



专题:基于 Android 系统的终端安全

基于 Android 的木马检测引擎的研究与实现

夏彬,邱峰

(中国科学院信息工程研究所,北京 100093)

摘要:近几年来,Android 手机木马病毒发展迅速,Android 手机安全问题成为大家关注的焦点,基于 Android 的木马检测引擎的研究与实现变得日益迫切。为此,提出了一套特征码提取检测算法 (FCPA),FCPA 通过调用 Android 系统库函数获取恶意文件的源路径,利用源路径找到相应文件并对文件进行散列处理,获取文件特征信息,生成一个唯一标识该木马病毒的特征值,然后构建特征码库。同时,设计并实现了木马检测引擎,其利用特征码提取算法快速扫描并检测出手机应用程序中的恶意程序。实验结果表明,该木马检测引擎能够有效检测恶意应用。

关键词: Android; 木马检测引擎; 特征码

中图分类号: TP311.56

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016257

Research and realization of the Trojan detection engine based on Android

XIA Bin, QIU Feng

Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract: During recent years, Trojan viruses on Android systems have greatly evolved, and the frequent security breach of Android systems is rapidly becoming a great concern of contemporary cyber security. The study of Trojan virus detection on Android engine and the application of its outcome has become increasingly significant. A feature code detection algorithm called FCPA got the sourceDir of the known malicious APK files through calling system API. It uses hash algorithm to process these files in order to get the feature information of the files so that the eigenvalue could get which could identify Trojan uniquely. The feature code library was composed of these eigenvalues. The Trojan detection engine called TDE, with the assistance of this feature code library, would be able to take out a quick scan among the files in the cell phone and detect malicious programs in mobile applications. Finally, the designed Trojan detection engine provided a secure environment for the user of the phone.

Key words: Android, Trojan detection engine, feature code

1 引言

近些年,木马病毒屡见不鲜,造成的破坏力日益增大。360 互联网安全中心在 2015 年共截获 Android 平台新增恶

意程序样本数 1 874.0 万个,平均每天截获的恶意程序样本数高达 51 342 个^[1]。由于移动终端安全研究发展时间较短,相对于 PC 端的安全防护体系,基于 Android 平台的安全防护体系相对滞后,目前大部分的手机杀毒软件都是直接

收稿日期:2016-08-20;修回日期:2016-09-28

基金项目:中国科学院战略性先导科技专项(A类)重点行业应用系统信息安防关键技术研究子课题(No. XDA06010703)

Foundation Item: Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA06010703)

将计算机病毒防护技术应用到移动端,没有结合 Android 手机自身的特点,所以杀毒软件的查杀效率并不高,而且还会影响手机其他方面的性能。综合以上几点原因,基于 Android 平台的木马检测引擎的研究与实现势在必行。

目前,国内外学者和知名杀毒软件厂商对 Android 系统的安全机制、恶意软件模型以及恶意文件的检测技术开始进行深入研究,针对 Android 平台的木马检测技术主要分为动态检测技术和静态检测技术。静态检测方法技术成熟、误报率低,是目前最成熟、可靠、有效的反木马技术,如参考文献[2]提出以 smali 源代码为主要研究对象,将 smali 命令抽象化后利用信息熵和字符串比较相似度;参考文献[3]使用了纯静态分析的方法,通过对 APK 权限和系统 API 调用的分析,构建了多维特征向量,并使用机器学习的办法对数据集合进行判断,以此来判断应用程序是否包含恶意行为;参考文献[4]提出了一种静态行为检测技术,检测 Android 平台的恶意软件,通过 readelf 命令输出 ELF 文件的静态信息,并通过最近邻(nearest neighbor)算法进行分类。动态检测技术特征库较小,能够防范未知的木马病毒,如参考文献[5]提出了一种沙箱结构,通过改变系统内核,实时监控系统的输入和输出以及系统 API 的调用情况。通过分析经过修改后的系统内核的日志文件,了解应用程序的实时行为,并生成向量;参考文献[6]提出基于沙盒的 Android 恶意软件动态行为分析方案,此方案为 Android 软件提供了一个真实的运行环境,让其暴露真实目的。

