

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

ĐỖ QUỐC LƯU

**PHÂN TÍCH MỨC ĐỘ AN TOÀN CỦA ỨNG DỤNG ANDROID
DỰA TRÊN HỌC MÁY**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, 2019

LỜI CẢM ƠN

Lời đầu tiên, em xin chân thành cảm ơn PGS. TS Phạm Thanh Giang người đã trực tiếp hướng dẫn em hoàn thành luận văn. Với những lời chỉ dẫn, những tài liệu, sự tận tình hướng dẫn và những lời động viên của thầy đã giúp em vượt qua nhiều khó khăn trong quá trình thực hiện luận văn này.

Em cũng xin cảm ơn quý thầy cô giảng dạy chương trình cao học "Khoa học máy tính" đã truyền dạy những kiến thức quý báu, những kiến thức này rất hữu ích và giúp em nhiều khi thực hiện nghiên cứu.

Cuối cùng, em xin gửi lời cảm ơn tới gia đình và bạn bè đã luôn ủng hộ động viên giúp đỡ em trong suốt những năm học vừa qua.

Em xin chân thành cảm ơn!

Thái Nguyên, ngày tháng năm 2019

Học viên

LỜI CAM ĐOAN

Em xin cam đoan: Luận văn này là công trình nghiên cứu thực sự của cá nhân, được thực hiện dưới sự hướng dẫn khoa học của **PGS. TS Phạm Thanh Giang**.

Các số liệu, những kết luận nghiên cứu được trình bày trong luận văn này trung thực và chưa từng được công bố dưới bất cứ hình thức nào.

Em xin chịu trách nhiệm về nghiên cứu của mình.

Học viên

MỤC LỤC

LỜI CẢM ƠN	1
LỜI CAM ĐOAN	ii
MỤC LỤC.....	iii
DANH MỤC CÁC TỪ VIẾT TẮT	v
DANH MỤC BẢNG.....	vi
DANH MỤC HÌNH ẢNH	vii
MỞ ĐẦU.....	1
CHƯƠNG 1: TỔNG QUAN VỀ AN TOÀN BẢO MẬT TRÊN HỆ ĐIỀU HÀNH ANDROID.....	6
1.1 Kiến trúc hệ điều hành Android.....	7
1.1.1 Tầng hạt nhân Linux (Linux Kernel)	8
1.1.2 Tầng Thư viện (Libraries) và Tiến trình Android (Android Runtime).....	9
1.1.3 Tầng Khung ứng dụng (Application Framework)	11
1.1.4 Tầng Ứng dụng (Applications)	13
1.1.5 Máy ảo Dalvik (DVM).....	14
1.1.6 Cấu trúc ứng dụng Android	17
1.2 An toàn bảo mật hệ điều hành Android	21
1.2.1 Mã độc.....	21
1.2.2 Biểu hiện của mã độc di động.....	22
1.2.3 Mã độc trong môi trường Android.....	22
1.2.4 Một số kỹ thuật phân tích mã độc	24
CHƯƠNG 2: GIỚI THIỆU HỌC MÁY VÀ CÁC MÔ HÌNH HỌC MÁY ..	36
2.1 Học máy là gì?	36
2.2 Phân loại kỹ thuật học máy	37
2.3 Các bước học máy.....	37

2.3.1 Thuật toán cây quyết định J48 (Decision Trees)	39
2.3.2 Thuật toán hồi quy logistics.....	42
CHƯƠNG 3: MÔ PHỎNG VÀ KIỂM THỬ	45
3.1 Mô phỏng (Phương pháp thực hiện)	48
3.1.1 Phương thức tính điểm.....	48
3.1.2 Thực nghiệm	51
3.2 Kết quả thực nghiệm	57
3.2.1 Kết quả	57
3.2.2 Đánh giá, tranh luận.....	58
KẾT LUẬN	60
TÀI LIỆU THAM KHẢO.....	61

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Từ gốc	Nghĩa tiếng Việt
AM	Activity Manager	Khối quản lý hoạt động
APK	Android Package	Gói ứng dụng Android
GPS	Global Positioning System	Hệ thống định vị toàn cầu
IDS	Intrusion Detection System	Hệ thống phát hiện xâm nhập
ID	Identification	Nhận dạng, nhận biết
AI	Artificial Intelligence	Trí tuệ nhân tạo
J48	Decision Trees	Cây quyết định
TP	True Positive	Số mẫu ác tính được phát hiện chính xác
FP	False Positive	Số mẫu ác tính bị phát hiện sai
TN	True Negative	Số mẫu lành tính được phát hiện chính xác
FN	False Negative	Số mẫu lành tính bị phát hiện sai
TPR	True Positive Rate	Tỷ lệ xác nhận chính xác mẫu ác tính
FPR	False Positive Rate	Tỷ lệ xác nhận sai mẫu ác tính
TNR	True Negative Rate	Tỷ lệ xác nhận chính xác mẫu lành tính
FNR	False Negative Rate	Tỷ lệ xác nhận sai mẫu lành tính
ACC	Overall Accuracy	Tỷ lệ xác nhận chính xác trên tổng số mẫu

DANH MỤC BẢNG

Bảng 3.1: Điểm Risk- score	49
Bảng 3.2: Điểm Protection- score	49
Bảng 3.3: Ví dụ tính điểm “quyền” ứng dụng	50
Bảng 3.4: Kết quả của phương pháp	57

