 TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

VÀ TRUYỀN THÔNG VIỆT - HÀN

**KHOA KHOA HỌC MÁY TÍNH**

****

**ĐỀ CƯƠNG**

|  |  |
| --- | --- |
| Tên đề tài: | Nghiên cứu về virus máy tính và phát triển phần mềm phát hiện virus máy tính ứng dụng trí tuệ nhân tạo |
| Giảng viên hướng dẫn: | TS. Trần Thế Sơn |
| Sinh viên thực hiện: | Đỗ Thanh Tùng Mã SV: 18IT186  Đỗ Tấn Tĩnh Mã SV: 18IT181 |

**NGHIÊN CỨU KHOA HỌC SINH VIÊN**

Năm học 2021

**MỤC LỤC**

[1. Thông tin chung 3](#_Toc70366303)

[2. Tổng quan tình hình nghiên cứu ở trong nước và nước ngoài 4](#_Toc70366304)

[3. Tính cấp thiết của đề tài 6](#_Toc70366305)

[4. Mục tiêu của đề tài 6](#_Toc70366306)

[5. Phương pháp nghiên cứu 6](#_Toc70366307)

[6. Đối tượng và phạm vi nghiên cứu 7](#_Toc70366308)

[7. Nội dung nghiên cứu 7](#_Toc70366309)

[8. Kết quả nghiên cứu 8](#_Toc70366310)

[9. Sản phẩm 8](#_Toc70366311)

[10. Về các đóng góp của đề tài đến giáo dục và đào tạo, kinh tế xã hội và an ninh quốc phòng 8](#_Toc70366312)

|  |  |
| --- | --- |
| ĐẠI HỌC ĐÀ NẴNG  **TRƯỜNG ĐẠI HỌC CNTT VÀ TRUYỀN THÔNG VIỆT - HÀN** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – tự do – hạnh phúc**  *Đà Nẵng, ngày 15 tháng 04 năm 2021* |

**ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC SINH VIÊN**

*(Năm học 2021)*

# Thông tin chung

+ Tên đề tài:

Nghiên cứu về virus máy tính và phát triển phần mềm phát hiện virus máy tính ứng dụng trí tuệ nhân tạo.

+ Chuyên ngành (*theo nội dung đề tài*):

Mạng máy tính và truyền thông (An ninh mạng)

+ Sinh viên chịu trách nhiệm chính:

* Họ và tên: Đỗ Tấn Tĩnh
* Mã số sinh viên: 18IT181
* Lớp: 18IT3 - Khoa: Khoa học máy tính
* Địa chỉ thường trú: K09/60 Hà Văn Trí, phường Khuê Trung, quận Cẩm Lệ, TP Đà Nẵng.
* Địa chỉ liên lạc: K09/60 Hà Văn Trí, phường Khuê Trung, quận Cẩm Lệ, TP Đà Nẵng.
* Số điện thoại: 0389909772
* Email: dttinh.18it3@vku.udn.vn

+ Các thành viên tham gia:

* Họ và tên: Đỗ Thanh Tùng
* Lớp: 18IT3 - Khoa: Khoa học máy tính
* Điện thoại: 0382352146
* Mã số sinh viên: 18IT186
* Email: [dttung.18it3@vku.udn.vn](mailto:dttung.18it3@vku.udn.vn)

# Tổng quan tình hình nghiên cứu ở trong nước và nước ngoài

Những năm gần đây đã chứng kiến sự phát triển nhanh chóng của các phần mềm độc hại cả về số lượng và chủng loại. Trong năm 1992 số lượng mã độc đã tăng từ 1000 lên 2300, năm 2002 có đến 60000 loại mã độc và các biến thể của chúng được phát hiện. Ngày này số lượng này đã tăng lên trên hơn 847 triệu mẫu mã độc tính đến cuối năm 2018. Trước sự lây lan nhanh chóng và mức độ nguy hiểm của các phần mềm độc hại, trên thế giới đã có rất nhiều nghiên cứu về các phương pháp phát hiện và loại bỏ các phần mềm độc hại ngay khi nó xuất hiện, nhưng nhìn chung các phương pháp này đều xoay quanh hai cách tiếp cận chính để phát hiện đó là dựa vào sự bất thường hay dị thường mà mã độc gây ra đối với hệ thống máy tính và hệ thống mạng hoặcdựa trên dấu hiệu nhận biết virus đó. Để có được các dữ liệu về sự bất thường hoặc dấu hiệu (signature) như đã nêu, người ta thường dựa vào hai kỹ thuật phân tích mã đôc chính: phân tích tĩnh và phân tích động.

