

## 基于众包和机器学习的移动应用隐私评级研究\*

张贤贤<sup>1</sup>, 王浩宇<sup>1+</sup>, 郭 耀<sup>2</sup>, 徐国爱<sup>3</sup>

1. 北京邮电大学 计算机学院 智能通信软件与多媒体北京市重点实验室, 北京 100876
2. 北京大学 信息科学技术学院 软件所 高可信软件技术教育部重点实验室, 北京 100871
3. 北京邮电大学 网络空间安全学院, 北京 100876

## Privacy Rating for Mobile Apps Based on Crowdsourcing and Machine-Learning Techniques\*

ZHANG Xianxian<sup>1</sup>, WANG Haoyu<sup>1+</sup>, GUO Yao<sup>2</sup>, XU Guo'ai<sup>3</sup>

1. Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
  2. Key Laboratory of High-Confidence Software Technologies (Ministry of Education), School of Electronics Engineering and Computer Science, Software Institute, Peking University, Beijing 100871, China
  3. School of Cyber Space Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
- + Corresponding author: E-mail: haoyuwang@bupt.edu.cn

ZHANG Xianxian, WANG Haoyu, GUO Yao, et al. Privacy rating for mobile Apps based on crowdsourcing and machine-learning techniques. *Journal of Frontiers of Computer Science and Technology*, 2018, 12(8): 1238-1251.

**Abstract:** Mobile Apps frequently request access to sensitive information without users' knowledge. Whether the

---

\* The National Natural Science Foundation of China under Grant Nos. 61401038, 61772042, 61702045 (国家自然科学基金); the National High Technology Research and Development Program of China under Grant No. 2015AA017202 (国家高技术研究发展计划(863计划)); the Frontier and Key Technology Innovation Project of Department of Science and Technology of Guangdong Province under Grant No. 2016B010110002 (广东省科学技术厅前沿与关键技术创新项目); the Key Technology Innovation Project of State Grid Corporation of China under Grant No. SGRJXTKJ[2017]265 (国家电网公司科技项目); the Youth Research and Innovation Program of BUPT under Grant No. 2017RC40 (北京邮电大学青年科研创新计划专项); the Project of Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia under Grant No. ITSM200601 (北京邮电大学智能通信软件与多媒体北京市重点实验室项目).

Received 2017-10, Accepted 2017-12.

CNKI网络出版: 2017-12-08, <http://kns.cnki.net/kcms/detail/11.5602.TP.20171208.1500.012.html>

sensitive information should be granted is related to the purpose of permission use. This paper proposes a privacy rating model to assess the privacy behavior of Android Apps based on the purpose of permission use and users' expectation. Based on 16651 crowdsourcing data of 421 users for the triple <app, permission, purpose>, this paper trains a privacy rating model based on machine-learning techniques. Then this paper uses static analysis to infer the purpose of permission use in the App, and grades the privacy by using privacy rating model. The experiment results show that the privacy rating model can achieve 80.7% accuracy. By applying the privacy rating model to 11931 Apps from Google Play, the results show that around 8% of Apps have serious privacy risks.

**Key words:** mobile privacy; permission; purpose; privacy score; machine learning

**摘 要:**移动平台上广泛存在权限滥用的问题,在用户不知情的情况下,很多应用会获取并泄露用户的隐私信息。隐私信息的使用是否合理与其使用意图相关。为了实现基于用户期望对应用的敏感行为进行隐私评分,提出一种基于应用敏感权限使用意图的隐私评级模型,基于众包数据中421个用户16 651个数据对不同的<应用,权限,意图>组合的评分,使用机器学习技术建立准确的隐私评级预测模型。通过静态分析应用使用敏感权限的意图,使用隐私评级模型对应用进行评分。实验结果表明,所建立的隐私评级模型能够达到80.7%的准确率。通过将隐私评级模型应用于来自谷歌商店的11 931个应用,结果表明大约8%的应用存在严重的隐私风险。

**关键词:**手机隐私;权限;意图;隐私评分;机器学习

**文献标志码:**A **中图分类号:**TP311

## 1 引言

在移动智能终端和多样的移动应用给用户带来便利的同时,移动平台上各种新的安全和隐私问题也日益凸显。安卓系统使用权限模型来控制对隐私信息的访问。然而,移动平台广泛存在权限滥用问题,很多应用经常申请不必要的敏感权限,使用户隐私信息面临被泄露的风险。很多应用会在用户不知情的情况下获取并泄露隐私信息。

近年来很多研究工作关注于应用分析和移动系统的隐私保护<sup>[1-6]</sup>,虽然大部分工作都可以检测隐私泄露或者进行隐私保护,但它们没有针对隐私信息使用的原因深入探究,导致用户的期望与应用行为的差距迟迟未能解决。应用的敏感行为是否恶意以及是否应该被允许都跟其使用权限的意图有关。例如,大部分用户会认为地图应用使用位置信息进行导航是正常行为,但并不愿意位置信息被用于提供定制化广告服务及第三方分析。因此相关工作<sup>[7]</sup>仅分析应用的敏感权限来对应用进行隐私评级的工作是不可靠的,功能丰富的应用使用的敏感权限很多,

导致隐私评分很低,但这些应用绝大多数都不是恶意的。

很多研究尝试解决用户期望与应用行为的差异。WHYPER<sup>[8]</sup>和AutoCog<sup>[9]</sup>基于用户所期望的应用行为,提出基于自然语言处理技术在应用描述与其申请权限之间建立映射关系,并用这种映射关系量化应用功能和行为之间的差异性。CHABADA (checking App behavior against descriptions of Apps)<sup>[10]</sup>基于描述对应用聚类,并以此分析同类应用中API调用异常的应用,进而寻找潜在的恶意应用。但研究表明,超过90%的应用都没有完整地在描述中说明其使用权限的原因<sup>[9]</sup>。

Lin等人<sup>[11-12]</sup>提出使用众包技术研究用户对不同的<应用,权限,意图>组合的接受程度。通过检测应用中隐私信息的使用位置(第三方库或应用核心代码),即可分析隐私信息使用的意图,即是否用于第三方库(例如广告推荐、社交网络、第三方分析等)还是应用核心功能需要。研究结果表明用户的期望和隐私使用的意图都会影响用户对应用隐私行为的接受

程度。本文受此工作启发,在用户众包数据的基础上,建立应用隐私评级模型,能够自动化对应用的隐私威胁分析和评级。

为实现准确的应用隐私评级模型,本文通过静态分析研究敏感权限的使用及其使用意图,基于众包数据中用户对不同的<应用,权限,意图>组合的评分,为应用提取多维度特征构建特征向量,并对数据集进行预处理,使用机器学习技术建立准确的隐私评级预测模型。

本文实现了一个应用隐私评级工具,能够准确地分析应用中实际使用的敏感权限及其使用意图并提取应用相关特征为应用构造特征向量,输入预测模型进行隐私评级。实验表明,模型预测准确率能够达到80.7%,通过将该预测模型应用于谷歌商店的11 931个应用,结果表明约8%的应用存在严重的隐私风险。

本文主要有以下贡献:

(1)建立了一个基于权限使用意图的移动应用隐私风险评级预测模型。该模型可以预测用户所关心的隐私信息使用的问题,并以评分等级的方式展现,直观且高效。

(2)实现了移动应用的权限使用意图的分析。首先获取应用中使用的敏感权限,然后为每一个敏感权限分析其使用意图,将敏感权限和使用意图作为预测模型的核心特征。

(3)实现一个自动化的隐私评级工具并将其应用于大量应用中验证,该工具可准确为每个应用预测隐私风险评级。

## 2 研究背景和相关工作

### 2.1 安卓的权限模型

本文的研究基于安卓权限模型,安卓权限框架从两方面保护用户隐私:(1)限制应用访问用户敏感资源;(2)在用户安装应用之前帮助用户做选择。安卓权限分为系统权限和特殊权限授权,系统权限包括普通类型和敏感类型,其中普通类型并不直接威胁到用户的隐私,直接在AndroidManifest.xml文件权限申请里注册,系统会默认授权。而敏感类型的权限可以让应用访问用户敏感数据,不仅需要

在AndroidManifest.xml中注册,同时在使用的时候需要向系统请求授权。安卓在6.0版本以前采用默认的授权模式,即所需权限一次性申请,用户在安装应用的时候系统采用默认授权,且一旦授权便不可撤销。这种授权模式没有考虑用户,用户要么选择接受所有的权限,要么为了拒绝授权不得已放弃安装应用。对用户来说,一方面用户体验很差,另一方面不能控制授予应用的权限是否会被合理使用,存在隐私泄露风险。安卓在6.0版本之后授权模式升级,新的授权模式只有需要授权的时候才请求用户是否授权,并且是在程序运行时授权而非安装时授权,这种模式赋予用户自主选择的权利,例如用户可以拒绝某些应用访问记录设备位置的权限。

授权模式升级之后用户拥有自主选择权,但这并不能解决用户隐私信息泄露的风险,安卓应用安全和隐私的设计依赖于用户能够理解所有权限,然而用户在没有长时间使用应用之前并不能了解应用的敏感行为,研究<sup>[13-14]</sup>表明,用户很少注意到权限相关的问题,Felt等人<sup>[13]</sup>进行的两次研究发现用户对权限的关注度和理解率都较低,这表明普通用户一方面缺乏对应用权限的关注,另一方面缺乏专业领域知识,对于应用所请求的权限并不了解其潜在风险,以及这些权限在应用中如何被使用。因此安卓使用的权限许可申请并不能帮助大多数用户做出良好的安全策略选择。

### 2.2 移动应用的隐私风险分析

针对隐私信息是否泄漏的问题,Enck等人<sup>[2]</sup>通过修改Dalvik虚拟机实现动态污点分析工具Taint-Droid。该工具可以将敏感数据标记为污点源,然后跟踪污点数据,根据污点数据是否被泄露来判断应用是否存在隐私泄露。RiskMon<sup>[6]</sup>提供一种连续而自动化的风险评估框架,通过收集用户对应用权限使用的反馈,从应用程序的元数据中构建模型,然后使用机器学习方法来评估应用风险。但RiskMon会在运行时跟踪应用API调用,耗时较多且大量占用手机资源,例如CPU和内存。

### 2.3 权限理解

文献[1,15-18]研究用户对权限的理解,用户通常会忽略安装应用时的安全警告<sup>[14-15]</sup>,同时由于对权



限理解不足<sup>[1,15]</sup>,以及并不了解应用所收集的隐私信息<sup>[15]</sup>,导致用户不能有效地对应用权限进行管理。

Liccardi 等人<sup>[17]</sup>提出修改 Google Play 的权限界面,为应用增加隐私泄露度量(即隐私评分),其目的是让无经验的用户能够理解应用权限。Sarma 等人<sup>[18]</sup>提出对应用中异常权限的使用产生安全警告并提醒用户的方法。如果应用请求的权限也被同类其他应用所请求,则说明该权限为应用所需,否则说明该应用的权限请求异常。Amini 等人<sup>[11]</sup>提出结合众包以及动态分析技术,帮助用户理解隐私信息的使用以及标记应用异常行为。Ismail 等人<sup>[19]</sup>使用众包的方法研究应用在不同权限设置下的可用性,及用户对应用可用性的接受程度,并以此为不同用户推荐权限设置。

## 2.4 用户期望与应用行为差距分析

基于用户能直观看到的应用描述、应用界面(UI)等信息,近期的研究工作尝试从用户角度出发,分析并解决用户期望与应用行为的差异。这些研究工作可以分为3类:应用描述与应用行为的一致性分析<sup>[8,20]</sup>,应用界面与应用行为的一致性分析<sup>[10]</sup>,以及应用功能与界面的一致性分析<sup>[21-24]</sup>。

基于应用描述的分析:WHYPER<sup>[8]</sup>基于用户心中所期望的应用行为,提出一种基于自然语言处理的方法在应用描述和应用申请权限之间建立一种映射关系,并用这种映射关系量化应用功能和应用真实行为之间的差异性。在此基础上,AutoCog<sup>[9]</sup>提出一种结合机器学习和自然语言处理的方法,利用大量数据生成应用描述和应用申请权限的关系模型,从而使分析结果更精准和全面。CHABADA<sup>[10]</sup>通过分析应用描述与应用实际功能的差别,寻找潜在恶意应用。基于应用描述进行聚类,然后找出同类别应用中API使用异常的应用。但当前应用描述更多的是关于应用的功能,而没有涉及到应用中隐私泄露的行为。在此基础上,Zhang 等人<sup>[20]</sup>提出自动化工具DESCRIBE ME,通过程序分析和自然语言处理技术为应用自动生成有关隐私信息泄露的描述。

基于应用UI界面的分析:PERUIM<sup>[21]</sup>(permission-UI mapping)和AsDroid<sup>[22]</sup>使用程序分析技术识别与应用界面元素相应的权限,分析UI权限与UI组件中

文本信息的差异,从而检测潜在的恶意应用。基于应用敏感行为相关的GUI操作序列,AppIntent<sup>[23]</sup>分析隐私信息的泄露是否为用户触发,从而检测潜在的恶意行为。尽管这些研究工作尝试从用户角度出发,分析并解决用户期望与应用行为的差异,但大部分应用并没有完整的应用描述或者UI描述信息。例如,超过90%的应用都没有完整地在描述中说明使用权限的原因<sup>[25]</sup>。

## 3 研究方法

本文基于Lin 等人<sup>[11]</sup>针对用户对移动应用隐私期望的众包数据集,为每一个应用提取多维度的特征,并结合<应用,权限,意图>三元组以及用户评分构造特征向量,以该数据集作为训练集构建并训练预测模型,并通过预测分数和真实分数的均方误差值来选择预测效果最好的模型,最后设计良好的隐私评分等级映射算法,将预测分数映射为评分等级,实现隐私评级预测工具,整体流程图如图1所示。

