ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

BÁO CÁO TÓM TẮT KẾT QUẢ NGHIÊN CỨU ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP ĐẠI HỌC

XÂY DỰNG PHẦN MỀM DIỆT VIRUS ICTUAV

Mã số: ĐH2016-TN07-01

Chủ nhiệm đề tài: ThS. Trịnh Minh Đức

THÁI NGUYÊN, 05/2019

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT&TT

DANH SÁCH THÀNH VIÊN THỰC HIỆN ĐỀ TÀI

TT	Họ và tên	Đơn vị công tác và lĩnh vực chuyên môn	Ghi chú
1	ThS. Trịnh Minh Đức	Đơn vị công tác: Trường Đại học Công nghệ thông tin và Truyền thông Chuyên môn: CNTT	
2	ThS. Lê Khánh Dương	Đơn vị công tác: Trường Đại học Công nghệ thông tin và Truyền thông Chuyên môn: CNTT	
3	ThS. Nguyễn Tuấn Hiệp	Đơn vị công tác: Trường Đại học Công nghệ thông tin và Truyền thông Chuyên môn: CNTT	
4	ThS. Vũ Việt Dũng	Đơn vị công tác: Trường Đại học Công nghệ thông tin và Truyền thông Chuyên môn: CNTT	
5	ThS. Võ Văn Trường	Đơn vị công tác: Trường Đại học Công nghệ thông tin và Truyền thông Chuyên môn: CNTT	

ĐƠN VỊ PHỐI HỢP CHÍNH

Tên đơn vị trong và ngoài nước	Nội dung phối hợp nghiên cứu	Họ và tên người đại diện đơn vị
Bộ môn An toàn hệ thống thông tin – Trường Đại học CNTT&TT	- Hỗ trợ cơ sở vật chất, thiết bị, phòng nghiên cứu. - Phối hợp nghiên cứu.	ThS. Lê Khánh Dương

MỤC LỤC

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CNTT&TT

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung

- Tên đề tài: XÂY DỰNG PHẦN MỀM DIỆT VIRUS ICTUAV
- Mã số: ĐH2016-TN07-01
- Chủ nhiệm: ThS. Trịnh Minh Đức
- Cơ quan chủ trì: Trường Đại học Công nghệ thông tin và Truyền thông
- Thời gian thực hiện: 24 tháng

2. Mục tiêu

Đưa ra đánh giá về các loại mã độc, tìm hiểu các phương pháp phân tích, trích rút đặc trưng mã độc từ đó xây dựng một phần mềm có khả năng phát hiện và loại bỏ mã độc

3. Kết quả nghiên cứu

- Đề tài đã thực hiện tìm hiểu về lập trình Win 32 API, Named Pipes.
- Nghiên cứu các vấn đề liên quan đến các loại mã độc, cách thức lây lan phá hoại của mã độc, các phương pháp phát hiện và loại bỏ mã độc đồng thời nghiên cứu về các phương pháp trích rút đặc trưng mã độc qua đó đề xuất được một giải pháp trích trọn đặc trưng cho bài toán phát hiện mã độc.
- Bên cạnh đó đề tài cũng đưa ra các kỹ thuật để xây dựng cơ sở dữ liệu mẫu mã độc.

4. Sản phẩm

- **Bài báo đăng tạp chí, kỷ yếu trong nước: 03**

1. Võ Văn Trường, Trịnh Minh Đức, Nguyễn Văn Vinh, Lê Khánh Dương (2016), "Đề xuất giải pháp trích chọn đặc trưng cho các thuật toán phân lớp dữ liệu trong kỹ thuật học máy giám sát và ứng dụng hiệu quả vào bài toán phát hiện mã độc", *Tạp chí Công nghệ thông tin và truyền thông – Bộ TT&TT,* 529 (719), tr. 38-46.

2. Võ Văn Trường, Trịnh Minh Đức, Nguyễn Văn Vinh, Lê Khánh Dương (2016), "Đề xuất giải pháp trích chọn đặc trưng cho các thuật toán phân lớp dữ liệu trong kỹ thuật học máy giám sát và ứng dụng hiệu quả vào bài toán phát hiện mã độc", *Kỷ yếu hội thảo Một số vấn đề chọn lọc về an toàn an ninh thông tin lần thứ 1*, Học viện kỹ thuật mật mã.

