

基于权限与行为的 Android 恶意软件检测研究

张骁敏¹, 刘静^{1,2,3}, 庄俊玺^{1,2,3}, 赖英旭^{1,2,3}

(1. 北京工业大学信息学部, 北京 100124;

2. 北京工业大学可信计算北京市重点实验室, 北京 100124;

3. 北京工业大学信息安全等级保护关键技术国家工程实验室, 北京 100124)

摘要: 针对 Android 平台恶意应用, 从 Android 自身权限机制入手, 提出了一种静态权限特征分析和动态行为分析相结合、将行为映射为权限特征的方法, 并采用关联分析算法挖掘出权限特征之间的关联规则, 把权限特征和行为特征作为朴素贝叶斯分类算法的输入, 建立了一个恶意应用检测模型, 最后通过实验验证了该方法的有效性和准确性。

关键词: Android 系统; 安全机制; 权限特征; 分类器

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00159

Research on Android malware detection based on permission and behavior

ZHANG Xiao-min¹, LUI Jing^{1,2,3}, ZHUANG Jun-xi^{1,2,3}, LAI Ying-xu^{1,2,3}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. Beijing Key Laboratory of Trusted Computing, Beijing University of Technology, Beijing 100124, China;

3. National Engineering Laboratory for Critical Technologies of Information Security Classified Protection,
Beijing University of Technology, Beijing 100124, China)

Abstract: For the Android platform malicious application, a method of mapping behavior to privilege characteristics by combining static privilege feature analysis and dynamic behavior analysis was proposed, and association analysis algorithm was used to dig out the association rules between privilege features. Feature as a naive Bayesian classification algorithm input, a malicious application detection model was established. Finally, the experiment verify the effectiveness and accuracy of the method.

Key words: Android system, security mechanism, permissions feature, classifier

1 引言

近年来, 随着智能手机的普及, Android 操作系统占据了手机操作系统的大部分市场份额。据市场研究公司 Strategy Analytics 公布的最新数据显示, 2016 年第三季度, Android 的全球智能手机市场份额已经达到了 88.0%, 较去年同期的

84.1% 份额略有上升^[1]。然而, 针对 Android 智能手机的恶意程序也以惊人的速度不断增长, 360 互联网安全中心发布的《2016 年第三季度中国互联网安全报告》显示, 2016 年第三季度, 360 安全中心平均每天拦截新增 Android 恶意应用程序样本近 3.8×10^4 个, 共拦截将近 3.49×10^6 个。Android 平台新增恶意程序主要是资费消耗,

收稿日期: 2016-12-02; 修回日期: 2017-03-02。通信作者: 赖英旭, laiyingxu@bjut.edu.cn

基金项目: 北京市自然科学基金资助项目 (No.4162006)

Foundation Item: The Natural Science Foundation of Beijing (No.4162006)

占比高达 73.5%，其次为恶意扣费（18.0%）、隐私窃取（3.8%）、流氓行为（2.8%）和远程控制（1.5%）^[2]，这些恶意程序给手机用户带来了巨大的安全隐患。

Android 平台恶意程序的迅猛增长，对恶意软件的检测研究十分重要。然而，现有的检测方法主要以基于权限特征的静态分析为主^[3]，这种方法只考虑单个权限特征，不足以反映 Android 应用程序的特性，且也缺乏针对 Android 应用程序的权限特征和行为的综合考虑。因此，本文提出了一种基于权限与行为的 Android 恶意软件检测方案。

2 相关工作

恶意软件检测存在 2 种方法^[4]：一是基于特征代码的检测方法；二是基于行为的检测方法。

基于行为的检测方法又分为动态分析方法和静态分析方法^[4]。张锐等^[5]采用卡方检验计算各权限特征对分类结果的影响，去除冗余权限特征，从权限相关性角度快速检测恶意软件，但没有考虑组合权限的威胁。重庆邮电大学黄梅根等^[6]以窃取 Android 系统隐私数据为切入点，将 Android 系统隐私数据划分为 4 类（通话、短信、地理位置信息、其他个人隐私信息），参考 Kirin 安全规则提出了一系列恶意权限组合，该方法的不足是某些权限规则对于某些正常软件和恶意软件都有可能被使用到，提出的组合权限没有科学依据。