本文利用传统的特征码检测技术,结合 Android 手机自身特点,提出一套特征码提取检测算法(feature code pick-up algorithm,FCPA);同时利用 FCPA 设计并实现了一套完整的木马检测引擎(Trojan detection engine,TDE),包括木马检测模块、对新安装程序扫描模块、定时扫描模块、更新模块;最后随机选取一些恶意软件,对木马检测引擎检测效果进行实验。

2 木马检测技术研究背景

2.1 特征码检测技术

特征码其实可以理解为木马病毒的“指纹”,当一种病毒出现时,安全人员就会从这个病毒程序中截取一段独一无二且足以表示这是病毒的二进制程序代码,以此当作检测程序辨别此病毒的依据。文件的特征码一般有 3 种格式:MD5、校验和字符串。特征码一般是在 PE(portable executable,可移植的执行体)文件中截取一段作为特征值,

常见的 EXE、DLL、OCX、SYS、COM 都是 PE 文件^[7]。目前常见的特征码扫描方式分为文件查杀、内存查杀和行为查杀。同时,针对上述查杀方式,特种木马程序往往选择通过加壳、加花指令,修改壳、特征码的方式,试图躲避防护系统查杀。为此,在对木马特征进行提取和分析的过程中,应采取多种方案并行处理的方式,准确寻找木马程序的特征代码部分,并存入特征码库中。为了提高静态扫描的准确性,确保涉密信息系统的安全性,特征码库的容量应保持在 300 条以上。目前大部分主流木马检测软件都是基于特征码的静态检测技术,各大杀毒厂商不断地从新出现的木马病毒中提取出特征码并加入病毒特征库中,木马检测软件将用户扫描本地文件的结果与特征库特征码相比对,判定文件是否为木马文件。

2.2 虚拟机技术

虚拟机技术是在特征码检测技术基础上发明的一种检测技术。虚拟机是一个运行在隔离环境中的计算机系统。通俗点说,就是软件在虚拟机上运行和在实际的系统中运行毫无差别。虚拟机技术主要应用于那些经过加壳和加花指令的木马病毒上,加壳和加花指令技术都是木马病毒打乱或者加密可执行文件的内部数据,使原本会被查杀的文件免于被查杀,而这些经过特殊处理的木马病毒会很好地隐藏自己的特征码,增加人们对特征码提取的难度。所以特征码检测技术对这类病毒的查杀效果会大大降低。虚拟机技术虽然能解决一些比较棘手的问题,但是由于计算机的处理能力有限,人们也无法预知木马病毒何时能开始工作,所以对于人力物力都是一种消耗。单纯的虚拟机技术不能高效地查杀木马,但是虚拟机技术将单纯的静态检测技术变为动静结合的检测技术,对木马检测技术的提升具有一定的意义。

3 基于特征码的木马检测算法

3.1 FCPA 原理

3.1.1 FCPA 的要求和目标

与 PC 机上的病毒防护相比,用于 Android 手机木马病毒检测的检测算法在追求高系统利用率和高性能的基础上,对系统的稳定性有更高的要求。FCPA 应能与 Android 手机系统自身特点相配合,充分发挥 Android 系统安全机制的自身特性,实现对木马病毒的高效快速查杀。市面上大部分安全软件运行时,需要调用 Android 核心库,长时间与系统内核通信,占用大量内存,造成手机卡顿甚



至死机。FCPA 应优化检测技术,在不影响检测成功率的前提下,降低软件与系统内核通信的时间和次数。

特征码检测技术是目前最成熟、最有效的木马检测技术,FCPA 根据传统的特征码检测算法原理,加入权限检测技术作为辅助,大大降低对感染木马病毒的恶意程序的检测误报率。同时利用散列摘要算法特征,增强特征码的不可逆性和唯一性。

