



移动应用安全批量化检测关键问题探讨

薛立宏¹, 郑京芳², 蒋永成¹, 王渭清²

(1. 中国电信股份有限公司上海研究院 上海 200122; 2. 中国电信集团公司 北京 100035)

摘要: 结合运营商应用商城业务运营安全需求, 分析了移动应用安全检测存在的主要问题, 提出了移动应用安全批量化检测的体系框架, 阐述了架构实施过程中的各项关键问题, 最后, 给出了安全检测服务平台建设策略。

关键词: 移动应用安全; 批量化检测; 安全模型

doi: 10.3969/j.issn.1000-0801.2014.12.020

Discussion on Key Issue of Batch Testing About the Security of Mobile Application

Xue Lihong¹, Zheng Jingfang², Jiang Yongcheng¹, Wang Weiqing²

(1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China; 2. China Telecom, Beijing 100035, China)

Abstract: Based on the service secure operating requirements for telecom carrier's mobile application store, the main problem of batch testing for the security of mobile application was analyzed and the architecture of batch testing was proposed. Then, the key issue about the architecture was discussed. Furthermore, some advices for the constructing of batch testing's service platform were given.

Key words: security of mobile application, batch testing, secure model

1 引言

近年来, 信息安全事件频发, 特别是“棱镜门”事件出现后, 保障网络信息安全已成为国家战略。建立国家级的网络空间安全战略是与美、欧争夺互联网网络空间安全控制权的必由之路, 是保障国家安全和建设更高效国家上层建筑的重要手段。NIST 最新统计显示, 90% 以上的攻击不是来源于网络层面, 而是源于网络层之上的应用层, 这说明移动应用安全已成为网络信息安全的热点问题, 移动应用安全形势也越来越严峻。当前, 移动互联网业务及相关安全问题呈现如下几个特点。

(1) 业务多样性

移动业务呈现多样性, 且云计算的引入带来了移动应用账号盗用、不良信息传播、业务使用安全、云端数据泄露

等问题, 某运营商内部统计显示, 其互联网应用基地每月受到移动应用安全威胁上万次, 发现恶意应用近千个。

(2) 网络开放性

随着移动网络能力开放和网络 IP 化, 传统网络 IP 的脆弱性引入了原来网络上的接入安全、非法访问、网络入侵、流量攻击、不健康内容扩散、匿名电话等一系列问题。

(3) 终端智能化

终端平台的智能化带来了类 PC 的恶意软件、木马、系统破坏、隐私数据泄露、钓鱼攻击等安全隐患和独有的垃圾短信、恶意 WAP-push (推入信息) 等安全问题, 67% 的移动终端使用者感染过病毒。

根据《2013 年我国互联网网络安全态势综述》^[1] 的数据可知, 移动互联网环境有所恶化, 其中 Android 平台恶意程序数量呈爆发式增长, 2013 年新增移动互联网恶意程



序样本达 70.3 万个,较 2012 年增长 3.3 倍,其中 99.5% 针对 Android 平台。移动应用商店、论坛、下载站点、经销商等生态系统上游环节污染,下游用户感染速度加快。移动应用商城作为用户下载应用的主渠道,需要加强相应的移动应用安全检测及运营管理服务。

本文将在分析上述安全形势基础上,结合移动应用的检测需求及检测技术现状,提出移动应用安全检测系统架构,探讨批量化安全检测实施过程中的关键技术,最后给出安全检测服务平台建设策略。

2 移动应用安全检测需求

2.1 移动应用安全检测技术现状

传统网络信息安全领域检测与防御技术主要包括基于特征码的病毒识别技术、基于网络行为的入侵检测技术、配置安全核查和基于防火墙的检测技术等。其中,防火墙技术和入侵检测技术主要面向 IP 网络,对于终端系统及应用的安全防范作用不大;基于特征码的病毒识别技术也主要用于病毒检测;移动互联网安全检测不仅包含上述内容,还包括个人隐私保护检测、信息内容过滤检测等新的层面。

目前,虽然各大应用商城已经开始重视移动应用的安全性,但总的来说移动应用的批量检测还处于起步阶段,检测质量参差不齐,与运营需求以及用户期望还有一定差距,主要需要在以下几个层面提升。