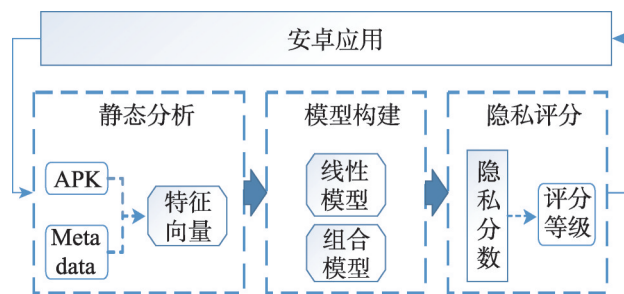


Fig.1 Framework of privacy rate prediction model

图1 隐私评级模型框架

### 3.1 众包数据集

本文的研究基于Lin 等人<sup>[11]</sup>针对用户对移动应用隐私期望的调研结果数据集。他们采用众包的方式调查用户对移动应用隐私信息使用期望的真实数据,因隐私策略的复杂性或是用户付出的时间与收益不平衡等原因,很少有用户去阅读终端用户许可协议或者隐私策略,但是众包技术可以很好地解决这些问题。通过提供清晰的解释来降低权限理解的复杂度,关注应用的哪些行为打破了用户的期望。首先要求参与者阅读由谷歌商店提供的关于应用的基本信息、截图和描述,然后一组参与者会被询问关

于应用权限使用期望相关问题,另一组则被告知权限的具体使用意图等信息,最后要求参与者为应用程序权限相关行为指定舒适度评分,评分范围从-2(非常不舒适)到+2(非常舒适)。

Lin 等人<sup>[11]</sup>在2014年2月研究了谷歌商店中根据星级评分排名前100的免费应用程序,采用众包的方式调查用户对移动应用隐私信息使用期望的真实数据。为获取更多数据,在2014年的8月进行第二轮研究,扩充众包数据集,使得数据集中任意一个敏感权限或者使用意图至少有20个<应用,权限,意图>三元组。Lin 等人<sup>[11]</sup>的两次研究共收集来自真实用户的44 676份有效数据,并对原始数据进行整理和清洗,使得最终的数据集中共包含来自421个真实用户关于414个应用的16 651份问卷结果,这也是本文所用的数据集。

### 3.2 隐私评级模型

隐私评级模型的实现分为3个主要步骤:第一步为应用构造特征向量。训练集中的数据只需要通过静态分析等技术从元信息中提取与应用相关的特征和用户对应用的反馈信息,这些特征数据结合众包数据集中的<应用,权限,意图>以及评分可以为应用构造出特征向量。但对于需要进行预测的应用应首先获取应用中实际使用的所有敏感权限,并为敏感权限分析使用意图,然后再提取应用相关的其他特征数据构造特征向量。第二步构建预测模型,利用不同的机器学习算法建立回归模型,通过参数调整和优化使模型稳定且实现最好的预测效果,通过对比分析每个回归模型的预测结果,选择性能最好的回归模型作为隐私评级工具的预测模型。第三步评分等级映射,建立合理的隐私评分计算以及等级映射算法,为应用最终确定一个隐私评级。

### 4 特征分析和提取

首先为每一个应用提取特征并构造特征向量。通过API分析和静态分析等技术获取应用相关的特征数据,包括权限分析、权限使用意图分析以及元数据的分析等,提取不同的特征为应用构建特征向量。特征数据来源于两部分,分别是使用爬虫从谷

歌应用商店爬取应用的apk文件和相关的元数据。通过反编译apk文件可以获取应用实际使用的敏感权限,元数据包含应用相关的一些特征数据,例如应用的描述、下载量、用户的评价量等特征。

#### 4.1 权限和使用意图

静态分析是移动应用分析中最常用的技术,通过静态分析可以实现敏感权限分析和使用意图分析。本文中使用反编译工具ApkTool将原始的apk文件反编译为中间代码。一方面可以获取到应用权限申请AndroidManifest.xml,该文件中申请的权限在应用代码中或第三方库代码中使用,因此通过解析该文件可以获取到该应用相关的一些特征,例如各个组件的数量、申请权限的数量、安装包大小等特征。此外还可以通过分析smali格式的中间代码中API的调用关系获取应用实际使用的敏感权限。隐私信息的使用是否合理与其使用意图相关,因此本文针对应用中频繁使用的11个敏感权限分析其不同应用中的使用意图,获取到<应用,权限,意图>三元组,表1列出本文中分析的11个应用常用的敏感权限。

Table 1 Sensitive permissions

表1 敏感权限

权限	描述
READ_PHONE_STATE	访问电话状态
ACCESS_FINE_LOCATION	获取精确位置
GET_ACCOUNTS	访问Gmail列表
CAMERA	访问摄像头
READ_CONTACTS	访问通讯录
MANAGE_ACCOUNTS	管理账户
RECORD_AUDIO	录音
SEND_SMS	发送短信
BLUETOOTH	使用蓝牙
AUTHENTICATE_ACCOUNTS	验证账户
ACCESS_COARSE_LOCATIONS	粗略获取位置信息

针对上述11种敏感权限,本文分析总结9种常用的使用意图,这个意图分类是经过相关工作<sup>[11-12]</sup>验证的常见权限使用意图,如表2所示,其中第三方库代码中权限的使用意图可以通过第三方分析工具LibRadar<sup>[26]</sup>分析提取。LibRadar是一种基于聚类的第三方库检测工具,使用该工具可以分析出应用中引用

的第三方库的数量以及不同的第三方库中使用的权限以及使用该权限的意图。但是一个权限仅在应用本身代码中被使用而没有在第三方库中被使用,则将其使用意图默认标记为 INTERNAL,表示该权限只在应用程序本身的代码中被使用。

Table 2 Purposes of permissions

表2 权限使用意图

使用意图	标记
功能实现	DEVELOPMENT_AID
广告	TARGETED_ADS
社交	SOCIAL_NETWORK
游戏	GAME_ENGINE
应用分析	MOBILE_ANALYTICS
工具	UTILITY
付款	PAYMENT
地图和位置服务	MAP
内部使用	INTERNAL

4.2 特征类型

特征向量可以唯一地表示一个应用的特定权限的使用意图,在创建预测模型之前首先创建特征向量集。本文提取多种特征数据为应用构造特征向量,其来源于两部分,一部分特征数据来自于“元数据(Metadata)”,即可以直接从应用商店获取与应用本身相关的特征数据,例如应用的下载量、评论数量、星级排名等数据;另一部分特征数据则需要通过静态分析从应用安装包文件中提取。本文将应用的特征分为3种类别,表3列出了特征数据的分类、名称及来源。

Table 3 Type of features

表3 特征类型

特征类型	特征名称	来源
受用户影响特征	评论数,星级(1,2,3,4,5)数,星级排名,下载量,内容评级	元数据(Metadata)
客观特征	是否提供开发者相关信息(网站,邮箱,隐私策略的链接),安卓组件(receivers, activities, service providers, services)的数量,描述,应用分类等	(1)元数据(Metadata) (2)静态分析(apk)
隐私相关特征	权限,使用意图,权限数量,第三方库数量	静态分析(apk)

