**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**SINH VIÊN: HOÀNG VĂN TUẤN**

**NGHIÊN CỨU, XÂY DỰNG HỆ THỐNG PHÂN TÍCH MALWARE**

**Chuyên nghành: An toàn thông tin**

**ĐỒ ÁN TỐT NGHIỆP**

**Hà Nội – Năm 2015**

**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**ĐỒ ÁN TỐT NGHIỆP**

**Tên đề tài:**

**Nghiên cứu, xây dựng hệ thống phân tích Malware**

**Thời gian:** 6 tháng

**Chuyên ngành**: An toàn thông tin

**Học viên thực hiện**: Hoàng Văn Tuấn

Nguyễn Thị Xoa

**Lớp**: ATTT10

**Cán bộ hướng dẫn:**

**Họ và tên**: Nguyễn Quang Uy

**Học hàm, học vị**: Tiến sĩ

**Đơn vị**: Bộ môn an ninh mạng - Khoa CNTT – HVKTQS

**Mục Lục**

[**Danh Mục hình ảnh** 2](#_Toc436306914)

[**Danh mục từ viết tắt** 3](#_Toc436306915)

[LỜI MỞ ĐẦU 1](#_Toc436306916)

[**Ý nghĩa, mục đích và yêu cầu** 1](#_Toc436306917)

[**Nội dung đề tài** 2](#_Toc436306918)

[**Chương 1: Tổng quan về hệ thống phân tích malware** 4](#_Toc436306919)

[**1.** **Giới thiệu:** 4](#_Toc436306920)

[**2.** **Tổng quan về malware** 4](#_Toc436306921)

[**Khái niệm**: 4](#_Toc436306922)

[**Mục đích của malware:** 4](#_Toc436306923)

[**Phân loại** 4](#_Toc436306924)

[**Cách thức lây nhiễm malware** 6](#_Toc436306925)

[**Các loại malware mới và xu hướng phát triển malware trong tương lai:** 6](#_Toc436306926)

[**3.** **Tổng quan về các kỹ thuật phân tích malware** 8](#_Toc436306927)

[**3.1.** **Kỹ thuật phân tích tĩnh:** 8](#_Toc436306928)

[**3.2.** **Kỹ thuật phân tích động** 8](#_Toc436306929)

[**3.3.** **So sánh các kỹ thuật phân tích malware** 8](#_Toc436306930)

[**4.** **Tổng quan về hệ thống phân tích malware** 9](#_Toc436306931)

[**4.1.** **Tổng quan và đánh giá các hệ thống hiện nay** 9](#_Toc436306932)

[**4.2.** **Nghiên cứu về sandbox và công cụ cuckoo sandbox** 11](#_Toc436306933)

[**4.3.** **Hệ thống cần xây dựng** 13](#_Toc436306934)

[**5.** **Kết luận chương** 13](#_Toc436306935)

[**Chương 2:** **Nghiên cứu các phương pháp phân tích Malware** 14](#_Toc436306936)

[1. Giới thiệu 14](#_Toc436306937)

[2. Kỹ thuật phát hiện malware của phần mềm phòng chống malware 14](#_Toc436306938)

[2.1. Nguyên lý phát hiện Malware 14](#_Toc436306939)

[2.2. Các kỹ thuật phát hiện Malware 14](#_Toc436306940)

[3. Kỹ thuật phân tích tĩnh Malware 14](#_Toc436306941)

[3.1. Kỹ thuật phân tích tĩnh Malware cơ bản 14](#_Toc436306942)

[3.2. Kỹ thuật phân tích tĩnh Malware nâng cao. 14](#_Toc436306943)

[4. Kỹ thuật phân tích động Malware 14](#_Toc436306944)

[4.1. Kỹ thuật phân tích tĩnh Malware cơ bản 14](#_Toc436306945)

[4.2. Kỹ thuật phân tích tĩnh Malware nâng cao. 14](#_Toc436306946)

[5. Kết luận chương 14](#_Toc436306947)

[**Chương 3:** **Xây dựng hệ thống phân tích Malware** 14](#_Toc436306948)

[1. Giới thiệu 14](#_Toc436306949)

[2. Thiết lập môi trường 14](#_Toc436306950)

[3. Nghiên cứu và sử dụng Cuckoo Sanbox 14](#_Toc436306951)

[4. Xây dựng website truy cập 14](#_Toc436306952)

[5. Kết nối hệ thống 14](#_Toc436306953)

[6. Kết luận chương 14](#_Toc436306954)

[**Chương 4: Kết quả thử nghiệm, đánh giá:** 15](#_Toc436306955)

[1. Kết quả thử nghiệm 15](#_Toc436306956)

[2. Đánh giá và nhận xét 15](#_Toc436306957)

[***Hướng phát triển:*** 15](#_Toc436306958)

