第37卷 第1期 计 算 机 学 报 Vol. 37 No. 1

2014年1月 CHINESE JOURNAL OF COMPUTERS Jan. 2014

基于多类特征的应用恶意行为检测系统

杨欢张玉清1胡予濮1刘奇旭2

1西安电子科技大学综合业务网理论及关键技术国家重点实验室西安710071)

2中国科学院大学国家计算机网络入侵防范中心北京100190)

**摘 要** 针对目前未知的Android恶意应用可以采用数据挖掘算法进行检测，但使用单一数据挖掘算法无法充分 发挥Android应用的多类行为特征在恶意代码检测上所起的不同作用.文中首次提出了一种综合考虑Android多类行为特征的三层混合系综算法THEA(Triple Hybrid Ensemble Algorithm)用于检测Android未知恶意应用.首先，采用动静态结合的方法提取可以反映八did应用恶意行为的组件、函数调用以及系统调用类特征；然后，针对上述3类特征设计了三层混合系综算法THEA该算法通过构建适合3类特征的最优分类器来综合评判Android的恶意行为；最后，基于THEA实现了Android应用恶意行为检测工具Androdect，并对现实中的 1126个恶意应用和2000个非恶意应用进行检测.实验结果表明，Android能够利用Android应用的多类行为特征有效检测Android未知恶意应用.并且与其它相关工作对比，Android在检测准确率和执行效率上表现更优.

关键词系综算法；nddd应用；多类特征；意代码检测；行为分析;数据挖掘；智能手机；络行为 中图法分类号 TP393 犇[号 10. 3724/SP.J. 1016.2014. 00015

**A Malware Behavior Detection System of Android Applications Based on**

**Multi-Class Features**

YANG Huan1 ZHANG Yu-Qing1)， HU Yu-Pu1 LIU Qi-Xu2)

1 ***iSiaLe Key Laboratory of InLegraLed Services NeLworks，Xidian University****，****Xi'n*** 710071**)**

***2 {National Computer Network Intrusion ProLecLion Center，University of Chinese Academy of Sciences，Beijing*** 100190**)**

**Abstract** At present，data mining algorithm is always used to detect unknown malicious applica—

tions of Android. While single data mining algorithm could not play the role of multi-class Android features in malware detection. For this problem， a Triple Hybrid Ensemble Algorithm (THEA) was first proposed，which considered multi-class Android features. First， we combined the static and dynamic methods to extract three classes of Android features，which could reflect malicious behavior effectively, such as components, function calls and system calls. Second, we designed THEA to build optimal classifier by handling three classes of features and then made a comprehensive judgment of unknown Android application. Finally, we implemented an automated tool named Androdect to detect 1126 malicious and 2000 non-malicious apps in real. The experi­mental results show that Androdect plays the role of multi-class Android features in unknown malware detection and it performs better than other related works on the availability，efficiency and accuracy.

**Keywords** ensemble algorithm； Android application； multi-class features； malicious

第37卷 第1期 计 算 机 学 报 Vol. 37 No. 1

2014年1月 CHINESE JOURNAL OF COMPUTERS Jan. 2014

code detec­

tion； behavior analysis； data mining； smartphone； network behavior

收稿日期：2013-04-13;最终修改稿收到日期**：**2013-11-19.本课题得到国家自然科学基金（1272481**)**、中国博士后科学基金资助项目 (201150041**，**20150152)、北京市自然科学基金（4122089)、国家发改委信息安全专项（发改办髙技[2012]1424号）资 助.杨欢，女， 1984年生，博士，助理工程师，主要研究方向为信息安全、漏洞挖掘和恶意代码检测.E-mail: yangh@nipc. org. cn.**张玉清，**男**，**1966年 生，博士，教授，中国计算机学会(CCF)会员，主要研究领域为网络与信息系统安全.**胡予濮**，男**，**1955年生，博士，教授，主要研究领域为 序列密码与分组密码、网络安全协议的设计与分析**.刘奇旭**，男，1984年生，博士，主要研究方向为信息安全、漏洞挖掘和漏洞等级评估.

1 引言

近年来,智能手机普及率迅速增加,智能手机 应用功能激增.全球最具权威的IT研究与顾问咨询公司Gartner的数据显示预期到2016年年底,将有23亿部电脑、平板电脑以及智能手机使用Android系统,而微软windows设备的数量为22.8亿. Android近年来的快速发展使其成为了一 个占主导地位的智能手机平台，拥有整个市场约三 分之二的份额.由于Android系统的开放性，它也成 为了众多恶意代码开发者的活跃地盘.