(1)受用户影响特征

受用户影响的特征这部分的数据主要来源于用户的反馈。用户可以在应用商店对应用进行星级投票,星级分为5个等级,五星级别最高,五星的数量越多表明该应用受欢迎程度越高,因此不同星级的数量可以在一定程度上反映用户对该应用的喜好程度。“下载量”特征更直接体现该应用的实际用户量。同类应用中,一个应用的下载量越大表明该应用在同类应用中越受欢迎,其使用用户隐私信息的可靠性更高,用户对该应用的评分可能会较高。特征与隐私评分的相关性分析表明下载量与隐私评分呈正相关且相关性较强。

(2)客观特征

客观特征与应用自身相关,应用是否提供介绍开发者的网站、姓名、开发者的邮箱等信息。同一个开发者可能参与开发过不止一个应用。例如,一个开发者开发过恶意应用,那么其他有该开发者参与开发的应用也应该引起用户对该应用中隐私信息使用情况的关注,与该开发者相关的应用评分倾向可能会比较相似。“描述”这个特征用最简洁的语言表明该应用的主要功能,该特征基本是在叙述应用的核心功能,很少体现出应用中权限的使用情况,因此本文在建模过程中不考虑“描述”特征。“应用分类”则表示应用的类型,例如游戏类或地图类,同类应用实现的功能会比较相似,因此可以认为同类型的应用可能会使用相似的敏感权限集合。

(3)隐私相关特征

隐私相关特征与用户隐私信息相关,应用中使用的敏感权限及其使用意图,一定程度决定用户对应用的接受程度以及评价,申请的权限可以在应用本身代码中使用也可在第三方库中使用,或是申请之后并不使用。因此,需通过分析应用中敏感API的使用,提取应用实际使用的敏感权限,及该应用第三方库中用到的敏感权限,并且分析每一个敏感权限分析在应用中的使用意图。

4.3 相关性分析

通过计算每个特征与评分的皮尔森系数分析每个特征与应用隐私评分之间的相关性,皮尔森系数



值范围为 $[-1,1]$ ,绝对值越大表示相关性越强,正值表示特征和评分之间是正相关,负值表示特征和评分之间是负相关。特征与隐私评分相关性分析结果如图2和图3所示,图2表明有74%的特征和隐私评分之间呈正相关性,只有26%的特征和隐私评分之间呈负相关性。

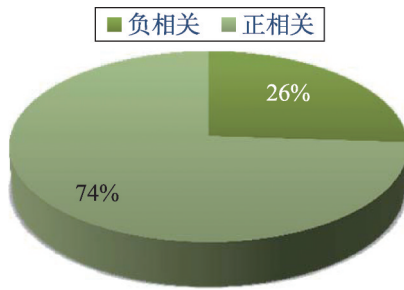


Fig.2 Features correlation

图2 特征相关性

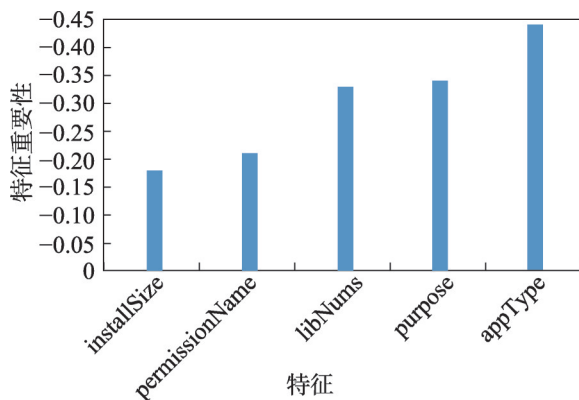
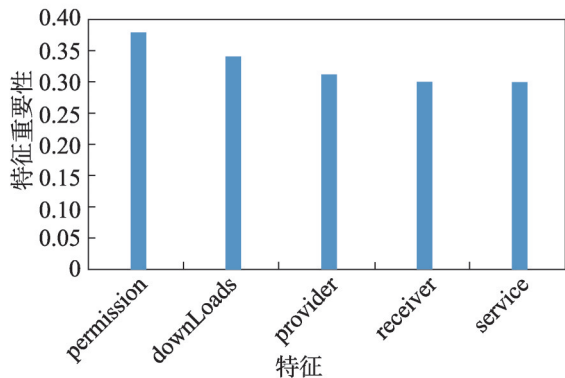
(a) Negative correlation between features and privacy score  
(a) 特征与隐私评分呈负相关(b) Positive correlation between features and privacy score  
(b) 特征与隐私评分呈正相关

Fig.3 Correlation between features and privacy score

图3 特征与隐私评分相关性

图3所示为负相关性最强的5个特征(a)图和正相关性最强的5个特征(b)图,其中负相关性最强的3个特征分别是应用类型(appType)、权限使用意图(purpose)和第三方库的引用数量(libNums)。权限使用意图与隐私评分呈负相关性,且相关性较强排第二,第三方库引用数量与隐私评分的相关性为-0.33,表明第三方库和隐私评分之间呈负相关且相关性较强,引用第三方库数量越多,对应用隐私评分所起的消极的作用越强。而正相关性最强的两个特征分别是权限数量(permission)和应用下载量(downLoads),应用下载量与隐私评分的相关性为0.34,表明下载量对应用隐私评分有着较为积极的作用,下载量越大,则应用受欢迎程度越高,其他3个特征(provider, receiver, service)都表示安卓组件的数量,安卓组件数量与应用隐私评分呈正相关性,且3种组件对隐私评分的影响程度无明显差异。

#### 4.4 特征向量预处理

本文采用长度为24的特征向量表示应用。由于特征数据类型既有数值类型特征也有非数值类型特征,需对原始的特征数据进行预处理,不同类型的特征采用不同处理方式。

##### (1) 归一化

归一化是一种简化计算的方式,即将有量纲的表达式经过变换化为无量纲的表达式。归一化后的数据会根据需要被限制在某一个范围之内。本文中用到的数据类型的特征数据级别差距较大,例如“下载量”这一特征值往往是上万甚至百万级别,而“组件数量”则是个位数级别。归一化处理一方面能减小某一维数据对结果影响太大,另一方面可以加快程序的运行速度。数据归一化的实现方式有多种,本文采用线性转换函数进行数值类型数据的归一化处理。

转换公式如式(1)所示,其中 $x$ 、 $y$ 分别表示转换前和转换后的值, $Mean$ 表示均值, $Std$ 表示标准差,归一化处理之后的数据范围限定在 $[-1,1]$ 之间,且均值为0。

$$y = \frac{x - Mean}{Std} \quad (1)$$

## (2) 标签化

非数值类型的特征数据有应用的分类、权限、使用意图等特征。本文对训练集中非数值化的特征进行标签化的处理之后参与对模型的训练,每一个需要预测评分应用的此类特征都需要预先进行标签化处理之后,才能输入模型进行隐私评分预测。本文采用独热编码的方式标签化非数值特征,对于每个非数值特征,假设存在  $N$  个可能值,经过独热编码后用长度为  $N$  且每位为 0 或 1 的码字表示,每个码字中只有一个位置的值可以为 1,表示对应值。