DANH MỤC HÌNH ẢNH

Hình 1.1: Sơ đồ kiến trúc hệ thống cấp thấp Android	8
Hình 1.2: Hộp thoại ứng dụng không phản hồi.....	12
Hình 1.3: Vị trí điện thoại được cung cấp cho google map	13
Hình 1.4: Cấu trúc tập tin .dex	15
Hình 1.5: Ứng dụng xin cấp quyền từ người dùng	17
Hình 1.6: Quy trình xây dựng 1 file ứng dụng Android	18
Hình 1.7: Mục tiêu của mã độc Android.....	23
Hình 1.8: Lời gọi hệ thống	24
Hình 1.9: Phân tích bất thường cố định	25
Hình 1.10: Phân tích mã nguồn.....	25
Hình 1.11: Phân tích quyền ứng dụng.....	27
Hình 1.12 : Các quyền đơn giản trong Android.....	31
Hình 1.13 : Các quyền nguy hiểm trong Android.....	32
Hình 1.14 : Giao diện file AndroidManifest	33
Hình 1.15 : Xin cấp quyền trên Android 6.0 Marshmallow	34
Hình 1.16 : Giao diện App Ops.....	35
Hình 1.17 : Giao diện kiểm soát quyền của ứng dụng Calender trên Android 6.0	35
Hình 2.1 : Kết quả thuật toán Cây quyết định.....	42
Hình 2.2 : Kết quả thuật toán Hồi quy Logistic	44
Hình 3.1: Phân tích mã nguồn.....	48
Hình 3.2: Xác định quyền	54
Hình 3.3: Protection- score	55
Hình 3.4: Risk- score.....	55
Hình 3.5: Ngưỡng đánh giá.....	56
Hình 3.6: Kết quả so sánh với phương pháp Ryo Sato, logictics	58

MỞ ĐẦU

1. Đặt vấn đề

Trong thời đại công nghệ hiện nay, khi tất cả mọi thứ đều được số hóa, cuộc sống, công việc trở nên khó khăn hơn nếu thiếu đi một chiếc điện thoại thông minh – smartphone. Dựa vào sự gia tăng về số lượng người dùng smartphone, các phần mềm độc hại với mục đích viết ra để trục lợi cá nhân cũng theo đó mà nhân lên hàng năm. Những phần mềm này ngày càng được tinh hóa hơn nhằm vượt qua các rào cản an ninh của Google. Chúng trở nên nhiều và đa dạng khiến những người dùng bất cẩn dễ dàng bị xâm hại. Sau thời gian nghiên cứu, tôi xin đề xuất việc kết hợp giữa phương pháp phân tích tĩnh và một số mô hình học máy vào việc phân tích mức độ đáng tin cậy của một ứng dụng Android. Nghiên cứu này sẽ giúp thiết bị di động của người dùng tránh khỏi việc lưu trữ một phần mềm nguy hiểm.

2. Đối tượng và phạm vi nghiên cứu

a. Đối tượng nghiên cứu:

- Nghiên cứu phương pháp phát hiện đánh giá mức độ mất an toàn của các phần mềm trên thiết bị di động Android
- Dữ liệu mẫu khai thác từ các nguồn công khai như Google Play và các trang nghiên cứu về bảo mật trên Android

- Drebin: <https://www.sec.cs.tu-bs.de/~danarp/drebin/>
- Contagio Mobile Malware: <http://contagiominedump.blogspot.com/>
- Các phần mềm lành tính được thu nhập trên cửa hàng Google Play

b. Phạm vi nghiên cứu:

- Trong hướng nghiên cứu của đề tài tập trung vào các ứng dụng trên nền tảng Android, tuy nhiên các kết quả nghiên cứu cũng là nền tảng để nghiên cứu về an toàn cho các ứng dụng trên nền tảng iOS.

3. Hướng nghiên cứu của đề tài

Hiện nay, số lượng smartphone đã vượt qua số lượng máy tính PC. Các thông tin quan trọng trên smartphone rất đa dạng, mang tính cá nhân. Các thông tin này thậm chí quan trọng hơn cả thông tin trên PC; bao gồm các tài liệu, các ghi chép, tài khoản email, tài khoản mạng xã hội, tin nhắn SMS/MMS, danh bạ, thông tin cuộc gọi, thông tin vị trí, hình ảnh...

Theo công bố gần đây của Kaspersky, năm 2012 là năm cho thấy sự tăng trưởng bùng nổ của phần mềm độc hại cho Android. Từ 8 chương trình độc hại vào tháng 1/2011, tỉ lệ phát hiện trung bình hàng tháng phần mềm độc hại mới cho Android mới trong năm 2011 đã lên đến hơn 800 mẫu. Trong năm 2012, Kaspersky Labs đã xác định trung bình 6.300 mẫu phần mềm độc hại trên thiết bị di động mới mỗi tháng. Nhìn chung, trong năm 2012 số lượng mẫu độc hại cho Android được biết đến tăng hơn 8 lần. Đến năm 2016, số malware được phát hiện lây nhiễm trên các thiết bị di động đã đã cán mốc 2 triệu. Các phần mềm độc hại có thể lợi dụng các lỗ hổng hoặc sự bất cẩn của người dùng để cài đặt vào smartphone. Một số phần mềm còn tìm cách vượt qua chính sách an ninh của hệ thống phân phối phần mềm (Market place) để giám sát và thu thập thông tin người dùng một cách tinh vi.