- 应用安全检测标准不统一。各个移动应用商城、安全检测厂商各自为战。以某运营商应用商城为例,分别与多家安全厂商合作,检测标准使用各安全厂商各自的标准,检测要求不一,最终结果难以评判。
- 检测工作缺乏流程化,人工检测所占比例较重。尤其是内容合规性检测还主要依靠人工判定,影响运营效率,需要进行检测工作的流程设计,采用批量化检测技术。
- 灰色行为的量化评估不完善。对一些“擦边球”的灰色行为,如广告以及推送链接内容判定困难,需要对广告、推送链接等进行自动量化评估,而非人工判定。

2.2 移动应用批量化安全检测需求

针对移动应用商城的安全运营,绝不仅仅是检测传统意义上“影响系统的病毒”,更需要对各种应用的业务合规

性、内容合规性、运行合规性进行检测。其内在的商业含义应该是为用户提供安全、合规的使用体验。因此,移动应用的批量化检测需求必须包含业务安全、内容安全、运行安全、数据安全等方面,对应用进行全方位的检测,各种检测要求的介绍如下。

- 业务安全要求:用户在使用应用过程中,在不知情或者未授权的情况下,不得出现恶意扣费、远程控制等影响程序可用性以及用户利益的业务行为。
- 内容安全要求:在应用使用过程中,不得通过文字、图片、视频等显示国家法律、法规禁止的内容。
- 运行安全要求:在软件的下载、安装、运行、更新以及卸载的过程中,不得出现影响移动终端操作系统正常运行的行为。
- 数据安全要求:除了应用功能实现的需求外,不得访问与应用无关的数据,尤其是用户隐私信息等。

3 移动应用批量化安全检测解决方案

3.1 批量化安全检测系统框架

基于检测能力处理与系统管理分离的总体要求,构建移动应用批量化检测的分层架构,实现系统可拓展、信息统一展现和业务快速灵活处理功能,从而适应快速变化的移动互联网安全环境。根据移动商城业务运营的核心需求,提出移动应用批量化安全检测的系统架构,如图 1 所示。

批量化检测的目标是通过引入智能检测工具,实现移动应用安全检测工作的自动化,提升检测精度和运营效率。相对于传统的移动应用安全检测方式,本系统在设计上通过增加爬虫服务器(应用样本自动采集)、内容合规智能检测、与应用商城业务运营系统的自动化接口等模块以及各检测工作流程化等方式,极大地减少了人工验证的环节,从而实现移动应用检测工作的批量化。移动应用批量化检测平台各模块主要功能如下。

- 爬虫服务器:主要用于自动获取外部移动应用样本。既可以抓取第三方应用商城中的应用样本,也可以抓取移动应用社区论坛上的应用样本。
- 应用样本调度系统:对来自爬虫服务器和运营商应用商城内部的移动应用样本进行统一存储,同时调用批量检测控制子系统接口,初步判断各应用样本的安全性,将问题移动应用样本自动提交到安全扫描引擎进行处理。