Enck 等^[7]设计实现了 Taint Droid 系统，该系统可以标记内存数据的流向，修改 Dalvik 虚拟机使在 Android 系统中应用传统的信息流追踪技术，在软件运行时对系统中敏感信息的传播进行监控，以达到安全防范的目的。然而，Taint Droid 采用了传统的动态污点传播技术，使处于运行时的应用程序性能开销明显增加，这在实时性要求非常高的移动平台上使用缺乏实用性。Zhang 等^[8]实现了软件动态分析平台 VetDroid，该平台可以分析出应用的权限使用行为，包括访问系统资源时如何申请使用的权限，以及进一步使用这些资源。但以上工作仅提供了一种辅助分析的平台工具，其本身无法分析恶意行为。

综上，基于特征代码的检测方法快速、准确

率高、误报率低，但是无法检测未知的恶意程序。基于行为的检测方法可以检测未知的恶意程序变种，缺点是误报率较高。因此本文从 Android 自身权限机制入手，提出了一种静态权限特征提取和动态行为分析相结合，将行为映射为权限特征的方法，采用关联分析算法挖掘出权限特征之间的关联规则，为恶意应用的检测奠定了坚实的基础。

3 恶意应用检测

恶意应用检测分为两大部分：建立恶意应用检测模型、测试恶意应用检测模型。

3.1 建立恶意应用检测模型

恶意 APK 和正常 APK 训练集是权限特征提取模块的分析依据，构建过程如图 1 所示，主要包括权限特征提取、权限特征关联分析和特征集合分类这 3 个模块。

3.1.1 权限特征提取

权限特征提取分为静态分析和动态分析。静态分析是对已知应用程序反编译的过程，本文静态分析是在开源项目 androwarn 的基础上进行改进，自动读取特定目录下应用程序申请的权限，该工具主要对反编译结果目录中的 AndroidManifest.xml 文件进行自动分析，提取训练集中应用程序申请的权限信息，如读取短信、打电话等权限信息。

静态权限特征提取的特点是应用程序无需执行，就可以提取应用程序申请的权限，如果使用这种方法对获取的权限信息进行恶意应用检测会导致误判率很高。例如，某应用程序申请了危险权限但其代码中并没有用到该权限，采用该方法很可能会将正常应用误判为恶意应用。因此，本文在静态分析的基础上，进一步结合动态行为分析提取权限特征。

本文的动态行为分析采用建立在 TaintDroid 基础之上的动态分析工具 DroidBox，其通过污染分析和 API 监控等手段分析函数的调用过程，提取训练集中每个应用程序的行为。DroidBox 的分析结果包括输入/输出网络数据、文件读取和写入操作、启动服务和通过 DexClassLoader 加载的类、通过网络/文件和 SMS 信息泄露、执行加密操作使用的 Android API、发送短信和拨打电话等行为。