3. Trịnh Minh Đức, Đinh Khánh Linh, Lê Khánh Dương, Võ Văn Trường (2019), "Một số kỹ thuật tạo cơ sở dữ liệu mẫu mã độc", *Tạp chí khoa học và công nghệ - Đại học Thái Nguyên* (đã được chấp nhận đăng).

- **Đề tài sinh viên NCKH: 01**

Nguyễn Hoàng Thắng. "Nghiên cứu và xây dựng hệ thống quét mã độc trực tuyến". Đề tài khoa học cấp sinh viên năm 2018. Kết quả nghiệm thu đạt loại Tốt theo QĐ số 31/QĐ-DH CNTT&TT, ngày 07/01/2019 của trường đại học CNTT&TT về việc công nhận kết quả thực hiện các đề tài KH&CN cấp cơ sở và sinh viên năm 2018.

- **Phần mềm diệt virus ICTUAV**

5. Hiệu quả

- Giáo dục, đào tạo: Sản phẩm của đề tài sẽ là tài liệu tham khảo cho sinh viên học các môn học liên quan đến chuyên ngành an toàn thông tin có thể hiểu được cơ chế hoạt động và các phương pháp phát hiện mã độc.
- Kinh tế, xã hội: Phần mềm có khả năng phát hiện và loại bỏ các loại mã độc, giúp bảo vệ người dùng tránh được các nguy cơ bị đánh cắp mật khẩu, tài khoản ngân hàng, thông tin cá nhân … tránh được các thiệt hại về kinh tế.

6. Khả năng áp dụng và phương thức chuyển giao kết quả nghiên cứu

Phần mềm đã được sử dụng tại công ty Lumi Việt Nam và đang tiếp tục được nâng cấp để có thể chuyển giao đến các đơn vị khác.

Ngày tháng năm 2019

Cơ quan chủ trì **Chủ nhiệm đề tài**

INFORMATION ON RESEARCH RESULTS

1. General information

- Project title: BUILDING THE ICTUAV ANTI-VIRUS SOFTWARE.
- Code number: ĐH2016-TN07-01.
- Coordinator: Msc. Trinh Minh Duc.
- Implementing institution: Thai Nguyen University of Information Technology and Communication.
- Duration: 24 months.

2. Objective(s)

Giving evaluations on types of malicious softwares, studying analysis methods, feature extractions of malwares from there building a software which is capable of detecting malwares.

3. Research results

- The subject studied about Win 32 API, Named Pipes
- Types of malwares, spreading and detructive actions of malwares
- Studying about methods for detecting and removing malwares
- Feature extraction methods
- Proposing a feature extraction algorithm for malware detection.

4. Products

- **Published articles in the country's magazines: 03**.

1. Vo Van Truong, Trinh Minh Duc, Nguyen Van Vinh, Le Khanh Duong (2016), "Proposing a feature extraction method for malware detection problem", *Journal of Information & Communications Technology*, 529 (719), pp. 38-46.
2. Vo Van Truong, Trinh Minh Duc, Nguyen Van Vinh, Le Khanh Duong (2016), "Proposing a feature extraction method for malware detection problem", *In proceedings of SoIS 2016*, ACT.
3. Trinh Minh Duc, Dinh Khanh Linh, Le Khanh Duong, Vo Van Truong (2019), "A number of techniques to create malware database", *TNU Journal of Science and Technology* (accepted to publish).

- **Student research topic: 01.**

Nguyen Hoang Thang. “Studying and building an online malware scanning system”. Student research topic 2018. Achieving the good result.

- **The anti-virus software ICTUAV**

5. Effects

- Education: Products of the project will be reference material sources for students who study subjects related to Information security
- Economy and society: The software is capable of detecting and removing malwares, which protects users from avoiding risks such as loss of passwords, bank accounts…

6. The ability to applying and methods of transferring the results of research

The software is used by Lumi Company and being updated more functions.

MỞ ĐẦU

1. Tổng quan tình hình nghiên cứu thuộc lĩnh vực của đề tài ở trong và ngoài nước

Ngoài nước: 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 phát triển nhanh chóng và sự tàn phá nặng nề 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, nhưng nhìn chung các nghiên cứu này đều xoay quanh hai phương pháp chính phát hiện dựa trên sự bất thường hay dị thường và dựa trên dấu hiệu.