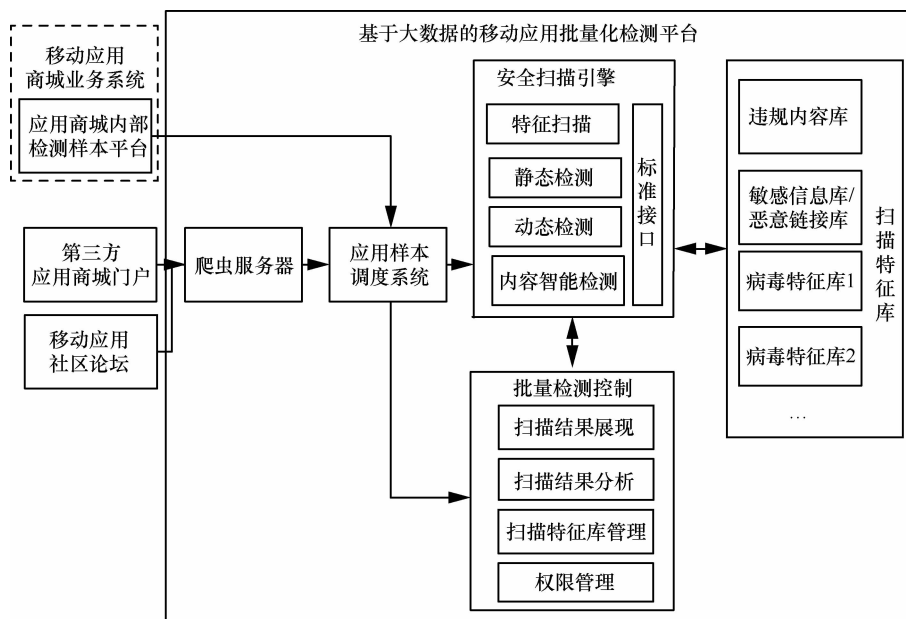


图1 移动应用批量化安全检测系统框架

- 安全扫描引擎:采用各检测工具流程化的方式实现批量检测,检测工具采用特征扫描、静态检测、动态检测与内容合规智能检测等技术手段,自动识别手机应用中的病毒、疑似恶意行为、敏感词,并对手机应用中的联网、发短信等行为进行记录。
- 批量检测控制模块:用于对扫描特征库中违规内容库中的内容进行查询、增加、删除、修改,从而实现内容检测个性化。同时通过标准接口,支持对敏感信息库/恶意链接库、病毒特征库的拓展,实现更加全面、准确的扫描。此外,检测控制模块还接收扫描引擎传来的扫描结果,形成检测报告。
- 扫描特征库:用于移动扫描时移动应用相关违规内容、恶意链接、广告、病毒等特征字段的匹配。特征库可通过开放接口,由第三方安全厂商提供,也可以为业务运营单位自建。

下面基于目前移动应用安全检测中的需求,结合实际的运营需要,重点分析上述移动应用批量化安全检测系统框架中涉及的几个关键技术问题。

3.2 批量化安全检测实施关键技术

3.2.1 面向安全检测的统一标准制订

目前,不同安全厂商对同一款应用的处理结果不一致,因此建设统一的扫描特征库势在必行,统一的特征库将为各种应用的检测提供统一的检测标准。统一的扫描特征库需要从内容统一和结构优化两个层面进行改造。

(1) 标准统一性

各方检测结果不一致,究其根源是检测的各个参与方对于恶意代码、关键字、敏感内容、判定策略等的理解不一致,因此,需要有统一的标准衡量以上工作,如图2所示。建立统一的移动应用安全检测标准主要有3个方面的依据:应用商城实际运营需要,如广告、恶意链接集中治理;政府监管需求,如违规内容过滤;国内、国际对于恶意代码、病毒等的相关标准,对病毒库特征库的规范程度进行判定。

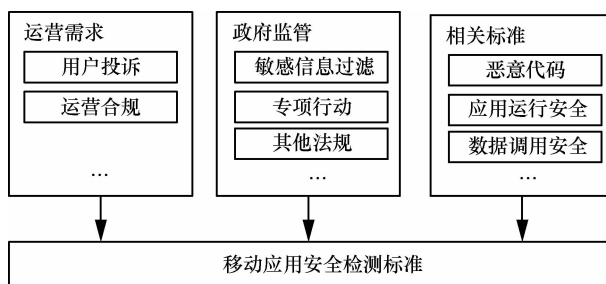


图2 建立统一的移动应用安全检测标准

(2) 结构优化

由过去传统的外包检测模式改为通过系统架构的统一开发接口,将第三方的安全检测能力纳入运营商的架构体系中,目的是当有多个安全检测厂商时,可以对其检测能力进行统一管理和调度,如图3所示。

安全检测统一标准在架构中主要体现在扫描特征库和移动批量检测控制模块的建设上,包括各类型字库、特

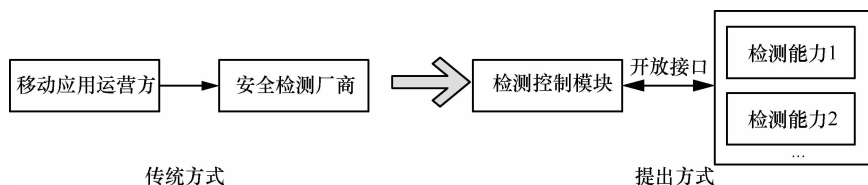


图3 通过开放接口实现检测平台对检测能力的统一管理

征库的全面性以及判定准则的一致性。其技术实现的关键是基于大数据处理技术(如 MapReduce、Hadoop),建立高效、统一的全息数据处理。

3.2.2 移动应用批量检测安全模型

基于国内外应用商城及应用开发者的最佳实践和移动互联网业务的发展趋势,从用户利益诉求、网络安全运行、国家政策要求等层面,针对业务安全、内容安全、运行安全以及数据安全要求,提出了移动应用批量化安全检测的分级指标框架,见表1。