Mục tiêu của phân tích tĩnh có thể là mã nhị phân hoặc mã nguồn (theo Christodorescu and Jha - 2003). Trước tiên, một tệp mã độc hại ở dạng thực thi nhị phân (portable executables – PE) cần được giải nén nếu nó được nén bằng công cụ nén nhị phân của bên thứ ba (ví dụ: UPX và ASPack Shell) hoặc được nhúng trong một trình đóng gói tự chế. Để dịch ngược các tệp thực thi, người ta có thể sử dụng các công cụ dịch ngược và kết xuất bộ nhớ. Các công cụ này (ví dụ: IDA Pro) hiển thị mã phần mềm độc hại dưới dạng ngôn ngữ bậc thấp (Assembly). Các công cụ kết xuất bộ nhớ (ví dụ: OllyDump và LordPE) được sử dụng để lấy các mã được bảo vệ nằm trong bộ nhớ chính và kết xuất chúng vào một tệp [cung cấp references]. Kết xuất bộ nhớ khá hữu ích để phân tích các mã độc thực thiđể thu thập các dấu hiệu của mã độc hay virus . Sau khi tệp thực thi được giải nén và giải mã, các câu lệnh gốc có thể được dịch ngược, chẳng hạn như lệnh gọi các hàm API của Windows, byte n-gram, chuỗi, mã opcodes (mã hoạt động) và đồ thị luồng điều khiển. Đó chính là các dấu hiệu tĩnh của một mã độc. Việc phát hiện mã độc dựa cách tiếp cận này được gọi là phát hiện mã độc dựa vào dấu hiệu (signature-based detection) [cung cấp reference]

Trên cơ sở dữ liệu có được từ phân tích tĩnh (sự bất thường hoặc dấu hiệu của virus máy tính), nhiều mô hình phát hiện mã độc dựa vào dấu hiệu đã được đề xuất như ……..[references]

Hầu hết các phần mềm diệt virus đều cung cấp cho người dùng chức năng phát hiện mã độc hoặc virus dựa vào dấu hiệu (ví dụ:…..)[reference]. Các tiếp cận này cho phép ứng dụng phát hiện rất nhanh các virus ngay khi nó xuất hiện. Tuy nhiên, việc xuất hiện các biến thể của virus máy tính có thể làm các dấu hiệu đặc trưng của virus bị sai lệch. Nếu phần mềm phát hiện mã độc chưa được cập nhật kịp thời các dấu hiệu mới thì virus có thể “qua mặt” được sự phát hiện của phần mềm diệt virus.

Vấn đề này có thể được giải quyết bằng một kỹ thuật phân tích mã độc khác, kỹ thuật phân tích động [reference]. Đây là kỹ thuật phân tích mã độc dựa vào việc quan sát mã độc thực thi (trên bộ xử lý thực hoặc ảo) để phát hiện ra các biểu hiện của máy tính khi virus đang hoạt động . Có thể có nhiều cách tiếp cận khác nhau đối với việc phân tích độnggiám sát việc gọi hàm, theo dõi luồng hoặc các tiến trình đang hoạt động trong máy tính, … Tuy nhiên việc thu thập các đáp ứng hoặc biểu hiện của máy tính bằng phân tích động đòi hỏi nhiều thời gian và chiếm tài nguyên của hệ thống vì phải kích hoặc mã độc khi phân tích [references]..

Nhiều phần mềm diệt virus cung cấp chức năng phát hiện mã độc dựa vào kết quả phân tích động cho phép phát hiện và diệt virus ở máy tính ngay cả khi máy tính đã bị nhiễm virus. Chức năng này đươc cung cấp qua lựa chọn deep scan hoặc chức năng tương tự ở rất nhiều các phần mềm diệt virus.

# Tính cấp thiết của đề tài

Ngày nay thời đại công nghệ thông tin, công việc thực hiện với máy tính cá nhân ngày càng nhiều, đặc biệt là việc trao đổi tệp tin với nhau ngày càng dễ dàng và phổ biến. Tuy nhiên sự thuận tiện ấy cũng tiềm ẩn nhiều nguy bị cơ tấn công từ các tổ chức bất chính. Theo số liệu thống kê của các tổ chức an ninh mạng như Kaspersky, Symantec, số lượng tấn công có chủ đích (APT) thông qua mã độc trong những năm gần đây ngày một gia tăng cả về số lượng lẫn mức độ nghiêm trọng. Vì vậy, việc phát hiện các mã độc để hạn chế tối đa tác hại của chúng lên hệ thống máy tính là rất cần thiết nhằm đảm bảo an toàn, an ninh mạng. Đặc biệt trong xu thế chuyển đổi số đang diễn ra mạnh mẽ ở nước ta trên nhiều lĩnh vực và cấp độ khác nhau, việc phát hiện và phòng ngừa các cuộc tấn công bằng mã độc có một vai trò hết sức quan trọng.