[***Tài Liệu Tham Khảo*** 16](#_Toc436306959)

# 

# **Danh Mục hình ảnh**

Hình 1.2: Thống kê các loại malware mới xuất hiện theo năm

Hình 1.4.1: Hệ thống phân tích Malware của Virustotal.com

Hình 1.4.2: Hệ thống phân tích Malware của Malwr.com

# **Danh mục từ viết tắt**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

LỜI MỞ ĐẦU

**Đặt vấn đề**

Trong giai đoạn hiện nay, khi công nghệ thông tin được ứng dụng vào hầu hết các lĩnh vực của cuộc sống, yêu cầu về an toàn thông tin càng trở nên quan trọng hơn. Chính vì vậy, việc phát triển các phần mềm phòng chống Malware (hay phần mềm Antivirus) ngày càng được quan tâm, chú trọng. Như vậy khẳng định tầm quan trọng trong việc phát hiện và ngăn chặn các loại malware. Theo ước tính của Trung tâm Nghiên cứu Chiến lược và Quốc tế (CSIS) Hoa Kỳ, trong năm 2014, cả thế giới tổn thất 445 tỷ USD do tội phạm mạng gây ra và 8.500 tỷ đồng là số tiền thiệt hại mà người sử dụng Việt Nam phải hứng chịu từ các loại malware máy tính.

Thông qua nội dung đề tài: “**Nghiên cứu và xây dựng hệ thống phân tích Malware tự động**”, nhóm nghiên cứu đã tìm hiểu các kiến thức về Malware, các kỹ thuật phân tích malware, nghiên cứu phương pháp để xây dựng hệ thống phân tích Malware tự động. Hệ thống phân tích Malware tự động là một hệ thống kết hợp của các kỹ thuật phân tích Malware (gồm kỹ thuật phân tích tĩnh và kỹ thuật phân tích động).

Hai kỹ thuật phân tích Malware đều có những ưu điểm riêng trong quá trình phân tích malware, nhưng sẽ không có kết quả tốt nhất nếu chỉ sử dụng một kỹ thuật đơn lẻ trong việc phân tích Malware. Như vậy, ta cần phải sử dụng kết hợp cả hai kỹ thuật phân tích này để có kết quả tốt nhất trong quá trình phân tích.

## **Ý nghĩa, mục đích và yêu cầu**

Ý nghĩa:

* Ý nghĩa khoa học:
  + Tìm hiểu tổng quan về Malware, phân loại và cách thức lây lan Malware.
  + Nghiên cứu về các kỹ thuật phân tích Malware.
  + Tìm hiểu một số kỹ thuật phát hiện malware của các phần mềm phòng chống malware.
  + Nghiên cứu phương pháp để xây dựng hệ thống phân tích malware tự động.
* Ý nghĩa thực tiễn:
  + Có khả năng ứng dụng trong các bài toán phát hiện, nhận dạng malware
  + Có khả năng ứng dụng vào các lĩnh vực liên quan đến an toàn thông tin

Mục đích:

* Nghiên cứu tổng quan về malware và các kỹ thuật phân tích Malware.
* Nghiên cứu và xây dựng hệ thống phân tích malware tự động.

Yêu cầu:

* Nghiên cứu tổng quan về malwarevà các kỹ thuật phân tích Malware.
* Xây dựng hệ thống thử nghiệm phân tích Malware tự động.

## **Nội dung đề tài**

Nội dung của đề tài được trình bày thành 7 phần:

***Lời mở đầu***

***Chương 1: Tổng quan về hệ thống phân tích malware***. Chương 1 trình bày khái quát về Malware, khái niệm, mục đích, phân loại và các hình thức lây lan của các loại Malware. Qua đó đánh giá mối đe họa và sự phát triển của malware trong tương lai. Tìm hiểu tổng quan về các kỹ thuật phân tích malware, và các hệ thống phân tích malware hiện nay.

***Chương 2: Nghiên cứu các phương pháp phân tích Malware***.

Chương 2 tập trung nghiên cứu hai kỹ thuật phân tích malware là kỹ thuật phân tích tĩnh và kỹ thuật phân tích động. Bên cạnh đó tìm hiểu thêm các kỹ thuật phát hiện malware mà phần mềm phòng chống malware sử dụng để phục vụ xây dựng hệ thống phân tích Malware tự động.

***Chương 3: Nghiên cứu và xây dựng hệ thống phân tích Malware tự động*.**

Chương 3 tập trung nghiên cứu phương pháp để xây dựng hệ thống phân tích malware dựa trên những nghiên cứu về các kỹ thuật phân tích malwre của các chương I, II. Nghiên cứu các hệ thống đã có, đánh giá và triển khai xây dựng hệ thống nghiên cứu. Thực hiện thiết lập môi trường, xây dựng hệ thống và website cho phép người dùng upload file lên hệ thống.

