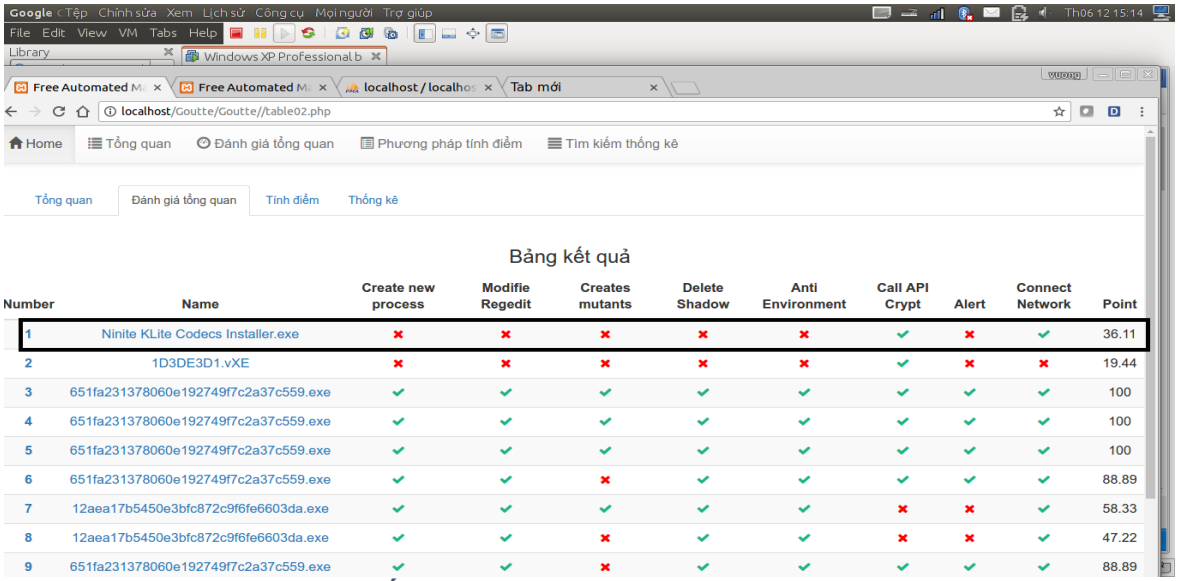
Thực hiện kiểm tra một tệp tin cài đặt có đuôi mở rộng .exe sạch được tải từ trang chủ https://ninite.com/ có tên là: “Ninite KLite Codecs Installer.exe”

Pha 1: Thực hiện chạy mẫu “Ninite KLite Codecs Installer.exe” trên chuyên trang phân tích virustotal cho kết quả đánh giá 0/60. Ý nghĩa của giá trị 0/60 thể hiện tập tin được quét bằng 60 phần mềm phát hiện mã độc và 0 thể hiện đánh giá đây là tệp tin thực thi sạch và không có phần mềm nào kết luận tệp tin có chứa mã độc.



Hình 3.7: Thử nghiệm quét tệp tin trên virustotal

Pha 2: Chạy tệp tin thực thi "Ninite KLite Codecs Installer.exe" [21] trong môi trướng Sandbox và chạy qua giải pháp phát hiện mã độc mã hóa dữ liệu Ransomware. Sau khi được đánh giá các hành vi mà phần mềm tác động lên hệ thống với phương pháp tính điểm của chương trình phát hiện mã độc Ransomware. Số điểm đánh giá phần mềm là 36.11.