## 5 构建模型

本研究的目的是建立机器学习模型,实现安卓应用隐私评分的预测,该模型的建立以应用中敏感权限的使用意图及相关的元数据为基础,是创建一个回归函数,该函数可以实现应用到评分的映射。构建预测模型是一个迭代的过程,需采用合理的方法选择模型并进行多次对比研究,并以实验结果数据为依据选择最优的回归模型作为预测模型。

### 5.1 建立模型

本文构建多个不同的模型进行回归训练,从最简单的线性模型到复杂的组合模型。线性回归模型实现简单,但容易出现过拟合等问题,对该问题可以使用正则化的方式优化。正则化是把额外的约束或者惩罚项加到已有模型的损失函数上,以防止过拟合并提高泛化能力,正则化又分 L1 正则化和 L2 正则化,相对应的产生了 Lasso 模型和 Ridge 模型。

将简单的基础模型组合可以实现较为复杂的组合模型,组合的方式有很多种,本文中采用随机化以及梯度下降的组合模型。随机化的组合方式结合决策树可以建立随机森林,随机森林由多棵决策树组成,每一棵决策树可以独立工作。梯度下降的方式则是指新模型是在之前建立的模型损失函数的梯度下降方向建立,如果建立的模型能够让损失函数持续下降,说明模型在不停地改进,最好的方式就是让损失函数在其梯度的方向上下降,以此来优化模型。

### 5.2 性能评估

模型预测能力的好坏需要一个评价指标,本文

采用模型对应用的预测评分和应用实际评分之间的均方误差值(mean squared error, MSE)作为模型预测能力的主要评估指标,均方误差值越小,表示模型预测能力越强,模型性能越好。此外在构建模型的时候使用  $k$ -折交叉验证对输入数据自动进行训练,  $k$  值取 1 至 10 之间的正整数,将根据实验结果,取使预测结果最优的  $k$  值。实验中将输入数据拆分为  $k$  组,其中一组保留用于测试,其他  $k-1$  组用于训练,此过程重复  $k$  次,使得每一组数据都有机会作为测试组,然后取  $k$  次训练的均方误差值的平均值作为最终的评估值。  $MSE_{k=i}$  表示第  $k=i$  组数据作为测试组训练模型得到的均方误差值,该值计算过程如式(2)所示,其中  $predict_i$  表示一个应用相关的第  $i$  个向量的预测评分,而  $real_i$  则表示真实评分。  $MSE$  表示模型的均方误差值,计算过程如式(3)所示。

$$MSE_{k=i} = \frac{1}{n} \sum_{t=1}^n (predict_i - real_i)^2 \quad (2)$$

$$MSE = \frac{1}{k} \sum_{i=1}^k MSE_{k=i} \quad (3)$$

### 5.3 隐私评级

一个应用中可能会使用多个敏感权限,同一个敏感权限可能会存在多个使用意图。因此每一个应用相关的<应用,权限,意图>三元组数量不同,这意味着预测模型对每个输入的应用输出一个数目不定的评分向量(appScore),需要确定一种合理的评分策略,根据预测所得应用的评分向量为应用确定一个最终的隐私评分(FinalScore),并根据最终的评分为应用确定一个隐私评分等级(PrivacyRate)。本文采取的方法是,首先对所有的预测值进行排序,然后查看是否存在负值。存在负值说明该应用存在不受用户欢迎的权限使用方式,则将所有的负值求和作为最终的隐私评分。如果不存在负值,表明该应用中敏感权限的使用较为合理,就将所有正值求平均值作为最终的隐私评分(FinalScore)。算法 1 描述了从 appScore 到 PrivacyRate 的计算过程。

#### 算法 1 评分等级策略

输入: 评分向量  $appScore = (s_1, s_2, \dots)$ , 表示一个应用的所有预测分数。

输出: 应用的隐私评分等级 PrivacyRate, [A, B, C, D]



4个等级之一。

```

appScore = (s1, s2, ...) sortedScore = (s1, s2, ...);
If (appScore 中存在负值) {
    FinalScore  $\leftarrow \frac{\text{sum}}{\text{average}}$  sortedScore = (s1, s2, ...);
} Else {
    FinalScore  $\leftarrow \frac{\text{average}}{\text{sum}}$  sortedScore = (s1, s2, ...);
}
FinalScore  $\xrightarrow{\text{algorithm2}}$  PrivacyRate;
Return PrivacyRate;

```

算法2描述从最终的隐私评分(FinalScore)到隐私评分等级(PrivacyRate)的映射。评分等级分为[A, B, C, D] 4个等级,其中A表示应用中权限的使用状况良好,不存在风险;B表示应用中权限的使用状况较好,虽然不存在隐私风险,但可能存在用户并不喜欢的权限使用;C表示应用中权限使用存在隐私风险;D表示应用中存在较为严重的隐私风险。一种情况下应用的评分等级会直接被判定为A。这种应用是指在权限分析的时候发现应用中并没有使用表1中提到的敏感权限,认为该应用不会去访问用户隐私数据,属于安全级别较高的应用,因此直接将其隐私风险等级判定为A。

#### 算法2 评分等级映射

输入:FinalScore,应用最终评分。

输出:PrivacyRate,[A, B, C, D] 4个等级之一。

```

If FinalScore  $\in$  [1,2]
    PrivacyRate  $\rightarrow$  A
Else if FinalScore  $\in$  [0,1):
    PrivacyRate  $\rightarrow$  B
Else if FinalScore  $\in$  (-1,0):
    PrivacyRate  $\rightarrow$  C
Else if FinalScore  $\in$  [-2,-1]:
    PrivacyRate  $\rightarrow$  D

```

## 6 实验和评估

### 6.1 预测准确率

从简单的线性回归模型到较为复杂的组合模型,本文训练8个不同的回归模型,每个模型预测能力使用预测值和实际值得均方误差(MSE)评估,图4展现了各个模型的MSE值,8个模型预测准确率由低到高排序依次为 BayesianRidge (BR)、Lasso (LS)、

Linear (LN)、LassoLars (LL)、Ridge (RD)、Random-Forest (RF)、ensemble of regression tree (ERT)、gradient boost regression tree (GBRT)。由图4可以看到,使用以<权限,意图>特征组为核心构建的模型中,前5个基础模型(BR、LS、LN、LL、RD)的预测结果准确率比较接近,MSE值均在0.4上下浮动,后面3个组合模型(RF、ERT、GBRT)的预测结果准确率比较接近,MSE值均在0.2左右,组合模型的预测能力明显优于基础模型的预测能力。其中渐进梯度回归树(GBRT)取得最好的预测效果,该模型的MSE值为0.193,表明渐进梯度回归树的预测准确率达到80.7%,因此最终选择渐进梯度回归树作为预测模型实现应用隐私评级工具。