为了能够更加直观地帮助运营商迅速判断移动应用安全能力,需要依据实际的运营情况设计一套运营指标体系与评分系统。

评分系统首先需要对各项三级指标设置相关的分值,并使用 S_i 表示,各三级指标的分值不同代表该类行为的恶意程度不同。如恶意代码类和违规内容类一旦被系统检测发现,该应用不允许上架;本地敏感功能调用行为属于灰色地带,可以根据应用商城业务运营系统的安全策略允许上线或者建议修改后上线。

针对二级指标设置相应的权重指标分别为 $\{W_{11} W_{12} W_{13}\}$ 和 $\{W_{21} W_{22} W_{23} W_{24} W_{25}\}$,则评测的得分可以用式(1)表示。

$$Score = W_{11} \sum_{i=1}^7 S_i + W_{12} \sum_{i=8}^9 S_i + W_{13} \sum_{i=10}^{11} S_i + \cdots + W_{25} \sum_{i=30}^{30} S_i \quad (1)$$

各检测项的分值设置以及所占权重均根据运营需求得出初值,并根据运营情况不断进行调整优化。

3.2.3 图片、视频不良信息的自动化处理

目前,主流的内容信息安全检测技术主要是基于文本的关键字检测,对于图像、音频、视频的安全检测技术尚不成熟。尤其是批量化的自动检测,其检测的误检率、正确率都不能保证实际商用,检测主要还是通过抽检方式进行人工判定。以某软件为例,其宣称图片的正检率大于90%,误检率小于7%,其功能实现的核心算法主要依赖于图片颜色分布以及内部图形的形状,在实际检测过程中具有较高的误检率。

因此,针对图片、音视频信息的批量化检测,其整个检测的核心是如何通过自动化检测技术,协助人工进行快速

表1 安全风险指标

一级指标	二级指标	三级指标
恶意软件 V_1	恶意代码 V_{11}	恶意扣费 S_1
		远程控制 S_2
		恶意传播 S_3
		恶意资费消耗 S_4
		系统破坏 S_5
		诱骗欺诈 S_6
		流氓行为 S_7
恶意联网行为 V_{12}	违规内容 V_{13}	自动联网 S_8
		反复请求联网 S_9
		政府监管 S_{10}
		运营限制 S_{11}
		电话拨打 S_{12}
		三方通话 S_{13}
		短/彩信发送 S_{14}
		邮件发送 S_{15}
		蜂窝网络数据连接 S_{16}
		WLAN 数据网络连接 S_{17}
风险软件 V_2	通信功能调用(经提示) V_{21}	蓝牙连接 S_{18}
		定位功能 S_{19}
		通话录音 S_{21}
		本地录音 S_{22}
		拍照摄像 S_{23}
		用户数据读写 S_{24}
		搜集用户行为 S_{25}
		用户上传数据 S_{26}
		插件安装 S_{27}
		软件捆绑 S_{28}
强制捆绑 V_{23}	加密通信 V_{24}	VPN 软件 S_{29}
		信令风暴 V_{25}
加密通信 V_{24}	信令风暴 V_{25}	频繁心跳连接 S_{30}

批量判定,从而减少人工工作量。图片、音视频内容检测的主要关注点见表2。

因此,针对图片、音视频的批量化检测可以从如下两个方面入手。

- 溯源性角度。其目的主要是获取图片、音视频文件的来源信息,并进一步对该来源的信用信息、违规信息等进行了解,从而让人工检测人员更准确地判

表 2 图片、音视频内容检测的关注点

内容可控	检测可执行
文件来源可控;内容信息可控	- 减少人工检测量; - 标注图片、音视频文件内容问题可能出现点; - 结合现有文本过滤系统

定需要检查的部分。图片、音视频溯源主要包括两个方面:一是应用开发者信息,通过对应用采取反编译等手段,获取应用者的数字签名证书,同时可以结合开发者提交给商城运营商的其他信息,判定该应用开发者是否可靠、是否有违规记录等;二是对于从其他应用商城中获取的应用,开发者信息可能获取不到,可以通过数字水印技术对图片、音视频等信息添加相关标识,在违规信息查处时,可以提供相关的来源信息,做到内容迅速定位。