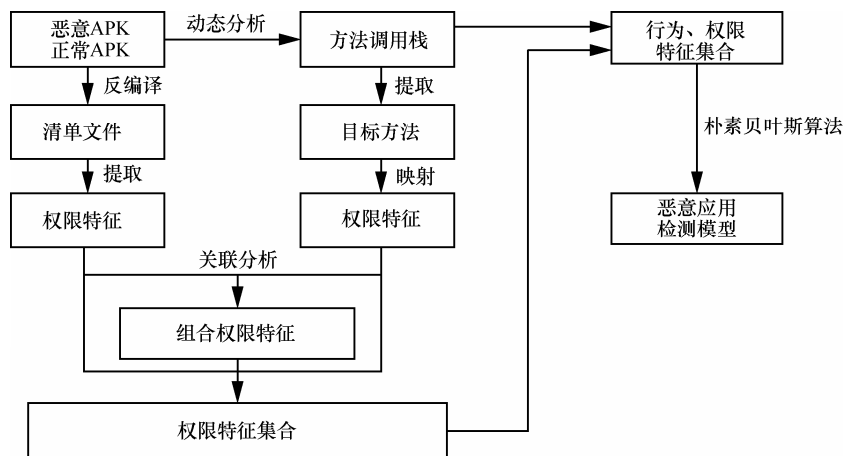


图1 恶意应用检测模型框架

通过 DroidBox 分析 sample1.apk 样本，得到的方法调用过程如图2所示。

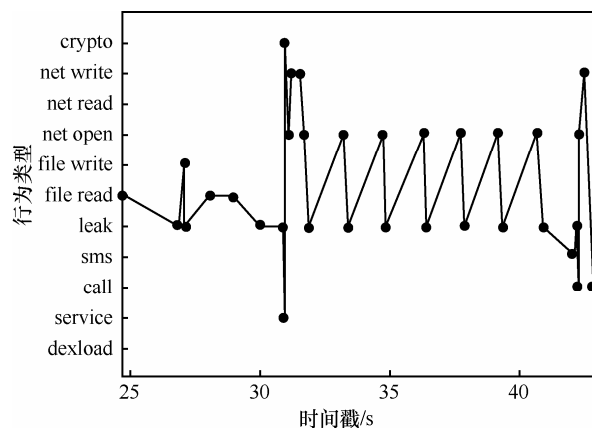


图2 sample1.apk 方法调用过程

每个应用程序动态行为分析的结果都是一个方法调用栈，包含方法的个数和调用顺序，该结果有2个用途：一是用于权限特征提取；二是和权限特征集合一并作为恶意应用检测模型的分类特征。

Android 应用程序通过声明权限获取沙盒未提供的行为，即一个行为的产生需要一个或多个权限的配合，所以应用程序的行为可以反映应用程序申请的权限。结合 Android 运行时权限，提出如下算法。

算法1 行为与权限特征映射

输入：方法调用栈，25种危险权限（包括在Android官方提供^[9]24种运行时权限基础上加上INTERNET权限）

输出：权限特征集合

- 1) $stack \leftarrow \text{Method Call Stack}$;
- 2) $pMSet \leftarrow 25 \text{ Dangerous Permissions}$;
- 3) $tm \leftarrow \text{Target Method}$
- 4) $pSet \leftarrow \text{set}()$;
- 5) $pSet \leftarrow \text{findPermission}(tm)$;
- 6) if $pSet$ is empty then
- 7) $N \leftarrow \text{size}(stack) - 1$;
- 8) $r \leftarrow \text{stringAnalysis}(stack[1 \cdots N])$;
- 9) $pSet \leftarrow pSet \cap r \cap pMSet$;
- 10) return $pSet$

算法1实现了行为与权限特征的转换，形成的行为与权限特征映射如表1所示。

静态分析提取权限特征只需要一步，即通过反编译提取权限特征信息，动态行为分析提取权限特征分为2步：一是根据函数调用提取训练集中每个应用程序的行为；二是利用表1将行为映射为权限特征。

3.1.2 权限特征关联分析

Android 应用程序申请的每个权限可以有效地反映出应用程序的行为模式，但对一些敏感权限，恶意应用和正常应用可能都会申请，不足以反映Android应用程序的特性，而且一个恶意行为的产生通常需要多个权限的配合^[10]。所以，本文通过静态分析和动态分析获取权限特征信息的同时，采用权限特征关联分析挖掘权限特征之间的关联规则，提高恶意应用检测的检测率和准确率。

关联分析是从海量数据集中，提取出有用的数据关联、频繁模式，为进一步的数据决策提供数据支持^[11]。本文采用挖掘布尔关联规则频繁项

表1 行为与权限特征映射

行为	权限特征
crypto	GET_ACCOUNTS
net_write	WRITE_CONTACTS
net_read	READ_PHONE_STATE、READ_CONTACTS
net_open	INTERNET、ACCESS_COARSE_LOCATION
file_write	WRITE_EXTERNAL_STORAGE、WRITE_CALENDAR
file_read	READ_EXTERNAL_STORAGE、READ_CALENDAR
leak	ACCESS_FINE_LOCATION、CAMERA
sms	READ_SMS、RECEIVE_SMS、SEND_SMS、RECEIVE_WAP_PUSH、RECEIVE_MMS
call	CALL_PHONE、READ_CALL_LOG、WRITE_CALL_LOG、ADD_VOICEMAIL、PROCESS_OUTGOING_CALLS
service	RECORD_AUDIO、USE_SIP
dexload	BODY_SENSORS

集的 Apriori 先验算法,挖掘应用程序训练集申请的权限之间的关联规则,该过程分为3个步骤。

1) 将权限特征提取模块提取到的每个应用程序申请的所有权限构建成权限特征集合。

2) 使用 python 语言编写代码对第1)步中的权限特征集合进行处理,统一为逗号分隔值(CSV, comma separated values)格式,以便后续进行频繁模式挖掘。