Android用户可以从Google Play**2**和第三方Android市场（例如：Amazon、hiapk、gfan等）下载应用程序.应用程序开发者可以上传自己的应用程序到第三方Android市场.Google Play(前名为Android Market),是一个由Google为Android设 备开发的在线应用程序商店,可以让用户浏览、下载及购买Google Play上的第三方应用程序3 . Google Play提供了Google’s Bouncer检测恶意程序，但是 实时性不够，恶意程序在被检测出来之前已经被用 户大量下载

根据独立安全机构Av-test在2013年1月使用 ESETT、Lookout: Antivirus S Security**、**Kaspersky**、** Qih**〇〇**360等22个手机杀毒软件对869个已知 Andrmd恶意应用的检测报告显示④各大手机杀毒 软件的平均查杀率为94%.而对于未知恶意应用的 检测，各大厂商没有公开杀毒软件的技术细节，本文 无法做详细分析.由于恶意代码的数量和种类越来 越多，加上代码混淆加密等技术的兴起，使得恶意 代码检测变得越来越困难.

目前，典型的恶意代码检测技术包括基于签名、 行为的检测方法.传统的基于签名的检测技术被普 遍使用，但是它必须拥有一类恶意应用的签名库后 才能检测该类恶意应用，因此无法有效检测未知的 恶意应用.

基于行为的恶意代码检测目前主要采用动态和 静态两种方法.动态方法是在系统运行过程中收集 应用程序的一些行为信息.优点是绕过了静态方法 遇到的代码混淆和加密等方面的问题，缺点是动态 测试代码覆盖率低，并且有些恶意程序可以防止自 身在模拟器下运

第37卷 第1期 计 算 机 学 报 Vol. 37 No. 1

2014年1月 CHINESE JOURNAL OF COMPUTERS Jan. 2014

行，当在模拟器下运行时会自动崩 溃；态方法主要研究使用反汇编反编译技术或者 在smali中间代码上进行控制流和数据流分析技术 来进行恶意代码检测.优点是代码覆盖率高，缺点是 要解决静态方法无法检测的代码混淆、加密以及在 动态执行中才解码恶意代码的问题.

由于数据挖掘技术可以从大量数据中挖掘出有 意义的信息2，有些研究者利用该技术挖掘出恶意 应用的行为特征，以此来检测未知恶意应用.但是， 目前的研究方法存在3方面的局限性：（1)不同的 数据挖掘算法针对同一类特征的检测效果不同，无 法预知哪个算法效果最优；⑵同一种算法对不同 类型的特征检测效果不一定都是最优；（3)使用单 一的算法不能充分发挥每类特征在Android恶意应 用检测时所起的不同作用.

本文针对以上方法的局限性，首次提出一种基 于三层混合系综算法的多类特征Andrmd应用恶意 行为检测系统，自动检测Andrmd平台的恶意应用 程序.该系统采用动静态结合的行为特征提取方法， 提取Andrmd应用的多类行为特征，再采用三层混 合系综算法建立检测模型，并对大量现实中的应用 程序进行检测.具体方法介绍如下：

首先，本文创建恶意应用库和非恶意应用库.恶 意应用库的创建参考文献[5]中所使用的部分恶意 应用.非恶意应用库的创建是通过修改AkdeniZ 的爬虫程序从Google Play上批量下载应用程序， 并使用多种已有检测软件、杀毒软件和人工进行检 测，确保数据库的准确性.

其次，在特征提取阶段，本文采用动静态结合的 分析技术.静态分析主要对AndroidManifest. xml文 件进行自动化分析.一是提取packet名称和activity 名称为动态分析提供参数；二是对permission**、** activity**、**service、receiver 和 provider 组件进行特征 提取；三是对使用native代码的应用程序中的i 库文件进行分析，提取应用程序的函数调用序列.动 态分析主要采用沙盒技术，自动完成启动模拟器、安 装应用程序、模拟用户行为和提取系统调用序列等 工作．

而后，在特征描述阶段，对提取的多类特征进行

在修改的过程中，遇到了很多没遇到过的问题，我也用了很多软件，比如将pdf文件转换成word文档，虽然不尽人意，但是还是努力地做了，我觉得通过这个上机作业，可能我懂得不是很多，但是至少让我明白了学计算机这一行，必须懂的还有很多很多，在今后的学习中，我会尽我所能的去学、去做。