- 可操作性角度。针对大量的图片、音视频,需要减少其人工检测时间、人工检测难度,进而提高检测效率。可操作性角度可以从两个方面考虑:一是针对视频文件,可以采取视频截帧的方法,通过一定的抽样概率,将视频转换为图片进行检查;二是针对含有文字的图片,可以通过文字、数字的挖取技术,识别图片中的文字,进而与违规内容库中的关键字进行匹配,判定图片的可疑程度。

通过上述的两个角度和提出的方法,可以让检测人员快速地判定违规信息,减少人员所耗精力、减少承担不必要的责任,进而提高检测效率。

3.2.4 应用更新分组批量化安全检测

基于实际运营要求,花费大量时间完整地检测一款移动应用虽在技术上是可行的,但实际情况并不允许。因此,批量化安全检测的设计理念应是“整体判定”,然后是“局部检查”。批量化安全检测主要包括 3 个部分:黑白名单匹配、静态检测和动态检测,如图 4 所示。

黑白名单匹配过程如图 5 所示。首先进行黑白名单匹配,在这一过程中运营商根据已发现的恶意软件、违规软件(黑名单模式)或已确定安全的软件(白名单模式)对检测应用进行直接匹配。如采用散列算法先对黑白名单中的应用计算散列值;然后,对待检测的软件计算散列值;最后,比较两个散列值,可直接判定是否是恶意(违规)软件。对于没有匹配上黑白名单的应用,将进入静态检测环节。

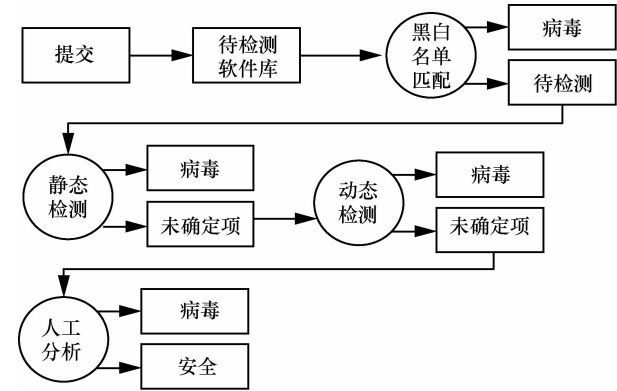


图 4 移动应用批量化安全检测步骤

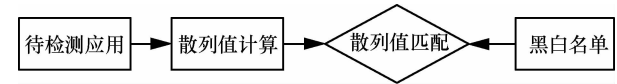


图 5 黑白名单匹配过程

静态检测匹配过程如图 6 所示。静态检测将移动应用分组通过逆向工程技术生成应用程序的源码,在源码中查找是否有敏感 API(application programming interface,应用程序编程接口)。敏感 API 包括发送短信,获取 IMEI 号、电话号码、IMSI 号、上网行为,使用 GPS 定位、打开摄像头、操作联系人、静默安装等。此外,逆向分析结果中还包括移动应用中包含的文字信息、图片、音视频、URL 链接等内容。将以上信息与扫描特征库中的内容进行匹配,判定是否为恶意(违规)应用。

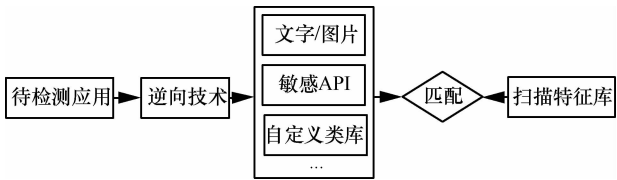


图 6 静态检测匹配过程

静态检测实现主要依赖的技术是反编译技术,基于 Android、iOS、Java 的反编译软件技术均在不断发展,如基于 Android 系统的反编译工具有 dex2jar、JAD Java Decompiler 等。

由于静态检测使用的扫描特征库存在一定局限性以及恶意代码自身的隐蔽性等,静态检测不能够完全过滤恶意软件,因此还需要基于用户行为分析的动态检测方法进行判定。

动态行为获取过程如图 7 所示。动态检测方法是指将待扫描的文件安装至沙盒后,捕捉其动态行为。捕捉的行为包括查询联系人、插入联系人、添加/删除联系人、更新联系人、发送彩信、发送短信、读取本机 IMSI 号、读取本机