Vì lẽ đó, đề tài nhằm mục đích nghiên cứu các loại mã độc, các kỹ thuật phát hiện (phân loại) mã độc dựa vào các đặc trưng của mã độc, trên cơ sở đó huấn luyện các mô hình trí tuệ nhân tạo để phát hiện mã độc để xây dựng một phần mềm có khả năng phát hiện và diệt mã độc khi xâm nhập vào máy tính

# Mục tiêu của đề tài

Hiểu rõ về virus máy tính, các đặc trưng của virus máy tính.

Phát triển phần mềm cho phép phát hiện sự tồn tại của virus trên máy tính dựa vào các đặc trưng của virus.

# Phương pháp nghiên cứu

Về mặt lý thuyết: Nghiên cứu tổng quan về các loại mã độc, cách thức lây lan phá hoại của chúng. Nghiên cứu các phương pháp phát hiện mã độc phổ biến hiện nay.

Về mặt thực nghiệm: dựa trên cơ sở lý thuyết đề xuất phương pháp trích rút đặc trưng cho bài toán phát hiện mã độc, đồng thời tiến hành xây dựng một phần mềm có khả năng phát hiện mã độc.

# Đối tượng và phạm vi nghiên cứu

1. Đối tượng nghiên cứu:

* Nghiên cứu các phương pháp phân tích mã độc
* Nghiên cứu các phương pháp phát hiện và loại bỏ mã độc.
* Nghiên cứu các phương pháp trích rút đặc trưng các mã độc.
* Nghiên cứu các phương pháp tạo cơ sở dữ liệu đặc trưng của mã độc.
* Tìm hiểu và nghiên cứu các mô hình học máy (trí tuệ nhân tạo) cho phép phát hiện mã độc hiệu quả.

1. Phạm vi nghiên cứu:

* Các vấn đề liên quan đến phân tích mã độc.
* Các vấn đề liên quan đến mã độc.
* Các vấn đề về trích rút đặc trưng mã độc.
* Các tiêu chí đánh giá hiệu quả, chất lượng của các phương pháp phát hiện và trích rút đặc trưng.

# Nội dung nghiên cứu

* Tìm hiểu lập trình Win32 API.
* Tìm hiểu về các loại mã độc, cách thức lây lan phá hoại của chúng.
* Tìm hiểu các phương pháp phát hiện mã độc.
* Tìm hiểu các phương pháp trích rút đặc trưng mã độc
* Phân tích và thiết kế hệ thống (biều đồ UC, SD, Class, ERD, …)
* Xây dựng CSDL về đặc trưng về các mã độc hiện có.
* Triển khai xây dựng phần mềm phát hiện virus dựa vào dấu hiệu và dựa vào đáp ứng của mã độc đối với hệ thống máy tính (giao diện, chức năng, ….)
* Kiểm thử - Đánh giá hoạt động
* Hiệu chỉnh và hoàn thiện chương trình.

# Kết quả nghiên cứu

Nắm được cơ chế hoạt động của mã độc và các mô hình phát hiện mã độc hiện đang được áp dụng. Từ đó xây dựng phần mềm phát hiện và tiêu diệt mã độc khi xâm nhập vào máy tính.

# Sản phẩm

Phần mềm diệt virus ứng dụng trí tuệ nhân tạo.

# Về các đóng góp của đề tài đến giáo dục và đào tạo, kinh tế xã hội và an ninh quốc phòng

Phát hiện và ngăn chặn việc lây nhiễm virus đối với máy tính và hệ thống mạng, góp phần đảm bảo an toàn an ninh mạng, ngăn chặn nguy cơ mất mát dữ liệu.

|  |  |
| --- | --- |
| **Người hướng dẫn** *(Ký và ghi rõ họ tên)* | **Sinh viên chịu trách nhiệm chính**  *(Ký và ghi rõ họ tên)* |
|  |  |
| TS. Trần Thế Sơn | Đỗ Tấn Tĩnh |