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 cám 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 cám ơn quý thầy cô giảng dạy chương trình cao học "Khoa hoc 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 cám 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 cám 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

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

DANH MỤC HÌNH ẢNH

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 hằng 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://contagiominidump.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 trí 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.