3) 使用 Apriori 算法挖掘权限特征之间的相互依赖性,即权限关联规则。

3.1.3 特征集合分类

特征集合包括权限特征和行为特征,特征集合分类采用朴素贝叶斯分类算法^[12],记 $X = \{x_1, x_2, \dots, x_n\}$ 表示用于分类的权限特征属性, $Y = \{y_1, y_2\}$ 表示分类结果为恶意应用和正常应用。其基本思想是通过训练样本对每个分类计算先验概率 $P(Y)$, 根据先验概率和样本数据信息获得确定事件的后验概率,最后把事件归于后验概率最大的类别。将权限特征集合和行为特征集合进行信息聚集,作为朴素贝叶斯分类算法的输入数据,建立一个恶意应用检测模型。

3.2 测试恶意应用检测模型

测试恶意应用检测模型的过程如图3所示。

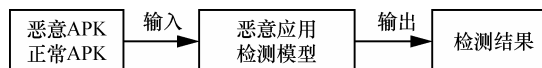


图3 测试恶意应用检测模型的过程

恶意 APK 和正常 APK 测试集是恶意应用

检测模型的测试依据,将测试集输入到恶意应用检测模型,判定目标应用程序是否为恶意应用程序。

4 实验与分析

4.1 实验说明

实验在内存为4GB,处理器为Intel(R) Core 2.20 GHz的机器和系统为Android 5.0.1的Nexus手机上完成,分别用Python、Java语言实现了如图1所示的恶意应用检测模型,为了衡量实验效果,定义如下衡量指标^[13]。

1) 真阳性 (TP, true positive): 检测正确的恶意样本数量。

2) 假阴性 (FN, false negative): 检测错误的恶意样本数量。

3) 真阳性率 (TPR, true positive rate):

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}, \text{表示命中率(检测率)}。$$

4) 假阳性 (FP, false positive): 检测错误的良性样本数量。

5) 真阴性 (TN, true negative): 检测正确的良性样本数量。

6) 假阳性率 (FPR, false positive rate):

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}, \text{表示错误命中率(误报率)}。$$

7) 准确率 (ACC, accuracy): $ACC = \frac{TP + TN}{N + P}$ 。

实验数据及来源如表 2 所示。

表 2 实验数据及来源

应用类型	训练集数目	测试集数目	来源
恶意应用	2 770	2 771	VirusShare ^[14]
正常应用	2 718	2 719	Google Play

4.2 实验结果

将测试集各分为 5 组作为实验数据，所有实验使用相同的训练集和测试集。

整个恶意应用检测模型的建立过程有 2 个重要的环节：一是提取权限特征时的关联分析；二是将权限特征和行为特征作为 2 个分类特征。

实验 1 权限特征提取未进行关联分析，分类特征仅为权限特征，实验结果如表 3 所示。

实验 2 权限特征提取进行关联分析，分类

特征仅为权限特征，实验结果如表 4 所示。

表 3 和表 4 的实验结果表明，仅通过单个的权限无法反映 Android 应用程序的特性，一个恶意行为的产生通常需要多个权限的配合，脱离组合的权限并没有太大的威胁，实验 2 在权限特征提取过程中进一步挖掘了权限特征之间的关联规则，使权限特征集合对恶意检测更有效，所以，权限特征关联分析明显提高了恶意应用检测模型的检测率和准确率。

但是，相对于表 3，表 4 中数据显示 *TPR* 有了明显的提高，而第 4 组数据则提高很少，通过观察恶意应用训练集权限组合和正常应用训练集权限组合，发现 INTERNET 是一个通用权限，在恶意应用和正常应用当中的使用率都很高，如表 5 和表 6 所示。

表 3 实验 1 的实验结果

衡量指标	组号					平均值
	1	2	3	4	5	
<i>TPR</i>	79.42%	84.66%	78.52%	81.23%	81.44%	81.05%
<i>FPR</i>	11.05%	10.29%	14.15%	12.87%	11.40%	11.95%
<i>ACC</i>	84.14%	87.16%	82.15%	84.15%	84.99%	84.52%

表 4 实验 2 的实验结果

衡量指标	组号					平均值
	1	2	3	4	5	
<i>TPR</i>	86.64%	85.74%	87.55%	82.12%	83.94%	85.10%
<i>FPR</i>	9.21%	8.27%	11.03%	11.95%	7.17%	9.53%
<i>ACC</i>	88.70%	88.71%	88.25%	85.06%	88.26%	87.80%