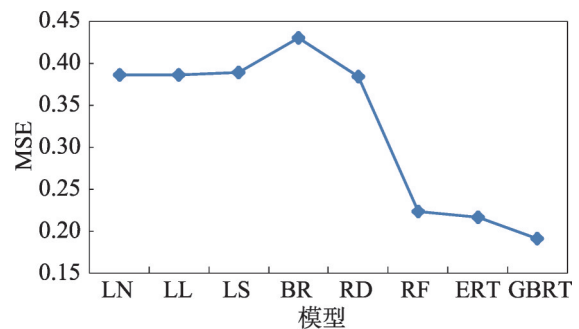


Fig.4 Prediction accuracy of models

图4 各个模型预测准确率

### 6.2 特征集优化

#### 6.2.1 特征重要性

各个模型的评估结果表明渐进梯度回归树的预测效果最好,为了进一步提高模型性能,分析不同的特征在模型中对预测结果的影响大小,去掉一些对预测结果贡献不大的特征,力求在降低特征维度的同时保持预测模型的准确率。首先计算每一个特征在模型中的重要性,如图5所示,permissionName(权限名称)和purpose(权限使用意图)为模型中最重要的两个特征,而dName(是否提供开发者姓名)和dEmail(是否提供开发者邮箱)两个特征的特征重要性值为0,说明dName和dEmail对模型预测能力几乎没有任何贡献,对应用的预测结果基本没有影响。分析训练集数据发现这两个特征为布尔类型的变量,表示应用是否有开发者的姓名和开发者的邮箱

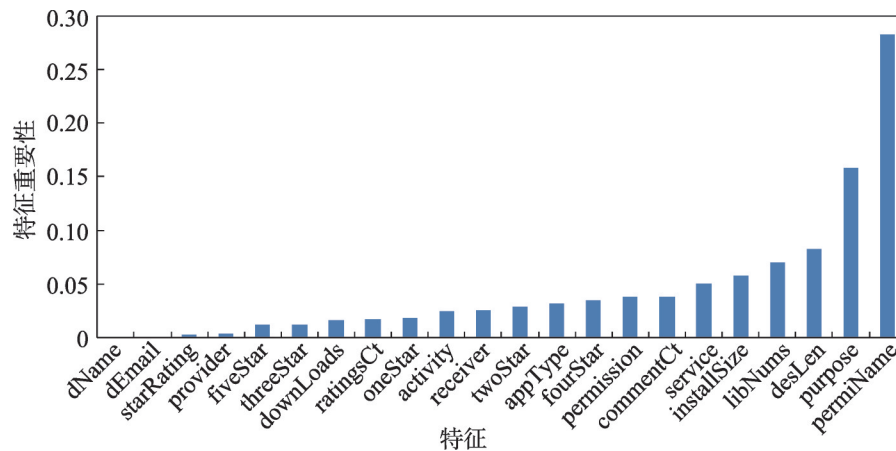


Fig.5 Features importance of GBRT

图5 GBRT 模型特征重要性

地址,与开发者相关的另一个特征是dWeb,表示是否提供了开发者的个人主页介绍,这个特征的重要性高于dName和dEmail。通过对训练集数据的统计分析发现这3个特征取值变化很小,如表4所示,这3个特征中值为1所占比例都高于94%,且dName和dEmail两个特征与隐私评分的皮尔森系数均小于0.1,表明这两个特征跟隐私评分的相关性很弱。因此从特征向量中删除这3个特征,并用剩余特征构建新模型,发现其MSE值为0.209,跟删除这3个特征之前模型的MSE值0.193相比,去掉这3个特征对预测结果准确率影响很小。

Table 4 Distribution of boolean type features

表4 布尔类型特征分布

特征	比例/% (值为1)	比例/% (值为0)	皮尔森 系数	p-value
dName	98	2	0.098	0.033 4
dEmail	96	4	0.076	0.095 6
dWeb	94	6	0.259	<0.000 1

### 6.2.2 特征子集

分析不包括应用开发者信息(dName, dEmail, dWeb)特征子集中特征的重要性,发现前8个最重要的特征中有5个特征来源于apk数据,只有3个特征来源于matadata,由于apk数据的获取更加灵活和方便,因此尝试只使用从apk文件中提取的特征子集为应用构造特征向量,并训练回归模型,分析每个使用特征子集训练模型的预测性能,实验结果如图6所

示,使用所有的特征构建的基础线性模型(蓝色)性能总要优于只使用apk中提取的特征子集构建的预测模型(红色),但是在组合模型中只使用apk中提取的特征子集构建的模型预测结果与使用所有特征构建的模型预测结果相近,ERT模型甚至达到了同样的预测准确率,只使用从apk中提取的特征子集构建的预测模型依然是GBRT模型得到最好的预测效果,且和使用所有特征数据训练的GBRT模型相比,准确率十分相近,因此最终选择只使用从apk中提取的特征子集构建GBRT预测模型,实现应用隐私评分预测工具。

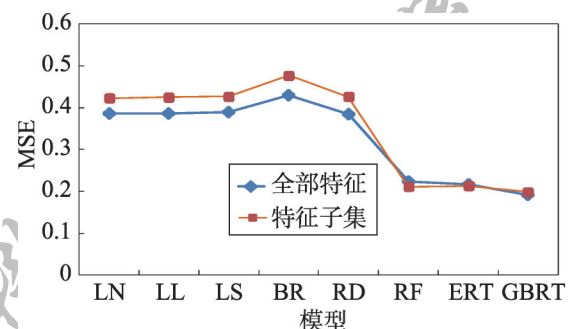


Fig.6 Predict results of feature sets

图6 特征子集的预测结果

### 6.3 大规模预测结果

将隐私评分预测工具应用到11 931个来自谷歌商店的安卓应用,依次反编译每个应用,然后分析每一个应用实际使用的敏感权限、使用意图以及其他的特征数据构造特征向量,通过数据预处理操作之

后输入预测模型确定一个隐私评分等级。预测结果分布如图7所示,4个等级的应用所占的比例不同,其中级别为A的应用所占的比例最高达到42%,评级为A的应用中,有相当一部分是没有使用敏感权限的应用,此类应用没有使用敏感权限即没有权限访问用户的隐私信息数据,因此此类应用安全级别较高,将其评分等级设置为最高级别。评级为B和C的应用

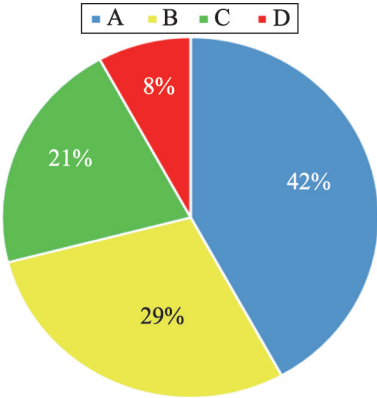


Fig.7 Distribution of privacy rating  
图7 隐私评级分布

所占的比例较为接近,分别为21%和29%,评级最低的D类应用所占比例为8%。

表5列出了4种不同级别的应用中敏感权限使用特点。针对每个评分级别分别列出了检测到的应用,并描述每个应用中实际访问的敏感权限及使用特点,并结合使用意图做出解释和说明。通过实际分析应用,发现以敏感权限的使用意图为应用进行隐私评分是合理的。