Trong [33] Schultz và nhóm tác giả đã lần đầu tiên giới thiệu khái niệm khai phá dữ liệu cho việc phát hiện mã độc. Họ đã sử dụng ba đặc trưng tĩnh khác nhau cho việc phân loại mã độc: Portable executable (PE), strings và các chuỗi byte.

Natajai [26] đã đề xuất một phương pháp trực quan hóa và phân loại mã độc sử dụng các kỹ thuật xử lý ảnh, phương pháp này hình dung các tập tin mã độc như là những ảnh xám. Sử dụng ký thuật K láng giềng gần nhất với khoảng cách Euclide cho việc phân loại mã độc. Trong [27] các tác giả đã so sánh sự phân tích dựa trên kết cấu nhị phân (dựa trên kỹ thuật xử lý ảnh) với phân tích động. Họ đã phát hiện ra rằng sự phân loại sử dụng phương pháp này nhanh hơn, khả năng mở rộng hơn và có thể so sánh với phân tích động về sự chính xác.

Kong và nhóm tác giả [23] đã trình bày một framework cho việc phân loại mã độc tự động dựa trên thông tin cấu trúc của mã độc.

Vấn đề quan trọng nhất để phát hiện đúng mã độc là cần phải có phương pháp trích rút đặc trưng thực sự hiệu quả, trong [38] Tony Abou-Assaleh đã đề xuất một phương pháp phát hiện mã độc mới sử dụng dấu hiệu đặc trưng byte n-grams, các thí nghiệm của Abou-Assaleh đã đạt được tỉ lệ chính xác 98%. Trong [32] Santos và nhóm tác giả đã sử dụng đặc trưng opcode n-gram cho phương pháp phát hiện của họ.

2. Tính cấp thiết của đề tài

Ngày nay mạng Internet đã trở nên phổ biến trên toàn thế giới, Internet đã lan tỏa vào mọi ngóc ngách của đời sống xã hội, bất kỳ ai cũng có thể truy cập Internet để tìm kiếm thông tin phục vụ cho nhu cầu của họ. Cùng với sự phát triển nhanh chóng của Internet là những mối nguy hiểm, rủi ro tiềm tàng khi sử dụng như đánh cắp mật khẩu, chiếm đoạt tài khoản ngân hàng, đánh cắp thông tin cá nhân, tống tiền…Những rủi ro này phần lớn bắt nguồn từ những phần mềm độc hại, các loại mã độc, virut máy tính… Chính vì thế, vấn đề nghiên cứu và xây dưng một phần mềm phát hiện và loại bỏ mã độc máy tính là một việc vô cùng cấp thiết.

Theo Báo cáo tổng kết của Kaspersky năm 2014 (Kaspersky Security Bulletin 2014), có 1.4 triệu vụ tấn công người dùng bằng mã độc trên Android năm 2014, tăng gấp 4 lần so với năm 2013. Trong đó, Việt Nam đứng thứ 6 trên toàn thế giới về số người dùng thiết bị di động bị mã độc tấn công. Nguy cơ mất an toàn thông tin đang ở mức đáng báo động khi Việt Nam có gần 50% số người dùng có nguy cơ nhiễm mã độc khi sử dụng Internet trên máy tính, xếp hạng 4 trên toàn thế giới; và đứng đầu thế giới với gần 70% người dùng máy tính dễ bị nhiễm mã độc, phần mềm độc hại cục bộ (qua USB, thẻ nhớ,…).

Ngoài ra, Microsoft ước tính rằng có khoảng 80% máy tính tại Việt Nam nhiễm các loại mã độc và phần mềm độc hại. Theo báo cáo gần đây của Hiệp hội An toàn Thông tin Việt Nam (VNISA), phần lớn các cơ quan tổ chức tại Việt Nam cho phép dùng thiết bị cá nhân (di động và máy tính bảng) truy cập vào mạng lưới tại nơi làm việc nhưng có tới 74% trong số thiết bị không hề sử dụng bất kỳ biện pháp bảo mật thông tin nào. Những thông số này đã dấy lên một mối lo ngại rất lớn và cũng đặt ra một áp lực không hề nhỏ cho các lãnh đạo, chuyên gia về công nghệ thông tin tìm ra giải pháp để đối phó với tình trạng mất an toàn thông tin trong môi trường hiện nay.