***Chương 4: Kết quả thử nghiệm và đánh giá***.

Trong chương này ta sẽ đi vào thử nghiệm và đánh giá hệ thống mà ta đã xây dựng ở chương V và đưa ra nhận xét và đánh giá những ưu điểm của hệ thống này cũng như những điểm yếu còn tồn tại để có định hướng phát triển trong tương lại.

***Kết luận và hướng phát triển.***

***Tài liệu tham khảo***

# **Chương 1: Tổng quan về hệ thống phân tích malware**

## **Giới thiệu:**

Trong chương này sẽ giới thiệu tổng quan về những vẫn đề mà đề tài nghiên cứu: khái niệm malware, phân loại và mục đích của malware. Tìm hiểu các loại malware mới, và nhận định hướng phát triển của malware trong tương lại. Tổng quan về các kỹ thuật phân tích malware và các hệ thống phân tích malware hiện nay. Từ đó, có định hướng cho việc xây dựng hệ thống phân tích malware.

1. **Tổng quan về malware**

### **Khái niệm**:

Malware là một thuật ngữ dùng để chỉ những phần mềm độc hại hay đoạn mã độc hại mà gây thiệt hại, làm mất kiểm soát, phá hoại , hoặc đánh cắp thông tin từ hệ thống máy tính[1]. Malware bao gồm các loại như: botnet, virus, worm, trojan horses, logic bomb, rootkit, bootkit, back-doors, spyware, và adware…

Hiện nay, malware 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…

### **Mục đích của malware:**

* Trò đùa giữa những người lập trình malware
* Khẳng định bản thân
* Học tập, nghiên cứu
* Phá hủy dữ liệu, hệ thống máy tính
* Theo dõi và đánh cắp thông tin máy tính, thông tin cá nhân người dùng…

### **Phân loại**

*Virus*

Virus là một dạng của malware có bản chất tự sao chép vào một tập tin thực thi. Nó được xây dựng để sửa đổi hoặc ngăn chặn các hoạt động của một máy tính. Khi chương trình được thực thi virus sẽ cố gắng để chèn một phiên bản của chính nó vào chương trình [2]. Virus không lây lan qua mạng. Virus có thể xác định được các vị trí của mình trong từng khu vực khác nhau của hệ thống. Virus có thể được chia thành nhiều dạng như: Encrypted Virus, Polymorphic Virus, Metamorphic Virus.

*Trojan*

Trojan là một dạng cuả malware, thường giả mạo là một phần mềm vô hại với người dùng nhưng nó thực hiện các hoạt động độc hại ở chế độ ẩn bên trong. Không như virus, trojan là một chương trình độc lập, nó không tự sao chép mà phải phụ thuộc vào người kích hoạt [3]. Sau khi cài đặt, trojan có thể tải về phần các phần mềm khác để bổ sung, sửa đổi các thiết lập của hệ thống hay lây nhiễu sang các file khác của hệ thống. Trojan thực sự rất nguy hiểm nếu như người dùng không hiểu biết về nó và nó chiếm số lượng lớn trong các hoạt động gây hại mà malware gây ra.

*Worm*

Worm là một dạng của malware, khác với virus nó có thể lây lan sang các máy tính khác thông qua mạng. Worm không ký sinh như virus mà nó hoạt động như một chương trình độc lập [4]

*Backdoor*

Backdoor là một dạng malware được sử dụng để vượt qua quy trình xác thực bảo mật thông thường. Sau khi đạt được quyền truy cập, nó sẽ kiểm soát hệ thống và có thể điều khiển từ xa, theo dõi ăn cắp thông tin cá nhân [5].