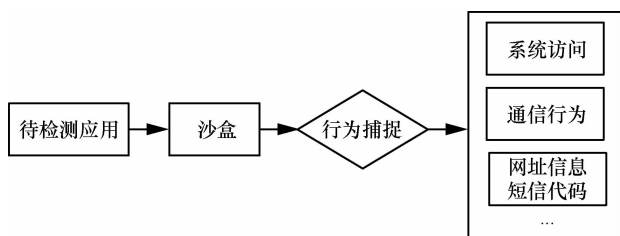


图7 动态行为获取过程

电话号码、读取本机 SIM 卡号、访问网址、访问蓝牙行为、打开摄像头预览、打开 GPS、访问位置信息等。

动态检测过程中,使用的主要关键技术是沙箱技术,该技术主要是模拟移动应用的运行环境(如 Android 操作系统环境),让移动应用在其模拟的环境中运行。在动态检测过程中,安全沙箱主要模拟如下 4 个方面的场景。

- 软件的逻辑控制,如应用运行的安装、卸载。
- 情景的逻辑控制,如模拟地理位置信息变化等。
- 行为的逻辑控制,如应用的联网动作、短信的发送动作等。
- 数据的逻辑控制,如本地文件数据的改动、存取等。

通过黑白名单匹配、静态检测和动态检测,对于还不能做出判定的移动应用,需要通过抽查的方式进行人工检测。

3.3 安全检测服务平台建设策略

移动应用上传至应用商城平台后,通过该商城业务运营系统中的安全检测触发接口,将检测任务传递给移动应用批量化检测平台,从而触发批量化检测平台进行应用检测。在其过程中,运营商可以通过查询接口进行任务进度查询,检测平台将其运行状态返回给运营商。在检测完成后,检测平台将扫描结果(如解析出的图片、链接等)返回给业务系统,同时业务系统通过接口获取批量化检测报告。

因此,移动应用商城业务系统与移动应用批量检测平台之间需要 4 种类型的接口:扫描调用接口、扫描进度查询接口、扫描任务反馈接口和报告查询接口,如图 8 所示。接口采用 HTTP+XML 的形式。

接口数据采用数据发送(post)请求方式,请求及返回的数据消息格式分别见表 3、表 4。

运营商在得到该应用的批量化检测结果后,再通过相关安全运营流程对相应的结果进行处理,如对图片或音视频等进行检测、对广告信息或恶意链接进行判定等。移动应用批量化安全检测系统需要很好地嵌入应用商城的业务运营流程中。

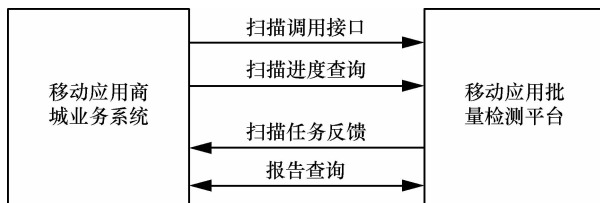


图8 移动应用商城批量化检测能力调用

表3 批量检测请求消息格式

标识	类型	描述
ID	string	检测批次 ID
FileRoute	string	待检测分组的目录路径
FileFolder	string	待检测分组文件夹
Type	int	全扫描; 静态扫描; 动态扫描; 进度查询; 报告查询

表4 批量检测请求返回消息格式

标识	类型	描述
ID	string	检测批次 ID
FileFolder	string	待检测分组文件夹
Vendor	string	能力提供者
Code	int	相应状态包括: 20000:成功; 20001:目录不存在; 20002:文件不存在; 20003:用户名、密码错误
Score	int	检测得分
Msg	string	返回结果消息

4 结束语

移动应用批量化检测系统既要考虑实际运营的需要,又要考虑相关政府的需求,同时还要满足国际、国内的安全标准,从整体上考虑其技术实现架构。在批量化检测平台投入运营之后,还需要不断对检测标准、关键技术、统一业务运营接口等方面持续优化,从而更好地提升移动应用安全检测准确性、检测效率和运营效率。

参考文献

- 1 CNCERT. 2013 年我国互联网网络安全态势综述, 2013
- 2 陈建民. 基于行为的移动应用程序安全检测方法研究. 计算机与工程设计, 2012, 33(12): 4476~4481
- 3 中国电信集团公司. 2013 网络信息安全技术发展白皮书, 2013
- 4 孙泽锋. 移动互联网发展技术与安全分析. 电信科学, 2011,

(下转第 152 页)