表 5 恶意应用程序使用较多的权限组合后的支持度和置信度

权限组合	支持度	置信度
READ_PHONE_STATE+INTERNET+SEND_SMS	0.21	0.86
INTERNET+WRITE_EXTERNAL_STORAGE+WRITE_CALENDAR	0.29	0.85
INTERNET+READ_EXTERNAL_STORAGE	0.25	0.85
ACCESS_FINE_LOCATION+INTERNET+GET_ACCOUNTS	0.40	0.82
INTERNET+READ_PHONE_STATE+RECORD_AUDIO	0.33	0.81

表 6 正常应用程序使用较多的权限组合后的支持度和置信度

权限组合	支持度	置信度
INTERNET+FLASHLIGHT+CHANGE_CONFIGURATION	0.23	0.88
RECEIVE_SMS+INTERNET+VIBRATE	0.30	0.86
INTERNET+SET_WALLPAPER+ACCESS_COARSE_LOCATION	0.21	0.84
INTERNET+ACCESS_WIFI_STATE+RESTART_PACKAGES	0.34	0.83
INTERNET+GET_TASKS+WRITE_SETTINGS	0.28	0.80

针对以上问题,对算法1进行了改进,将INTERNET权限从危险权限中去除,生成新的行为与权限特征映射表,进行实验。

实验3 权限特征提取进行关联分析,算法1去除INTERNET权限,分类特征仅为权限特征,实验结果如表7所示。

从表7可以看出,实验3较实验2检测率和准确率都略有提高,但是,第1组和第3组中TPR和ACC明显低于平均值,究其原因,是由于第1组和第3组的测试样本集中存在一些非典型恶意应用,其特点是申请了很少的权限,但威胁程度极大,仅依靠权限不能够准确检测出这种恶意应用,这是基于权限恶意应用检测的不足之处。针对非典型恶意应用进行如下实验。

实验4 在实验3的基础上加入了行为特征,即如图1所示的实验方案,实验结果如表8所示。

实验结果表明,在静态权限特征分析和动态行为分析相结合的基础上,进行权限特征关联分析的方案的检测率和准确率高于没有进行权限特征关联分析的方案,该方案弥补Kirin方案不足,同时还挖掘出了权限之间的关联规则。通过观察恶意应用和正常应用所使用的权限组合,发现INTERNET是通用权限,进而改进了算法1,最后,针对申请很少权限但威胁极大的非典型应用程序,将权限特征和行为特征作为分类特征,使最终输出的恶意应用检测模型具有较高的检测率和准确率,对测试集的检测能力有了明显提升。

5 结束语

基于静态权限特征分析的不足,本文提出了静态权限特征分析和动态行为分析相结合的方法,并将行为映射为权限特征,进一步优化权限特征,然后利用权限特征关联分析挖掘出权限之间的关联规则,解决了对申请敏感权限的正常应用程序的误判问题。并且针对申请较少权限的非典型应用程序,将权限特征和行为特征作为分类特征,提高了检测模型的检测能力。因此,该恶意应用检测模型对恶意应用检测有较高的准确率,但还存在一定的不足,如果再结合其他的特征,扩大样本数量,优化分类算法,可以进一步提高准确率。

参考文献:

- [1] Strategy analytics: Android captures record 88 percent share of global smartphone shipments in Q3 2016[EB/OL]. <https://www.strategyanalytics.com/strategy-analytics/news/strategy-analytics-press-releases/strategy-analytics-press-release/2016/11/02/strategy-analytics-android-captures-record-88-percent-share-of-global-smartphone-shipments-in-q3-2016?slid=90814&spg=3#.WG2VCIV951s>.
- [2] 2016年第三季度中国互联网安全报告[EB/OL]. <http://www.docin.com/p-1787390122.html&formDaily=1>.
The Chinese Internet security report in the third quarter of 2016[EB/OL]. <http://www.docin.com/p-1787390122.html&formDaily=1>.
- [3] 张怡婷, 张扬, 张涛, 等. 基于朴素贝叶斯的Android软件恶意行为智能识别[J]. 东南大学学报, 2015, 45(2): 224-230.
ZHANG Y T, ZHANG Y, ZHANG T, et al. Based on naive bayesian Android software malicious behavior intelligent identification[J].