*Spyware*

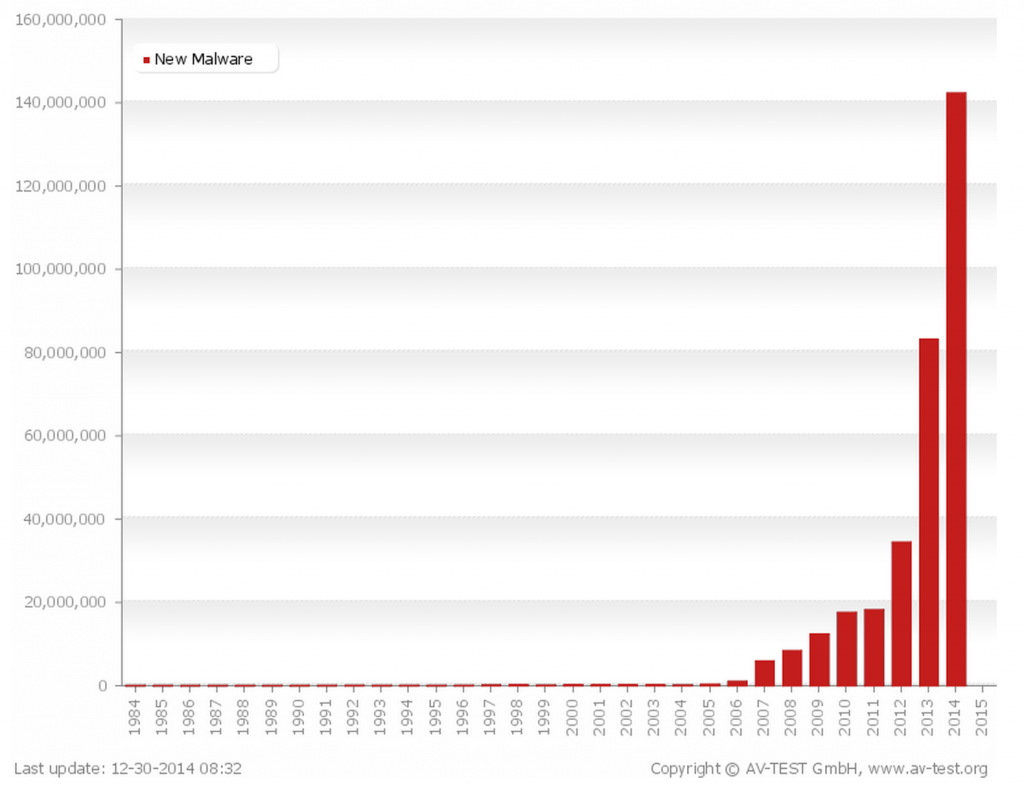
Spyware là phần mềm gián điệp ăn cắp thông tin nhạy cảm từ hệ thống của nạn nhân và chuyển thông tin này về cho những kẻ tấn công. Các thông tin có thể phục vụ tốt cho kẻ tấn công bao gồm tài khoản các hệ thống máy tính, tài khoản ngân hàng hay các nội dung nhạy cảm khác.

### **Cách thức lây nhiễm malware**

* *Lây nhiễm theo cách cổ điển*: malware lây nhiễm từ máy tính này sang máy tính khác, từ thiết bị này sang thiết bị khác thông qua các phương tiện lưu trữ dữ liệu như: đĩa CD, USB flash, Ổ cứng di động…
* *Lây nhiễm qua thư điện tử:* malware được gắn vào các thư điện tử, các tệp tin đính kèm, các đường link ẩn… người sử dụng khi nhận được thư điện tử đã không kiểm tra kỹ, gián tiếp thực thi malware.
* *Lây nhiễm qua truy cập trang web*: malware được gắn trên các banner, quảng cáo của các trang web, người sử dùng sẽ trực tiếp hoặc gián tiếp truy cập vào các trang web bị dính malware và bị lây nhiễm.
* *Lây nhiễm qua các phần mềm cài đặt*: malware được đính kèm trong các phần mềm cài đăt, người sử dụng sẽ bị lây nhiễm khi thực hiện quá trình cài đặt hoặc sau khi cài đặt phần mềm. Các lây nhiễm này thường dựa vào sự thiếu hiểu biết hoặc không đọc kỹ các yêu cầu cài đặt của phần mềm.
* *Lây nhiễm qua khai thác lỗi bảo mật*: thông qua việc khai thác các lỗi bảo mật trên hệ điều hành, cấu trúc tệp tin… để cài các malware nên máy tính nạn nhân.

### **Các loại malware mới và xu hướng phát triển malware trong tương lai:**

Theo AV-Test, có đến hơn 140 triệu loại malware mới xuất hiện trong năm 2014, nhiều hơn hẳn so với các năm trước. Bên cạnh sự vượt trội về số lượng, cũng có thể thấy năm 2014 xuất hiện nhiều malware có đặc điểm khác biệt so với các malware trước đây. Các loại malware mới ngày càng được thiết kế tinh vi hơn, đặc biệt các malware được thiết kế cho mục đích tấn công có chủ đích nhằm vào cá nhân hoặc các tổ chức lớn.



Hình 1.2: Thống kê các loại malware mới xuất hiện theo năm.