Theo báo cáo gần đây của Hiệp hội An toàn Thông tin Việt Nam phần lớn các cơ quan tổ chức tại Việt Nam cho phép dùng thiết bị cá nhân (di động và máy tính bảng) truy cập vào mạng lưới tại nơi làm việc nhưng có tới 74% trong số thiết bị không hề sử dụng bất kỳ biện pháp bảo mật thông tin nào. Từ thực tế này, tôi đề xuất đề tài: “Xây dựng phần mềm diệt virus ICTUAV”. Đây sẽ là một đề tài có ý nghĩa thực tiễn và ứng dụng rất cao, với phần mềm này chúng tôi hy vọng sẽ giúp bảo vệ người dùng trước những nguy hiểm, rủi ro khi truy cập máy tính và Internet.

3. Mục tiêu của đề tài

Đưa ra đánh giá về các loại mã độc, virus máy tính, tìm hiểu các phương pháp phân tích, trích rút đặc trưng mã độc từ đó xây dựng một phần mềm có khả năng phát hiện và loại bỏ mã độc.

4. Cách tiếp cận và phương pháp nghiên cứu

4.1. Cách tiếp cận

- Nghiên cứu trong tài liệu (từ các bài báo, tạp chí...)
- Tham gia các diễn đàn, hội thảo, xê mi na về an toan thông tin trong nước.
- Thử nghiệm trên một hệ thống thực tế.

4.2. 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.

5. Đối tượng và phạm vi nghiên cứu

5.1. 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 mã độc
- Nghiên cứu các phương pháp tạo cơ sở dữ liệu mẫu mã độc
- Nghiên cứu phần mềm mã nguồn mở ClamAV

5.2. 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

6. Nội dung nghiên cứu

- Tìm hiểu lập trình Win32 API
- Tìm hiểu Named Pipes
- Tìm hiểu về các loại mã độc, cách thức lây lan phá hoại của chúng.
- Tìm hiểu các phương pháp phát hiện và loại bỏ mã độc.
- Tìm hiểu các phương pháp trích rút đặc trưng mã độc
- Xây dựng CSDL mã độc
- Nghiên cứu và tìm hiểu về Clamav, một phần mềm tự do nguồn mở cho việc phát hiện và loại bỏ các loại mã độc.
- Xây dựng phần mềm phát hiện và loại bỏ mã độc dựa trên Clamav

I. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH CỦA ĐỀ TÀI

1. Mức độ hoàn thành mục tiêu đề tài

Mục tiêu của đề tài là Đưa ra đánh giá về các loại mã độc, virus máy tính, tìm hiểu các phương pháp phân tích, trích rút đặc trưng mã độc từ đó xây dựng một phần mềm có khả năng phát hiện và loại bỏ mã độc.

Đề tài đã đề xuất một giải pháp trích chọn đặc trưng cho bài toán phát hiện mã độc, đưa ra một số kỹ thuật để tạo cơ sơ dữ liệu mẫu mã độc, đồng thời xây dựng thành công phần mềm diệt virus ICTUAV.

2. Mức độ hoàn thành các nội dung

Đề tài đã hoàn thành đầy đủ các nội dung đưa ra trong thuyết minh:

- Trình bày các loại mã độc, cách thức lây lan phá hoại của chúng.
- Tìm hiểu các phương pháp trích rút đặc trưng mã độc
- Xây dựng CSDL mã độc
- Nghiên cứu và tìm hiểu về Clamav, một phần mềm tự do nguồn mở cho việc phát hiện và loại bỏ các loại mã độc.
- Xây dựng phần mềm phát hiện và loại bỏ mã độc dựa trên Clamav

3. Số lượng, chất lượng sản phẩm đạt được so với đăng ký

Kết quả của đề tài đáp ứng về số lượng và chất lượng so với thuyết minh phê duyệt của hội đồng.