表7

实验3的实验结果

衡量指标	组号					平均值
	1	2	3	4	5	
TPR	87.73%	87.36%	88.23%	85.74%	86.13%	87.04%
FPR	8.66%	7.72%	10.48%	10.85%	7.00%	8.94%
ACC	89.52%	89.80%	88.89%	87.43%	89.54%	89.03%

表8

实验4的实验结果

衡量指标	组号					平均值
	1	2	3	4	5	
TPR	90.25%	90.61%	91.16%	89.35%	89.78%	90.22%
FPR	6.62%	6.43%	6.99%	7.35%	6.45%	6.78%
ACC	91.70%	92.17%	92.08%	90.98%	91.63%	91.71%

- Journal of Southeast University, 2015, 45(2):224-230.
- [4] 陈宏伟. 基于关联分析的 Android 权限滥用攻击检测系统研究[D]. 合肥: 中国科学技术大学, 2016.
- CHEN H W. Research on Android rights abuse attack detection system based on correlation analysis[D]. Hefei: China University of Science and Technology, 2016.
- [5] 张锐, 杨吉云. 基于权限相关性的 Android 恶意软件检测[J]. 计算机应用, 2014, 34(5):1322-1325.
- ZHANG R, YANG J Y. Android malware detection based on permission relevance[J]. Journal of Computer Applications, 2014, 34(5):1322-1325.
- [6] 黄梅根, 曾云科. 基于权限组合的 Android 窃取隐私恶意应用检测方法[J]. 计算机应用与软件, 2016, 33(9):320-333.
- HUANG M G, ZENG Y K. Based on the combination of authority Android steal privacy application detection method[J]. Computer Applications and Software, 2016, 33(9):320-333.
- [7] ENCK W, GILBERT P, CHUN B G, et al. TaintDroid: an information flow tracking system for real-time privacy monitoring on smartphones[C]//Usenix Symposium on Operating Systems Design and Implementation(OSDI 2010). 2010:393-407.
- [8] ZHANG Y, YANG M, XU B, et al. Vetting undesirable behaviors in Android apps with permission use analysis[C]//The 20th ACM Conference on Computer and Communications Security. 2013: 611-622.
- [9] The developer's guide[EB/OL]. <https://developer.android.google.cn/guide/index.html>.
- [10] 杨欢, 张玉清, 胡予濮, 等. 基于权限频繁模式挖掘算法的 Android 恶意应用检测方法[J]. 通信学报, 2013, 34(Z1):106-115.
- YANG H, ZHANG Y Q, HU Y P, et al. Android malicious application detection method based on privilege frequent pattern mining algorithm[J]. Journal on Communications, 2013, 34(Z1):106-115.
- [11] HUANG J J, ZHANG X Y, TAN L. Detecting sensitive data disclosure via bi-directional text correlation analysis[C]//The 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering. 2016:169-180.
- [12] AKSHAY N, PRATEEK S. The curse of 140 characters: evaluating the efficacy of SMS spam detection on Android[C]//The 3rd ACM workshop on Security and Privacy in Smartphones & Mobile Devices. 2013:33-42.
- [13] 蔡泽廷, 姜梅. 基于权限的朴素贝叶斯 Android 恶意软件检测研究[J]. 电脑知识与技术, 2013, 9(14):3288-3291.
- CAI Z T, JIANG M. Research on naive bayesian Android malware detection based on permission[J]. Computer Knowledge and Technology, 2013, 9(14):3288-3291.
- [14] VirusShare.com. Because sharing is caring[EB/OL]. <https://virusshare.com/>.

作者简介:



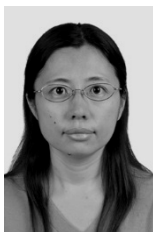
张骁敏(1990-), 男, 山西洪洞人, 北京工业大学硕士生, 主要研究方向为 Android 软件安全。



刘静(1978-), 女, 北京人, 硕士, 北京工业大学讲师, 主要研究方向为网络安全、可信计算。



庄俊玺(1981-), 女, 河南新乡人, 北京工业大学讲师, 主要研究方向为网络安全、可信计算等。



赖英旭(1973-), 女, 辽宁抚顺人, 博士, 北京工业大学教授, 主要研究方向为网络接入控制、病毒防御技术、网络安全工程、可信计算理论及其应用。