Dự đoán trong những năm tới xu hướng phát triển của malware sẽ tiếp tục xuất hiện các phần mềm gián điệp tinh vi nhằm vào hạ tầng trọng yếu của các quốc gia để đánh cắp và phá huỷ. Các malware trên di động tiếp tục tăng nhanh và sẽ xuất hiện nhiều biến thể phần mềm mã hoá tống tiền (ransomware) trên di động. Quyền riêng tư của người sử dụng tiếp tục là vấn đề nóng đặc biệt là trên các hãng sản xuất thiết bị di động hoặc các ứng dụng di động. Có thể xuất hiện nhiều vụ lọ lọt dữ liệu của người sử dụng tương tự như vụ các diễn viên nổi tiếng bị lộ ảnh nóng trên iCloud. Các dịch vụ lưu trữ dữ liệu đám mây gặp phải sự nghi ngại của người sử dụng do vấn đề đảm bảo an toàn. Mạng xã hội như Facebook sẽ trở thành con đường chủ yếu để những kẻ lừa đảo hoạt động. Các cuộc tấn công mạng mang màu sắc chính trị gia tăng giữa nhóm hacker của các quốc gia. Hacker tiếp tục nhằm vào các giao thức mạng hoặc thư viện quan trọng với hy vọng tìm ra những lỗ hổng tương tự Heartbleed hay Shellshock. Doanh nghiệp đối đầu với các cuộc tấn công đánh cắp dữ liệu ngày càng tinh vi. Đặc biệt là các tổ chức tài chính và cung cấp dịch vụ internet.

1. **Tổng quan về các kỹ thuật phân tích malware**
   1. **Kỹ thuật phân tích tĩnh:**

Kỹ thuật phân tích tĩnh là kỹ thuật phân tích code và cấu trúc của chương trình mà không phải thực thi [6]. Như vậy, khi phân tích ta không cần phải thực thi tập tin mà vẫn cho phép ta đưa ra những thông tin tổng quan nhất về một tập tin bất kỳ (gồm địnhg dạng file, ngày tạo,…), từ đó nhận định tập tin đó có phải là Malware hay không dựa vào những dấu hiệu đặc trưng.

Kỹ thuật này được áp dụng trong tất cả các phần mềm phòng chống Malware hiện nay. Nhưng nếu chỉ sử dụng kỹ thuật phân tích tĩnh sẽ không đưa ra được một kết luận đúng nhất, vì kỹ thuật này sẽ không đưa ra được kết luận chính xác khi ta phân tích một loại malware mới chưa có đặc điểm nhận dạng.

* 1. **Kỹ thuật phân tích động**

Kỹ thuật phân tích động là kỹ thuật phân tích hoạt động của một tập tin khi nó được thực thi trên bộ nhớ [7]. Kỹ thuật này sử dụng các công cụ debug chương trình, công cụ theo dõi tiến trình hoạt động, công cụ giám sát mạng,…Sử dụng kỹ thuật này cho phép ta nhận định chính xác được malware khi quan sát các hành vi trong quá trình thực thi từ đó có thể kết luận chính xác. Nhưng khi thực thi tập tin Malware nguy hiểm sẽ có những nguy cơ ảnh hưởng đến máy tính của người phân tích và có thể sẽ không thu được kết quả gì mà máy tính đã bị Malware phá hoại.

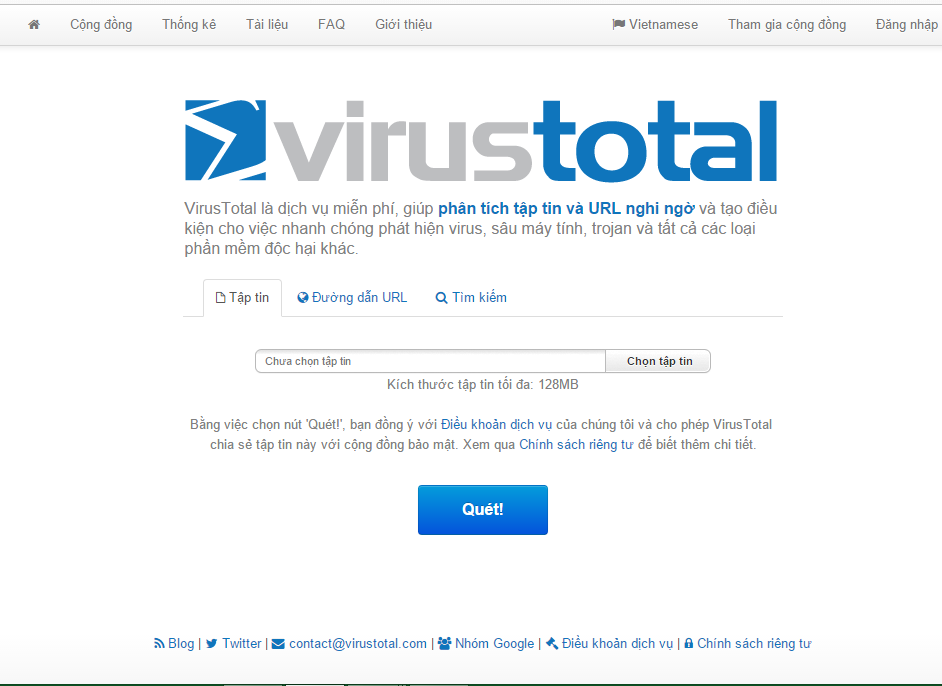
* 1. **So sánh các kỹ thuật phân tích malware**