6.3.1 敏感权限分析

统计分析每一类评分等级中应用的各个敏感权限出现的次数在每个评级类别中所占比例,各个类别的统计数据如表6所示,表6的最后一列表示每一个权限在所有应用中出现的次数总和,对该列排序可以看出出现次数最高的权限是读取手机状态的READ\_PHONE\_STATE和获取用户精准位置信息ACCESS\_FINE\_LOCATION,ACCESS\_COARSE\_LOCATIONS和READ\_CONTACTS所占比例均低于1%。分析表6中获取用户位置的权限ACCESS\_FINE\_LOCATIONS和ACCESS\_COARSE\_LOCATIONS。

Table 5 Sensitive behavior of different rating applications

表5 不同评级应用的敏感行为

评级	权限使用特点	应用名称	功能描述	敏感权限	使用意图和说明
A	没有使用敏感权限	timetable	日程管理, 时间提醒	无	无
	使用敏感权限,且使用意图合理	flashlight	手电筒	CAMERA	功能实现 使用该权限为了实现应用本身功能,使用意图比较合理
B	(1)敏感权限数量较少 (2)存在少数让用户接受度不高的使用意图	wPainTimer	计时器	READ_PHONE_STATE ACCESS_FINE_LOCATION	功能实现 ACCESS_FINE_LOCATION并不是必须的,没有此权限不影响功能实现,该权限的使用可能会降低用户好感度
C	(1)敏感权限数量较多 (2)存在让用户感觉不安全或者不舒服的使用意图	Halloween Ringtones	敲门铃音	MANAGE_ACCOUNTS READ_PHONE_STATE ACCESS_FINE_LOCATION	功能实现 使用敏感权限数量较多且实现功能并不需要位置权限 ACCESS_FINE_LOCATION
D	(1)敏感权限数量较多 (2)一个敏感权限会有多个使用意图 (3)使用超出应用本身功能的权限使用	Perspective	3D增强现实游戏	RECORD_AUDIO READ_PHONE_STATE ACCESS_FINE_LOCATION CAMERA	GAME_ENGINE TARGETED_ADS 其中有两个敏感权限的使用意图是广告,且访问位置信息的权限并不是必须的



Table 6 Sensitive permissions of different privacy rates  
表6 不同评分等级敏感权限使用情况

权限名称	A/%	B/%	C/%	D/%	总次数
READ_PHONE_STATE	10.86	14.74	27.34	25.19	4 287
ACCESS_FINE_LOCATION	20.86	31.43	26.63	23.28	5 058
GET_ACCOUNTS	11.60	7.30	7.24	8.86	1 585
CAMERA	14.88	8.61	9.72	11.57	2 052
READ_CONTACTS	0.07	0.08	0.20	0.26	33
MANAGE_ACCOUNTS	23.24	23.45	11.97	11.91	3 063
RECORD_AUDIO	11.12	10.17	10.97	14.00	2 218
SEND_SMS	1.45	2.32	1.73	3.46	418
BLUETOOTH	5.06	1.77	4.01	1.47	621
AUTHENTICATE_ACCOUNTS	0.86	0.13	0.19	0	45
ACCESS_COARSE_LOCATIONS	0	0	0	0	0

ACCESS\_FINE\_LOCATIONS 表示获取用户准确位置,通过GPS卫星定位精确度达到10 m以内,该权限在各个评分级别中所占比例均高于20%,而获取用户粗略位置信息的权限所得的统计数量均为0,这些应用在想要获取用户位置的时候一定是想要更精确的位置,因此会偏向于使用ACCESS\_FINE\_LOCATION而非ACCESS\_COARSE\_LOCATION,而每一个评分等级中ACCESS\_FINE\_LOCATION权限所占的比例也相对较高。因此访问用户详细位置信息的权限ACCESS\_FINE\_LOCATION出现次数总和最高,达到5 058次。

### 6.3.2 使用意图分析

表7的统计数据展示了各个使用意图在每一个评分等级类别的应用中所占比例,在A、B、C 3个评分等级中所占比例最高的使用意图是INTERNAL,该标记表示某一敏感权限只在应用程序内部使用,而D级别中所占比例最高的使用意图则是DEVELOPMENT\_AID。横向对比4种评分等级的应用发现在D级别中,TARGETED\_ADS(用于广告),PAYMENT(用于付款),GAME\_ENGINE(游戏)和DEVELOPMENT\_AID(应用分析)这几种使用意图所占比例要高于其他类别中所占比例,这几种使用意图中尤其是TARGETED\_ADS(用于广告)可能会

降低用户体验,让用户感觉不舒服。PAYMENT(用于付款)可能会威胁用户财产安全,因此会降低应用的评分等级。UTILITY(工具)则在A和B中所占比例较高,这表明当敏感权限的使用意图为帮用户解决问题的时候,用户对应用隐私信息使用的接受度较高。

Table 7 Permissions purposes of different privacy rates  
表7 不同评分等级使用意图分布 %

使用意图	A	B	C	D
DEVELOPMENT_AID	28.42	34.18	37.42	65.63
GAME_ENGINE	0.43	1.15	2.02	2.56
INTERNAL	67.29	56.27	46.84	16.07
MAP	0.43	0.55	0.38	0.45
MOBILE_ANALYTICS	1.19	2.64	4.09	3.30
PAYMENT	0.32	0.32	1.03	2.42
SOCIAL_NETWORK	0.00	0.02	0.10	0.62
TARGETED_ADS	0.79	0.62	7.32	8.30
UTILITY	1.12	4.24	0.79	0.65

## 7 结束语

本文提出了一种基于权限使用意图和用于期望的移动应用隐私评级方法,并实现了一个评级工具。通过使用静态分析技术获取每个应用实际使用的敏感权限并分析其使用意图,结合应用其他维度的特征构造特征向量,然后利用机器学习方法构建回归模型。实验结果表明,所构建的预测模型准确率可以达到80%以上,将其应用于11 931个来自谷歌商店的应用,结果表明约8%的应用存在严重的隐私风险。

## References:

- [1] Chin E, Felt A P, Sekar V, et al. Measuring user confidence in smartphone security and privacy[C]//Proceedings of the 8th Symposium on Usable Privacy and Security, Washington, Jul 11-13, 2012. New York: ACM, 2012: 1.
- [2] Enck W, Gilbert P, Chun B G, et al. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones[J]. ACM Transactions on Computer Systems, 2014, 32(2): 5.
- [3] Egelman S, Felt A P, Wagner D A. Choice architecture and