3.1. Bài báo đăng trên tạp chí, kỷ yếu trong nước

1. Võ Văn Trường, Trịnh Minh Đức, Nguyễn Văn Vinh, Lê Khánh Dương (2016), "Đề xuất giải pháp trích chọn đặc trưng cho các thuật toán phân lớp dữ liệu trong kỹ thuật học máy giám sát và ứng dụng hiệu quả vào bài toán phát hiện mã độc", *Tạp chí Công nghệ thông tin và truyền thông – Bộ TT&TT*, 529 (719), tr. 38-46.

2. Võ Văn Trường, Trịnh Minh Đức, Nguyễn Văn Vinh, Lê Khánh Dương (2016), "Đề xuất giải pháp trích chọn đặc trưng cho các thuật toán phân lớp dữ liệu trong kỹ thuật học máy giám sát và ứng dụng hiệu quả vào bài toán phát hiện mã độc", *Kỷ yếu hội thảo Một số vấn đề chọn lọc về an toàn an ninh thông tin lần thứ 1*, Học Viện Kỹ Thuật Mật Mã.

3. Trịnh Minh Đức, Đinh Khánh Linh, Lê Khánh Dương, Võ Văn Trường (2019), "Một số kỹ thuật tạo cơ sở dữ liệu mẫu mã độc", *Tạp chí khoa học và công nghệ - Đại học Thái Nguyên* (đã được chấp nhận đăng).

3.2. Sản phẩm đào tạo

- 01 đề tài khoa học cấp sinh viên
  Nguyễn Hoàng Thắng. "Nghiên cứu và xây dựng hệ thống quét mã độc trực tuyến". Đề tài khoa học cấp sinh viên năm 2018. Kết quả nghiệm thu đạt loại tốt theo QĐ số 31/QĐ-DH CNTT&TT, ngày 07/01/2019 của trường đại học CNTT&TT về việc công nhận kết quả thực hiện các đề tài KH&CN cấp cơ sở và sinh viên năm 2018.
- **Sản phẩm ứng dụng:** Phần mềm diệt virus ICTUAV.

II. ĐÁNH GIÁ GIÁ TRỊ KHOA HỌC VÀ THỰC TIỄN CỦA KẾT QUẢ NGHIÊN CỨU

Kết quả của đề tài đã đưa ra được các đánh giá về các loại mã độc, cách thức lây lan phá hoại của chúng, các phương pháp phát hiện mã độc hiện nay, qua đó chúng tôi đã đề xuất giải pháp trích chọn đặc trưng cho bài toán phát hiện mã độc và đã tiến hành các thử nghiệm để chứng tỏ hiệu quả của giải pháp đề xuất. Đồng thời chúng tôi cũng đưa ra các kỹ thuật để tạo cơ sở dữ liệu mẫu mã độc. Kết quả chính của đề tài là phần mềm diệt virus ICTUAV, phần mềm này của chúng tôi đã được chuyển giao cho công ty Lumi và đã nhận được những phản hồi rất tích cực.

Kết quả của đề tài là nguồn tài liệu tham khảo cho sinh viên học các môn học liên quan đến chuyên ngành an toàn thông tin, công nghệ thông tin có thể hiểu được cơ chế hoạt động và các phương pháp phát hiện mã độc; các nghiên cứu về mã độc có sử dụng quá trình trích rút đặc trưng.

KẾT LUẬN VÀ KIẾN NGHỊ

Đề tài đã xây dựng thành công phần mềm diệt virus ICTUAV, trong đề tài chúng tôi đã đề xuất sử dụng giải pháp trích chọn đặc trưng để tăng khả năng phát hiện mã độc của phần mềm, đây là một đóng góp quan trọng của đề tài. Tuy nhiên phần mềm vẫn còn một nhược điểm là chưa có cơ chế bảo vệ thời gian thực. Trong thời gian tới chúng tôi sẽ khắc phục nhược điểm này và tiếp tục tập trung nâng cấp phần mềm, cải thiện giải pháp trích chọn đặc trưng để nâng cao khả năng phát hiện mã độc.

Để kết quả của đề tài có thể tiếp tục được phát triển cần có một phòng thí nghiệm chuyên phân tích về mã độc, qua đó có thể xây dựng được cơ sở dữ liệu mã độc cho phần mềm.