Hình 3.8: Kết quả chạy trên chương trình thử nghiệm

Hình 3.9: Hành vi của phần mềm thực thi

Giải thích điểm số đánh giá phần mềm

Phần mềm có hành vi gọi các hàm mã hóa và giải mã vì đây là một file thực thi đã được mã hóa trước khi tải về từ trang web https://ninite.com. Tệp tin cần phải giải mã để cài đặt vào máy người dùng chính vì vậy sử dụng thư viện giải mã CryptUnprotectData. Sau khi phát hiện hành vi này hệ thống tính điểm căn cứ theo hành vi số 7 trong bảng 6 sẽ tính điểm cho hành vi này là 7/36. Phần mềm "Ninite KLite Codecs Installer.exe" tiếp tục gửi lênh GET đến trang web http://www.majorgeeks.com/ để tải tệp tin "k_lite_codec_pack_full" và cài đặt vào máy tính người dùng đây là hành vi GET hoặc POST dữ liệu nào đó từ bên ngoài tương ứng với hành vi số 8 trong bảng 6 và nhận điểm số 6/36. Như vậy tổng điểm số đánh giá tệp tin "Ninite KLite Codecs Installer.exe" sẽ được tính theo công thức (1) và có số điểm là 36,11. Với số điểm này hệ thống đánh giá mối nguy hại từ phần mềm ở mức thấp. Sau khi phân tích có thể kết luận được đây là tệp tin thực thi sạch.

3.2.3. Đánh giá thử nghiệm và kết luận

Quá trình thử nghiệm thực hiện phân tích 2 mẫu, một mã độc và một mẫu không phải là mã độc, kết quả chương trình phát hiện được chính xác được tệp tin độc hại và đạt yêu cầu bài toán đặt ra.

Giải pháp đã đề xuất được đánh giá tóm tắt trong bảng sau.

Đạt mục tiêu như thiết kế	Chưa đạt được như mục tiêu
1. Hệ thống có khả năng xử lý chính xác các dữ liệu cần lấy theo các tiêu chí đã được đặt ra. 2. Hệ thống có khả năng phát hiện các hành vi nguy hiểm dựa vào dữ liệu báo cáo phân tích từ hệ thống Sandbox. 3. Có khả năng giảm thiểu được phát hiện sai thông qua cơ chế tính điểm theo mức độ nguy hiểm của từng hành vi.	1. Thời gian xử lý chậm, chưa có chức năng báo cáo cho quản trị viên. 2. Chưa thu thập được nhiều họ mã độc để thực hiện phân loại mã độc cũng như so sánh các hành vi để tạo bộ dữ liệu mẫu đủ lớn, đủ đa dạng, giảm thiểu phát hiện sai và áp dụng học máy để tăng chính xác và khả năng phân tích số lượng lớn mã độc.

.

KẾT LUẬN

1. Kết quả đạt được

Qua quá trình nghiên cứu, luận văn đã đạt được các kết quả nghiên cứu chính như sau:

- Đã nghiên cứu về mã độc Ransomware, biện pháp nhận biết và phòng chống mã độc, một số phương pháp phát hiện nhanh mã độc.
- Đã nghiên cứu, phân tích, đánh giá hai phương pháp phân tích mã độc điển hình là: phân tích tĩnh và phân tích động.
- Phân tích, lựa chọn công cụ, phương pháp phân tích hành vi mã độc Ransomware. Nghiên cứu thiết lập môi trường phân tích mã độc. Đề xuất được một mô hình cụ thể cho phân tích hành vi mã độc.
- Thu thập mẫu mã độc, nghiên cứu các hành vi, hoạt động của một số loại mã độc. Đưa ra mức ưu tiên và tiêu chí, cách tính điểm tiêu chí đánh giá mức độ nguy hiểm của mã độc.
- Xây dựng được một giải pháp phát hiện mã độc Ransomware, cụ thể là một phần mềm phân tích dựa trên hành vi và phân tích heuristic gồm các mô đun: xử lý dữ liệu mẫu mã độc thu thập được, lưu giữ mẫu, đánh giá mức nguy hiểm và phát hiện mã độc.
- Thử nghiệm giải pháp.
- Trong quá trình nghiên cứu, học viên đã viết và gửi bài báo khoa học có tiêu đề "Phương pháp kết hợp cho phân tích mã độc Ransomeware" đến tạp chí an toàn thông tin ngày 21-5-2017.