|  |  |  |
| --- | --- | --- |
|  | Ưu điểm | Nhược điểm |
| Kỹ thuật phân tích tĩnh | + Không thực hiện chạy Malware, giảm nguy cơ lây lan, phá hoại của malware.  + Dựa trên cấu trúc file để phân tích, tìm điểm nghi ngờ trong file để xác định malware. | + Khó phân tích khi file nghi ngờ malware bị nén hay mã hóa.  + Không phát hiện được loại malware mới. |
| Kỹ thuật phân tích động | + Giám sát được các hành vi thực thi của file malware trên hệ thống từ đó xác định chính xác loại malware.  + Đánh giá được các ảnh hưởng thực tế đến hệ thống. | + Khi phân tích động thường phải thực thi chương trình nên tiềm ẩn nguy cơ ảnh hưởng đến hệ thống.  + Một số loại malware mới có cơ chế chống debug, phát hiện máy ảo. |

1. **Tổng quan về hệ thống phân tích malware**
   1. **Tổng quan và đánh giá các hệ thống hiện nay**

Trên thực tế đã có những hệ thống được triển khai áp dụng những kỹ thuật phân tích trên như:

**a. virustotal**.com (<https://www.virustotal.com/> ), hệ thống này cho phép người dùng có thể gửi mẫu lên server và server thực hiện quá trình phân tích và trả về kết quả cho người dùng xem liệu tập tin đó có phải là một tập tin malware hay không.



*Hình 1.4.1: Hệ thống phân tích Malware của Virustotal.com*

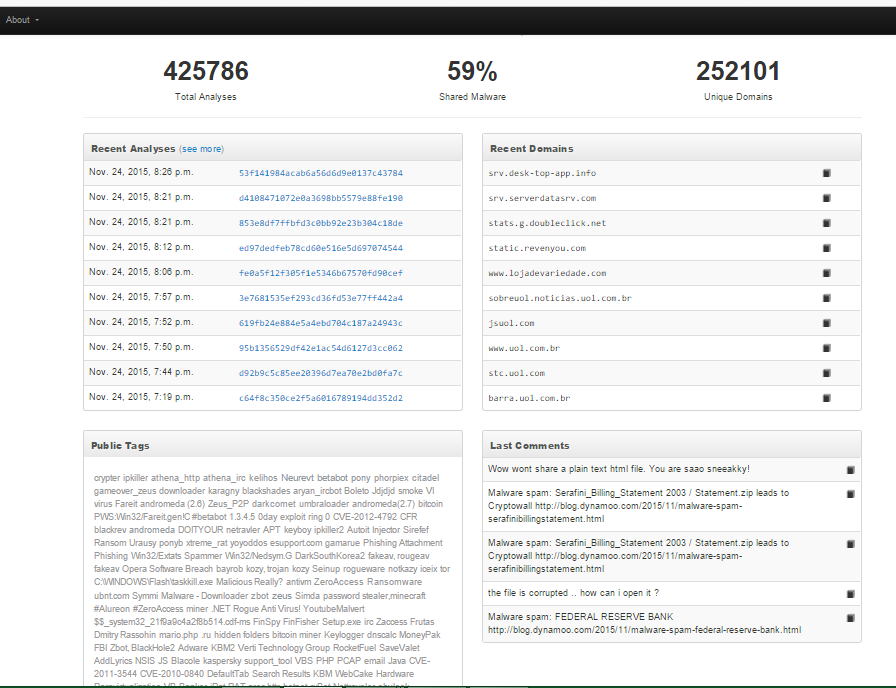
**Ưu điểm:**

Hệ thống này sử dụng rất nhiều các phần mềm antivirus để quét tập tin và trả về kết quả cho người dùng ,việc sử dụng rất nhiều các phần mềm antivirus như vậy sẽ đưa ra được một kết quả khách quan hơn, và việc đánh giá sẽ tốt hơn, sự tin cậy của người dùng sẽ cao hơn.

**Nhược điểm:**

Nhưng thực tế, hệ thống này có nhiều nhược điểm, các antivirus đưa ra kết quả rất khác nhau gây ra tình trạng không định danh rõ ràng được tập tin đó là loại malware nào, và khi có loại malware mới trong khi các phân mềm antivirus chưa cập nhật được mẫu đó trong cơ sở dữ liệu và sẽ kết luận nhầm. Đặc biệt, hệ thống này không đưa ra được thông tin chi tiết về các hành vi gây hại của tập tin đó.

**b. malwr**.com ( <https://malwr.com/> ), hệ thống này cũng là một hệ thống phân tích malware tự động có ưu điểm hơn hệ thống của virustotal.com. Nhóm nghiên cứu dựa trên hệ thống này để có thể xây dựng một hệ thống phân tích malware tự động. Hệ thống này cũng cho phép người dùng gửi file lên để phân tích, và có thể tìm kiếm thông tin về những loại malware khác nhau.