3.1.2 FCPA 设计

在 root 过的手机中,可以发现 APK 文件保存到 Android 手机的路径和格式是固定的,系统程序存储在 system/app 目录下,用户程序存储在 data/app 目录下。利用此特点,通过查询源路径 sourceDir 可以获得 APK 文件安装全路径。通过 sourceDir 可以提取出恶意应用程序的 APK 文件,利用散列算法的特性,将文件信息进行散列摘要处理,最后将处理的结果经过 MD5 函数处理。FCPA 原理如图 1 所示。FCPA 的主要设计思路如下:

- (1)调用 Android 系统 API 函数,获取恶意程序的存储路径 sourceDir;
- (2)将 APK 完整路径传入 getFileMd5() 函数;
- (3)函数生成一个 md5 数字摘要器;
- (4)通过循环将 APK 文件的代码经过散列摘要处理,得到此文件的签名信息;
- (5)将签名信息经过处理后生成 16 位十六进制的特征码;
- (6)将待检测程序的 APK 文件解压缩,获得二进制的 AndroidManifest.xml,并通过转换为文本格式读取或者二进制格式读取的方法查找 AndroidManifest.xml 文件中相应的声明权限,然后根据建立起来的威胁判定规则对这些权限进行威胁判定。

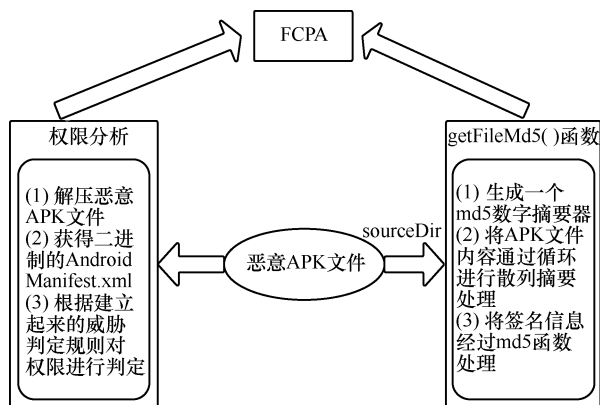


图1 FCPA 原理

3.2 FCPA 特征码数据库设计

FCPA 设计的一个很大用处便是利用 FCPA 高效快速地完成 FCPA 特征码库的构建,特征码库的辨识性越高、越庞大,木马检测引擎的检测效率也就越高。

表 1 是 FCPA 特征码库结构。特征库共有 5 个字段:md5 字段存储的是病毒特征码;type 字段存储的是病毒类型;name 字段是恶意 APK 程序的名称;desc 字段是对病毒行为的描述。特征码库应设计得简洁:一方面便于特征码的存放;另一方面使特征码库更新速度快、占用存储空间小。

表 1 FCPA 特征码库的结构

字段名称	数据类型	是否允许空	是否为主键
_id	integer	否	是
md5	text	是	否
type	numeric	是	否
name	text	是	否
desc	text	是	否

目前 FCPA 特征码库共有 1 万多条木马病毒数据记录,且还在不断完善中,FCPA 特征码库部分截图如图 2 所示。

RecNo	_id	md5	type	name	desc
Click here to define a filter					
601	1656	cbded7c976700eb5ac874827028fc13b	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
602	1657	9476e0000e3b5bf0e5017fb2a67b6610	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
603	1658	84c16e2c53f2665749377391e1ad9f3b	6	Android.Hack.i22kt.a	恶意后台扣费,病毒木马程序
604	1659	5a83a5b2ce0a48695f7102f15d316e65	6	Android.Troj.GacBlocker.b	恶意后台扣费,病毒木马程序
605	1660	09a10fc8e72b188d0739ca78fb8886b	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
606	1661	0918d93dcfac2390070a2df5ee87f4	6	Android.Troj.KungFu.a	恶意后台扣费,病毒木马程序
607	1662	ad3d15d3642904f3c77fa167e97d0a5	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
608	1663	f744f90623d6fc673a2dea04d14c52b	6	Android.Troj.AirAD.a	恶意后台扣费,病毒木马程序
609	1664	4952e036c8392d7eb5d033d1149600d8	6	Android.Troj.SmsSender.a	恶意后台扣费,病毒木马程序
610	1665	330acab26d296398480f26c5b231432e	6	Android.Troj.KillAll.b	恶意后台扣费,病毒木马程序
611	1666	58522ddce84abab1743c3fb2a10f6b7f	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
612	1667	c06697091a431674f131322befe5f42	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
613	1668	7008051b65836f71ac520054080438	6	Android.Troj.AirAD.a	恶意后台扣费,病毒木马程序
614	1669	cf063008bd303eb44861c0270fe1b8ef	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
615	1670	aa2199bfa78981733ecb46726b1fd739	6	Android.Adware.AirAD.a	恶意后台扣费,病毒木马程序
616	1671	c7b241046016e7444d1e42060cfcc076	6	Android.Troj.AirAD.a	恶意后台扣费,病毒木马程序