- smartphone privacy: there's a price for that[M]//Böhme R. The Economics of Information Security and Privacy. Berlin, Heidelberg: Springer, 2013: 211-236.
- [4] Choe E K, Jung J, Lee B, et al. Nudging people away from privacy-invasive mobile Apps through visual framing[C]//LNCS 8119: Proceedings of the 14th International Conference on Human-Computer Interaction, Cape Town, Sep 2-6, 2013. Berlin, Heidelberg: Springer, 2013: 74-91.
- [5] Harbach M, Hettig M, Weber S, et al. Using personal examples to improve risk communication for security & privacy decisions[C]//Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems, Toronto, Apr 26-May 1, 2014. New York: ACM, 2014: 2647-2656.
- [6] Jing Yiming, Ahn G J, Zhao Ziming, et al. RiskMon: continuous and automated risk assessment of mobile applications [C]//Proceedings of the 4th ACM Conference on Data and Application Security and Privacy, San Antonio, Mar 3-5, 2014. New York: ACM, 2014: 99-110.
- [7] Zhu Konglin, He Xiaoman, Xiang Bin, et al. How dangerous are your smartphones? App usage recommendation with privacy preserving[J]. Mobile Information Systems, 2016: 6804379.
- [8] Pandita R, Xiao Xusheng, Yang Wei, et al. WHYPER: towards automating risk assessment of mobile applications[C]//Proceedings of the 22nd USENIX Security Symposium, Washington, Aug 14-16, 2013. Berkeley: USENIX Association, 2013: 527-542.
- [9] Qu Zhengyang, Rastogi V, Zhang Xinyi, et al. AutoCog: measuring the description-to-permission fidelity in Android applications[C]//Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, Nov 3-7, 2014. New York: ACM, 2014: 1354-1365.
- [10] Gorla A, Tavecchia I, Gross F, et al. Checking App behavior against App descriptions[C]//Proceedings of the 36th International Conference on Software Engineering, Hyderabad, May 31-Jun 7, 2014. New York: ACM, 2014: 1025-1035.
- [11] Lin Jialiu, Amini S, Hong J I, et al. Expectation and purpose: understanding users' mental models of mobile App privacy through crowdsourcing[C]//Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, Sep 5-8, 2012. New York: ACM, 2012: 501-510.
- [12] Liu Bin, Lin Jialiu, Sadeh N, et al. Reconciling mobile app privacy and usability on smartphones: could user privacy profiles help?[C]//Proceedings of the 23rd International Conference on World Wide Web, Seoul, Apr 7-11, 2014. New York: ACM, 2014: 201-212.
- [13] Felt A P, Ha E, Egelman S, et al. Android permissions: user attention, comprehension, and behavior[C]//Proceedings of the 8th Symposium on Usable Privacy and Security, Washington, Jul 11-13, 2012. New York: ACM, 2012: 3.
- [14] Kelley P G, Consolvo S, Cranor L F, et al. A conundrum of permissions: installing applications on an Android smartphone[C]//LNCS 7398: Proceedings of the 16th International Conference on Financial Cryptography and Data Security, Kralendijk, Mar 2, 2012. Berlin, Heidelberg: Springer, 2012: 68-79.
- [15] Shklovski I, Mainwaring S D, Skúladóttir H H, et al. Leakiness and creepiness in App space: perceptions of privacy and mobile App use[C]//Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems, Toronto, Apr 26-May 1, 2014. New York: ACM, 2014: 2347-2356.
- [16] Mylonas A, Kastania A, Gritzalis D. Delegate the smartphone user? Security awareness in smartphone platforms [J]. Computers & Security, 2013, 34(3): 47-66.
- [17] Liccardi I, Pato J, Weitzner D J, et al. No technical understanding required: helping users make informed choices about access to their personal data[C]//Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, London, Dec 2-5, 2014. Brussels, Belgium: ICST, 2014: 140-150.
- [18] Peng Hao, Gates C, Sarma B P, et al. Using probabilistic generative models for ranking risks of Android Apps[C]//Proceedings of the 2012 ACM Conference on Computer and Communications Security, Raleigh, Oct 16-18, 2012. New York: ACM, 2012: 241-252.
- [19] Ismail Q, Ahmed T, Kapadia A, et al. Crowdsourced exploration of security configurations[C]//Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Apr 18-23, 2015. New York: ACM, 2015: 467-476.
- [20] Zhang Mu, Duan Yue, Feng Qian, et al. Towards automatic generation of security-centric descriptions for Android Apps [C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, Oct 12-16, 2015. New York: ACM, 2015: 518-529.

- [21] Li Yuanchun, Guo Yao, Chen Xiangqun. PERUIM: understanding mobile application privacy with permission-UI mapping[C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Sep 12-16, 2016. New York: ACM, 2016: 682-693.
- [22] Huang Jianjun, Zhang Xiangyu, Tan Lin, et al. AsDroid: detecting stealthy behaviors in Android applications by user interface and program behavior contradiction[C]//Proceedings of the 36th International Conference on Software Engineering, Hyderabad, May 31-Jun 7, 2014. New York: ACM, 2014: 1036-1046.
- [23] Yang Zheming, Yang Min, Zhang Yuan, et al. AppIntent: analyzing sensitive data transmission in Android for privacy leakage detection[C]//Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, Berlin, Nov 4-8, 2013. New York: ACM, 2013: 1043-1054.
- [24] Good N, Dhamija R, Grossklags J, et al. Stopping spyware at the gate: a user study of privacy, notice and spyware[C]//Proceedings of the 1st Symposium on Usable Privacy and Security, Pittsburgh, Jul 6-8, 2005. New York: ACM, 2005: 43-52.
- [25] Yang Wei, Xiao Xusheng, Li Dengfeng, et al. Security analytics for mobile Apps: achievements and challenges[J]. Journal of Cyber Security, 2016, 1(2): 1-14.
- [26] Ma Ziang, Wang Haoyu, Guo Yao, et al. LibRadar: fast and accurate detection of third-party libraries in Android Apps [C]//Proceedings of the 38th International Conference on Software Engineering Companion, Austin, May 14-22, 2016. New York: ACM, 2016: 653-656.

### 附中文参考文献:

- [25] 杨威, 肖旭生, 李邓锋, 等. 移动应用安全解析学: 成果与挑战[J]. 信息安全学报, 2016, 1(2): 1-14.



ZHANG Xianxian was born in 1993. She is an M.S. candidate at Beijing University of Posts and Telecommunications. Her research interest is mobile security.

张贤贤(1993—),女,甘肃天水人,北京邮电大学计算机学院硕士研究生,主要研究领域为移动安全。



WANG Haoyu was born in 1991. He received the Ph.D. degree in computer science from Peking University in 2016. Now he is a lecturer at Beijing University of Posts and Telecommunications. His research interests include mobile security and software engineering, etc.

王浩宇(1991—),男,河南周口人,2016年于北京大学获得博士学位,现为北京邮电大学讲师,主要研究领域为移动安全,软件工程等。



GUO Yao was born in 1976. He received the Ph.D. degree in computer engineering from University of Massachusetts at Amherst in 2007. Now he is an associate professor at Peking University. His research interests include software engineering and system software, etc.

郭耀(1976—),男,山西人,2007年于马萨诸塞大学阿姆斯特分校获得博士学位,现为北京大学副教授,主要研究领域为软件工程,系统软件等。



XU Guo'ai was born in 1972. He received the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications in 2002. Now he is a professor at Beijing University of Posts and Telecommunications. His research interests include software security and mobile security, etc.

徐国爱(1972—),男,江西鄱阳人,2002年于北京邮电大学获得博士学位,现为北京邮电大学教授,主要研究领域为软件安全,移动安全等。