*Hình 1.4.2: Hệ thống phân tích Malware của Malwr.com*

**Ưu điểm:**

Ưu điểm của hệ thống này so với hệ thống virustotal là ta có thể xem chi tiết các các hành vi thực thi của tập tin ta gửi lên, như vậy có thể nhận định chính xác hơn về mực độ gây hại cho máy tính, đánh giá được tổn thương nhất định nếu tập tin đó được thực thi thực sự trên máy tính nạn nhân, từ đó ta có thể đưa ra được cách thức loại bỏ loại Malware đó.

**Nhược điểm:**

Hệ thống này không đưa ra cho người dùng được kết luận cuối cùng về định danh của loại Malware đó, mà chỉ đưa thông tin phân tích cho người dùng, như một bản báo cáo các hành vi của tập tin đó.

* 1. **Nghiên cứu về sandbox và công cụ cuckoo sandbox**

**Sandbox**

Sandbox là một cơ chế bảo mật dùng để tách các chương trình đang chạy, thường được sử dụng để thực thi mã chưa được kiểm tra, hoặc chương trình không đáng tin cậy và chưa được xác minh từ các bên thứ ba, các nhà cung cấp phân mềm, và các trang web không đáng tin cậy[8].

Sandbox được áp dụng trong việc phân tích phần mềm độc hại, dùng để chạy các ứng dụng không rõ nguồn gốc và không đáng tin cậy bên trong một môi trường cô lập và sau đó theo dõi và nhận được thông tin về các hành vi thực thi của chương trình đó.

Sử dụng sandbox có nhiều ưu điểm tốt, nó hỗ trợ việc phân tích động rất tốt, nhưng có khuyết điểm là ta sẽ phải cô lập môi trường chạy của malware chỉ trong sandbox, nên sẽ không có được tất cả thông tin về hành vi hoạt động của malware. Như vậy, ta cần một hệ thống phải sử dụng cả kỹ thuật phân tích động và phân tích tĩnh, để có được thật nhiều thông tin về malware hơn.

**Cuckoo**

Là hệ thống mã nguồn mở về phân tích malware, được sử dụng để tự động thực thi và phân tích các file và thu thập kết quả phân tích của malware trong quá trình file đó được thực thi trong hệ điều hành windows [9]. Kết quả thu được bao gồm:

* Các hàm API win32 được sử dụng trong quá trình thực thi phần mềm độc hại.
* Các tập tin được tạo ra, xóa và được tải về bởi malware trong quá trình thực thi của nó
* Sự trao đổi mạng trong quá trình thực thi
* Ảnh chụp màn hình windows khi thực thi các phần mềm độc hại.
* Dữ liệu của bộ nhớ khi các tiến trình của phần mềm độc hại thực thi.

**Loại file có thể phân tích**

* File thực thi trong windows
* File DLL
* Tài liệu .PDF
* Tài liệu Microsoft Office
* URL và các tập tin HTML
* PHP script
* File CPL
* File Visual Basic (VB)
* File ZIP
* Java JAR
* File Python…
  1. **Hệ thống cần xây dựng**
* Hệ thống phân tích malware này sử dụng kết hợp hai kỹ thuật phân tích tĩnh và kỹ thuật phân tích động.
* Sử dụng sandbox để tạo môi trường thực thi cô lập malware và ghi lại thông tin về hành vi thực thi của malware
* Sử dụng mã nguồn mở Cuckoo Sanbox để hỗ trợ việc phân tích những thông tin thu được từ sandbox ta thiết lập và đưa ra bản báo cáo cho người dùng.
* Ở đây, hệ thống sẽ xây dựng một website tương tác giữa người dùng và hệ thống, sẽ trực quan hơn dễ dàng cho nhiều người có thể tương tác với hệ thống. Hệ thống hỗ trợ người dùng có thể vào web để gửi file lên hệ thống. Từ đó, hệ thống nhận file và tiến hành phân tích và trả về kết quả cho người dùng.

1. **Kết luận chương**

Từ thực tế hiện nay về sự phát triển không ngừng của các loại malware mới, sự cần thiết phải trong việc phân tích, phát hiện, loại bỏ malware. Hiện nay các nhà phân tích malware vẫn dựa vào các kỹ thuật phân tích tĩnh và phân tích động để có thể phân tích các loại malware. Việc đó sẽ rất tốn thời gian để có thể phân tích một mẫu malware mới. Như vậy, một hệ thống phân tích malware tự động sẽ giúp ích rất lớn trong việc phân tích malware hiện nay.