图2 FCPA 特征码库部分截图

4 TDE 设计与实现

4.1 TDE 设计要求和目标

TDE 是本文利用 FCPA 开发出的一套运行在 Android 4.2 版本以上系统的安全软件,目标是能够高效快速地检测出手机系统程序和用户程序木马感染情况。同时,由于对新安装程序进行木马病毒扫描是保护手机的第一道防线,所以本文应设有新安装程序扫描模块。目前大部分手机安全软件都存在占用内存高,造成手机卡顿的情况,所以,本文设计了定时扫描模块,用户可以根据自己使用手机情况设

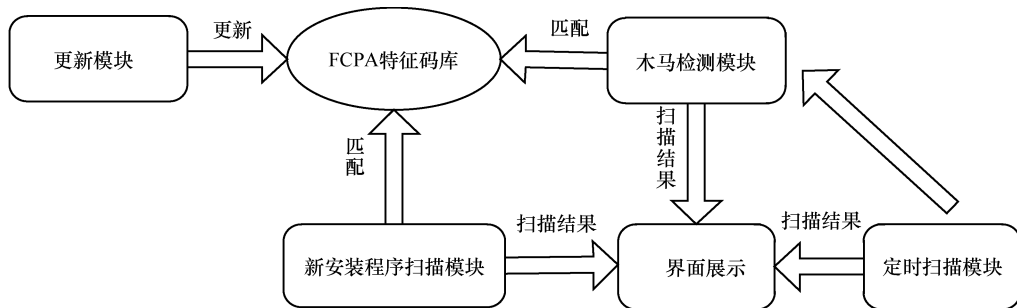


图3 TDE 系统总体架构

定扫描时间。最后及时更新 FCPA 特征码库, 可以加强 TDE 的检测成功率。

4.2 TDE 系统整体框架

根据第 4.1 节的分析可得, TDE 系统主要分为 4 个模块: 木马检测模块、更新模块、新安装程序扫描模块、定时扫描模块。木马检测模块主要工作是利用 FCPA 检测算法, 对 Android 手机已安装程序进行扫描; 更新模块主要工作是更新 FCPA 特征码库; 新安装程序扫描模块主要工作是对新下载的软件进行木马病毒扫描; 定时扫描模块主要功能是可以设置检测时间。系统总体架构如图 3 所示。

4.3 木马检测模块设计

木马检测模块采用 FCPA 检测算法。检测引擎分为 3 个模块: 特征码加载模块、扫描模块和特征码匹配模块。各模块的关系如图 4 所示。

特征码加载模块, 主要负责加载 FCPA 特征码库, 在木马检测系统运行时, 将特征码病毒库装入内存, 根据命令获取病毒库对象; 扫描模块, 扫描手机所有安装程序, 包括系统程序和应用程序; 特征码匹配模块, 利用 FCPA 算出手机安装程序的特征码, 同时将 FCPA 特征码库载入内存中, 和恶意代码扫描模块算出的特征码逐个比对, 如果与特征码库匹配, 则提示用户该程序是危险程序, 如果不匹配, 提示扫描安全。木马检测模块流程如图 5 所示。