Kết quả thực hiện đề tài nghiên cứu có ý nghĩa về mặt khoa học và thực tiễn. Về mặt khoa học, luận văn đã có nghiên cứu về cơ sở lý thuyết phân tích mã độc, so sánh đánh giá hai phương pháp phân tích động và tĩnh; đề xuất một mô hình cụ thể và chương trình phân tích hành vi mã độc, phát hiện mã độc theo mức ưu tiên và đánh giá mức nguy hiểm.

Về mặt thực tiễn, kết quả nghiên cứu đưa ra một giải pháp phân tích, phát hiện mã độc Ransomware hiệu quả, thay thế một phần kiến thức chuyên gia, dễ sử dụng, có khả năng làm chủ công nghệ, áp dụng được vào thực tiễn.

2. Một số hạn chế

- Hạn chế số lượng các mẫu mã độc, chưa đa dạng về chủng loại, điều này có khả năng gây ảnh hưởng đến kết quả.

- Luận văn tập trung phân tích hai mẫu Ransomware phổ biến tại Việt Nam và chưa thực hiện được trên một số mẫu Ransomware khác.

3. Hướng phát triển

- Thu thập nhiều mẫu và biến thể nhằm xây dựng bộ hành vi đặc trưng hoàn chỉnh hơn. Khi đã có bộ dữ liệu đủ lớn có khả năng áp dụng học máy để tăng hiệu xuất phát hiện Ransomware.

- Nghiên cứu công nghệ phân tích tĩnh tự động để nâng cao hiệu xuất cũng như chất lượng của tập hành vi mẫu, làm cơ sở phát triển các giải pháp ngăn chặn Ransomware.

DANH MỤC TÀI LIỆU THAM KHẢO

[1] Pearson Education, Inc (2004), The Tao Of Network Security Monitoring

[2] Practical Malware Analysis (2012), The Hands-On Guide to Dissecting Malicious Software by Michael Sikorski and Andrew Honig.

[3] AAE Elhadi, MA Maarof, et.al. Malware detection based on hybrid signature behaviour application programming interface call graph. American Journal of Applied Sciences, 2012

[4] AD Schmidt, SA Camtepe, S Albayrak. Static smartphone malware detection. eprints.qut.edu.au. 2010.

[5] M Wagner, F Fischer, R Luh, A Haberson. A Survey of Visualization Systems for Malware Analysis. Eurographics Conference on Visualization (EuroVis) 2015.

[6] Y Cao, Q Miao, J Liu, W Li. Osiris: a malware behavior capturing system implemented at virtual machine monitor layer. Mathematical Problems in Engineering, 2013.

[7] KS Han, BJ Kang, EG Im. Malware analysis using visualized image matrices. The Scientific World Journal, 2014

[8] S Cesare, Y Xiang, W Zhou. Malwise&# x2014; an effective and efficient classification system for packed and polymorphic malware. IEEE Transactions on Computers, 2013

[9] LX Min, QH Cao. Runtime-based behavior dynamic analysis system for android malwaredetection. Advanced Materials Research, 2013.

[10] S Feldman, D Stadther, B Wang. Manilyzer: automated android malware detection through manifest analysis. IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2014 .

[11] J Jang, HK Kim. Function-Oriented Mobile Malware Analysis as First Aid. Mobile Information Systems, 2016.

[12] DF Guo, JJ Hu, AF Sui, GZ Lin, T Guo . The Abnormal Mobile Malware Analysis Based on Behavior Categorization. Advanced Materials Research (Volumes 765-767). 9/2013.

[13] K Rami, V Desai . Performance Base Static Analysis of Malware on Android. International Journal of Computer Science and Mobile Computing. 9/2013, p.247-255.

[14] S Naval, V Laxmi, MS Gaur, P Vinod . ESCAPE: Entropy score analysis of packed executable. Proceedings of the Fifth International Conference on Security of Information and Networks. Pages 197-200. 2012.

[15] Robert J. Bagnall, Geoffrey French: The Malware Rating System (MRS). 2015 (Track7/105_tr7)

Một số website

[17] https://www.vxstream-sandbox.com

[18] https://www.wireshark.org

[19] https://www.hex-rays.com/products/ida/

[20] https://blog.kaspersky.com/tag/ransomware/

[21] https://ninite.com