Hệ thống phân tích malware này sẽ kết hợp các kỹ thuật phân tích malware để có tốc độ phân tích nhanh và kết quả phân tích chính xác hơn, sử dụng mã nguồn mở CuckooSandbox để phát triển hệ thống. Và để có thể xây dựng hệ thống cần phải nghiên cứu sâu hơn về hai kỹ thuật phân tích tĩnh và phân tích động, như vậy hệ thống mới có thể hoàn thiện.

# **Chương 2: Nghiên cứu các phương pháp phân tích Malware**

1. Giới thiệu
2. Kỹ thuật phát hiện malware của phần mềm phòng chống malware
   1. Nguyên lý phát hiện Malware

2.1.1. Phát hiện Malware dựa vào dấu hiệu đặc trưng

2.1.2. Phát hiện mã độc dựa vào điểm bất thường.

* 1. Các kỹ thuật phát hiện Malware

2.2.1. Checksumers

2.2.2. Fuzzy Hashing

2.2.3. Scan String

2.2.4. Code Emulation

2.2.5. Static Heuristic Analysis

1. Kỹ thuật phân tích tĩnh Malware
   1. Kỹ thuật phân tích tĩnh Malware cơ bản
   2. Kỹ thuật phân tích tĩnh Malware nâng cao.
2. Kỹ thuật phân tích động Malware
   1. Kỹ thuật phân tích tĩnh Malware cơ bản
   2. Kỹ thuật phân tích tĩnh Malware nâng cao.
3. Kết luận chương

# **Chương 3: Xây dựng hệ thống phân tích Malware**

1. Giới thiệu
2. Thiết lập môi trường
3. Nghiên cứu và sử dụng Cuckoo Sanbox
4. Xây dựng website truy cập
5. Kết nối hệ thống
6. Kết luận chương

# **Chương 4: Kết quả thử nghiệm, đánh giá:**

1. Kết quả thử nghiệm

Kết quả cần đạt được:

* + Xây dựng hệ thống phân tích malware đầy đủ chức năng: cho phép người dùng upload file lên hệ thống, cho phép tìm kiếm các báo cáo phân tích của những loại malware được lưu trữ trong cơ sở dữ liệu, hiển thị thông tin trực quan của quá trình phân tích file đó.
  + Hệ thống được cài đặt trên hê điều hành Ubuntu, sandbox là các máy ảo được cài hệ điều hành windows.
  + Đảm bảo các kết nối của hệ thống luôn được hoạt động.
  + Duy trì kết nối giữa website và server luôn được đảm bảo.
  + Giao diện web tương tác với người dùng hợp lý và đẹp mắt.
  + Tốc độ phân tích của hệ thống nhanh, khoảng 2 phút.

1. Đánh giá và nhận xét

# ***Hướng phát triển:***

* Trong nghiên cứu này xây dựng hệ thống để phân tích malware tự động, giúp hỗ trợ việc phân tích malware một cách nhanh chóng hơn so với các kỹ thuật phân tích cũ và có thể triển khai thực tế. Như vậy, khi triển khai thực tế thì hệ thống sẽ giúp ta thu thập các mẫu malware mới mà người dùng gửi lên, phân tích và cập nhật vào cơ sở dữ liệu của hệ thống. Điều này rất cần cho các hãng sản xuất phần mềm phòng chống malware hiện nay, việc thường xuyên cập nhật cơ sở dữ liệu malware sẽ làm tăng khả năng phát hiện ra các loại malware nguy hiểm có thể gây hại cho người dùng.
* Ngoài ra hệ thống cũng có thể áp dụng cho các phân mềm phòng chống malware, bằng việc kết nối phần mềm với hệ thống khi có loại malware mới không có trong cơ sở dữ liệu của phần mềm phòng chống malware này thì nó sẽ gửi file này lên hệ thống, từ đó hệ thống sẽ phân tích file này và đưa ra đánh giá file này có phải là malware hay không và định danh cho nó, cũng như đưa ra được phương pháp loại bỏ malware này ra khỏi máy tính.
* Sau khi có được một cơ sở dữ liệu đủ lớn ta có thể tích hợp thêm vào hệ thống các thuật toán phát hiện malware, từ đó có thể phát triển thành hệ thống phát hiện malware với tốc độ cao.

# ***Tài Liệu Tham Khảo***

**[1] John Wiley & Sons – 2012 - Modern Malware For Dummie - tr5.**

**[2], [3], [4], [5] Swapna Vemparala – 2015 - Malware Detection Using Dynamic Analysis – tr4, tr7, tr7, tr8.**

**[6], [7] Michael Sikorski and Andrew Honig – 2012 - PRACTICAL MALWARE ANALYSIS - The Hands-On Guide to Dissecting Malicious Software – tr9.**

**[8], [9] - 02/2015 - Cuckoo Sandbox Book – tr7, tr8.**