特征码加载模块分为两部分, 第一部分是 FCPA 特征码库加载到工程中。FCPA 特征码库被放到 Android 工程 assets 目录下。如果用户是第一次运行程序, 在运行开机欢迎页面 SpalshActivity 时便将数据库加载到系统中, 之后再次运行程序, 若系统检测到数据库存在, 特征码库不会被重复加载。

扫描模块和特征码匹配模块是木马检测模块的核心模块, 主要负责对所有已安装程序进行扫描, 然后和病毒库进行比对。如果匹配成功, 则证明程序为木马程序, 并在

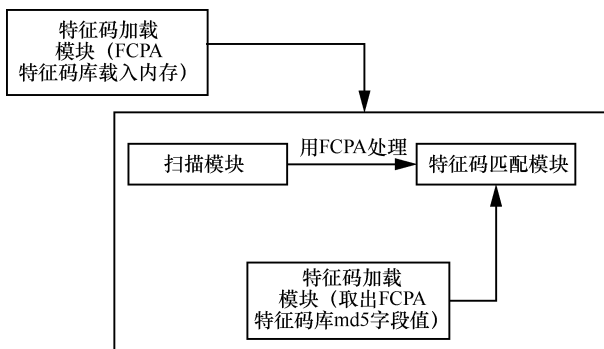


图4 木马检测模块的关系

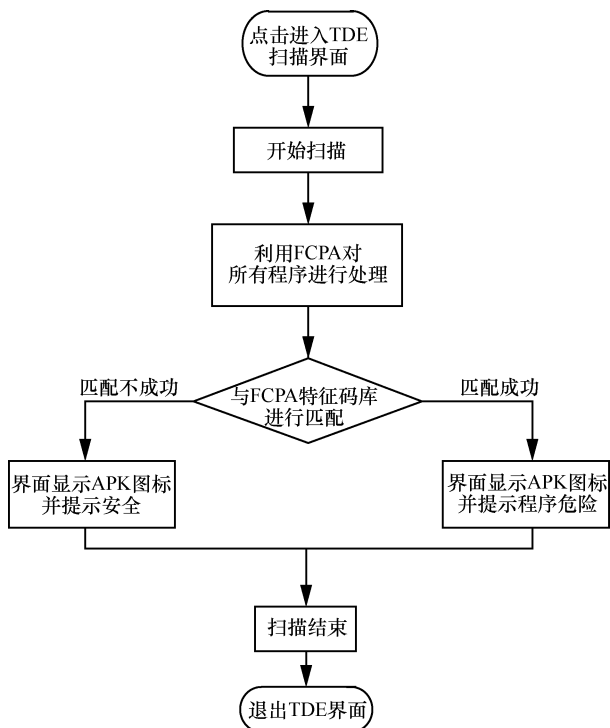


图5 木马检测模块流程

前台界面显示, 点击可以卸载, 如果扫描完成后都没有病毒, 则显示安全。



5 TDE 测试及结果

5.1 测试环境和测试方法

为了能更真实地测试本文所设计的 TDE 系统的稳定性和有效性,测试环境采用真机测试,测试用机为 Galaxy Nexus,手机类型为智能手机,操作系统为 Android 4.3 版本,核心数为双核,基带版本为 I9250XXKK6,版本号为 JWR66Y。

实验对 TDE 检测木马成功率进行测试和记录,测试流程如下。

(1)安装 TDE1.0 版本、360 手机安全卫士。

(2)将 100 个待测程序(70 个恶意程序、30 个正常程序)安装到待测真机上,安装软件的时候利用 ADB 安装命令。本文用“星辰变全本”程序进行举例。图 6 为软件安装过程。

```
C:\Users\Administrator>adb -d install F:\android_malware_fr_xuxian\samples\Cruise
Win\438e0b566eca22e7168711931a958736d9a50118.apk
adb server is out of date. killing...
* daemon started successfully *
77 KB/s (36368 bytes in 0.365s)
pkg: /data/local/tmp/438e0b566eca22e7168711931a958736d9a50118.apk
Success
```

图 6 ADB 安装程序

(3)安装程序时,TDE 新安装程序扫描模块在后台通知用户程序高危,本文为了测试木马检测模块检测效率,先忽视该通知,继续安装。

(4)运行 TDE 和 360 手机安全卫士,等待扫描结果。

(5)后台更新 FCPA 特征码库、TDE,重新安装 TDE 2.0 版本,进行扫描。

5.2 测试结果与分析

部分测试结果见表 2。

在 100 个测试软件中,TDE 1.0 版本共检测出 65 个恶意程序,检测效率达 95%,经过更新的 TDE 2.0 版本,检测效率达 100%,可以有效地检测恶意应用。

6 结束语

Android 操作系统用户数量庞大,已经成为黑客攻击的重要目标。作为攻击 Android 操作系统的主要手段,特洛伊木马具有良好的隐藏性、欺骗性和破坏性。在各类非法程序中已经占到了极大的比重,由木马造成的破坏和损失也越来越严重。因此针对 Android 的木马检测技术的研究也越来越有必要。

本文利用传统的特征码检测技术,设计 FCPA 并开发一套木马检测引擎(TDE),TDE 系统包括木马检测模块、定时扫描模块、新安装程序扫描模块和更新模块,用户体验性好。最后经过 100 个应用程序测试,TDE 检测效率良好,能够有效检测恶意应用。本系统虽然在手机防护上具有一定的优势,但是对特征码库依赖程度仍然较高,下一步需要将动态分析结合到系统中。

参考文献:

- [1] 360 Internet Security Center. China mobile security situation report [R/OL]. [2016-01-29]. <http://zt.360.cn/1101061855.php?did=1101593997&dtid=1101061451.html>.
- [2] GORFIELD P. Similarities for fun and profit [J]. Phrack, 2012(68): 2-20.
- [3] PEIRAVIAN N, ZHU X. Learning for Android malware detection using permission and API calls [C]// IEEE 25th International Conference on Tools with Artificial Intelligence,

表 2 部分测试结果

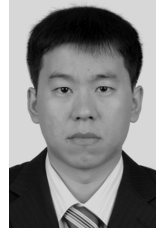
文件名	所包含病毒	检测结果		
		TDE		360 安全卫士
		1.0 版本	2.0 版本	1.0 版本
I'go Reader	a.privacy.adrdataxuan	显示病毒	显示病毒	显示病毒
魔女殿	a.payment.keji.j	显示病毒	显示病毒	显示病毒
万阅公寓	a.payment.keji.j	显示病毒	显示病毒	显示病毒
星辰变全本	a.system.safesys.zf	显示病毒	显示病毒	显示病毒
The Holy Fucking Bible	a.payment.theworld	未显示病毒	显示病毒	显示病毒
Gone in 60 s	a.privacy.gone60.b	未显示病毒	显示病毒	显示病毒
酷 6 视频	a.payment.fakeku6.a	显示病毒	显示病毒	显示病毒
应用宝	无	未显示病毒	未显示病毒	未显示病毒

- Nov 7-9, 2011, Boca Raton, Florida, USA. New Jersey: IEEE Press, 2013: 1082-3409.
- [4] TONG Z F. Static industry of Android malware detection [J]. Jiangsu Communications, 2011, 28(1): 39-42.
- [5] BLASTING T, BATYUK L, SCHMIDT A D. An Android application sandbox system for suspicious software detection[C]// 2010 International Conference on Unwanted Software, Oct 19-20, 2010, Nancy, France. New Jersey: IEEE Press, 2011: 55-62.
- [6] ZHAO Y, HU L, XIONG H. Android malware dynamic analysis scheme based on sandbox information network security [J]. Information Network Security, 2014, 30(12): 21-26.
- [7] JIN Q, WUG X, LI D. Antivirus engine and automatic signature extraction algorithm research [J]. Computer Engineering and Design, 2007, 3(24): 5863-5866.

[作者简介]



夏彬(1994-),男,中国科学院信息工程研究所硕士生,主要研究方向为云安全、大数据安全、移动互联网安全。



邱峰(1983-),男,博士,中国科学院信息工程研究所高级工程师、硕士生导师,主要研究方向为云安全、大数据安全、移动互联网安全。