APP 隐私泄露风险评估与保护方案

王新宇[1,2]，牛犇[1]，李凤华[1,2]，贺坤[1,2]
（1. 中国科学院信息工程研究所，北京 100093；2. 中国科学院大学网络空间安全学院，北京 100049）

摘 要：针对 APP 中第三方服务提供商非法采集用户隐私信息的问题，提出了一种 APP 隐私信息泄露风险评估方案 PRAS。该方案通过统计第三方服务提供商从不同 APP 获取的权限，并考虑权限组合对隐私泄露风险带来的非线性影响，构建模型来评估隐私泄露风险。基于风险评估结果，在服务质量与隐私保护之间进行均衡分析，最终给出系统整体的权限管理方案，在保证服务质量的同时，降低隐私信息泄露风险。实验结果表明，PRAS 将 APP 整体的隐私泄露风险平均降低了 18.5%。

关键词：安卓；隐私保护；风险评估；权限管理
中图分类号：TN 929
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2019085

Risk assessing and privacy-preserving scheme for privacy leakage in APP

WANG Xinyu[1,2], NIU Ben[1], LI Fenghua[1,2], HE Kun[1,2]
1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The APP in smartphone contain various third-party services. However, the service providers illegally read the user's private information. To address this problem, a privacy risk assessing scheme called PRAS was proposed. Firstly, a model was built to assess the risk of privacy leakage, by counting all the permissions acquired by each service providers and considering the non-linear impact of the permissions combination on privacy leakage. Then, by analyzing the balance between service quality and privacy-preserving, an optimal model was used to minimized the risk of private information leakage, and a permission management method was given to protect the privacy information among APP. The experiment results show that PRAS reduces the risk of privacy leakage by an average of 18.5%.

Key words: Android, privacy-preserving, risk assessment, permission management

1 引言

随着信息化服务的快速普及和移动互联网相关技术的发展，智能设备的使用已经渗透人们的日常生活。用户在享受智能设备带来便捷服务的同时，也承受着日益严峻的隐私信息泄露的风险[1]。据调查，85.2%的受访者曾因使用 APP 导致个人隐私信息泄露[2]；Android 平台申请使用“读取通讯录”

收稿日期：2018–11–06；**修回日期**：2019–02–28
通信作者：李凤华，lifenghua@iie.ac.cn
基金项目：国家重点研发计划基金资助项目（No.2017YFB0802203）；国家自然科学基金资助项目（No.U1401251, No.61672515, No. 61872441）；中国科学院青年创新促进会人才基金资助项目；工业和信息化部 2018 工业互联网创新发展工程基金资助项目：工业互联网标识解析数据管理技术标准制定与试验验证
Foundation Items: The National Key Research and Development Program of China (No. 2017YFB0802203), The National Natural Science Foundation of China (No. U1401251, No.61672515, No.61872441), Youth Innovation Promotion Association CAS, Industrial Internet Innovation and Development Project of China: Technical Standard Formulation and Verification of Identifier Data Management for Identification and Resolution System

权限的 APP 中，89.69%的 APP 存在滥用该权限的情况[3]。在上述调查中，由于智能设备中 APP 过度申请权限，导致用户隐私信息泄露事件的频发。因此，亟需针对权限配置的管理方案，保护用户的个人隐私数据不被泄露。

APP 中包含的第三方服务提供商可以在用户没有察觉的情况下获取隐私信息。目前，智能设备的操作系统（例如 Android、iOS 等）使用权限管理机制保护设备中的隐私数据。以 APP 为单位控制访问权限，只有获得相应权限的 APP 才能读取用户隐私数据。但现实中，由于 APP 集成了来自第三方服务（例如广告、支付等服务），使权限管理机制的保护效果降低。第三方服务以库文件的形式被包入宿主 APP，这些库文件为服务提供商（SP, service provider）发布的二进制文件，常见的文件后缀名有“.jar”“.so”“.tdb”等，且同一服务可以被打包到多个 APP 中，在权限管理机制下，其拥有与宿主 APP 相同的权限。若从每个 APP 获得部分敏感权限，则 SP 能获得用户全部与隐私信息相关的敏感权限；进而绘制完整的用户画像，并以此牟利。因为操作系统没有向用户提示是宿主 APP 还是第三方服务在申请使用权限，所以用户对发生的隐私信息泄露事件一无所知。

为解决上述问题，本文提出了一种系统整体的隐私信息泄露风险评估及保护方案 PRAS (privacy risk assessing scheme)。该方案从隐私信息传播的广度和深度构建模型来评估隐私泄露风险，最终给出权限管理方案。其中，传播广度是指隐私信息传播给了那些服务提供商，一个服务提供商可能从多个 APP 中获取隐私信息；传播深度是指用户隐私信息泄露的程度，Android 设备中通过权限机制管理隐私信息，本文以服务提供商获得的权限数量和其敏感度来计算隐私信息泄露的深度。PRAS 中量化了权限的敏感度，在量化过程中考虑权限组合对隐私泄露风险带来的非线性影响。由于搭载 Android 系统的智能设备数量巨大，且研究成果具有普遍性，本文以 Android 系统为实现平台设计方案，验证了所提方案的有效性。本文的主要贡献如下。

1) 提出了一种新的隐私信息泄露风险评估模型，该模型重点考虑了第三方服务提供商在用户不知情的情况下获取隐私信息的情况。

2) 基于恶意侵犯隐私的 APP 集合与正常 APP 集合之间申请权限的差异，量化了权限的敏感度，加入了权限组合对隐私泄露带来的非线性影响。

3) 均衡服务质量与隐私保护，建立了个性化最优模型，提供了整体的权限管理方案，在保证服务质量的同时，尽量降低隐私信息泄露风险。

2 相关工作

2.1 第三方服务库识别

早期研究通过白名单机制识别第三方服务库。Grace 等[4]利用白名单识别出 100 个广告库，白名单机制虽然简单易行，但抗干扰能力差。Chen 等[5]在检测 APP 克隆时列出了 73 个第三方服务库的模块名，假如模块名被修改后，这种方法就失效了。

由于第三方服务库会被重用，可以通过提取代码特征或比对代码的相似性来识别。在提取代码特征的方案[6-11]中，Narayanan 等[6]和 Liu 等[7]对 APP 进行分析后，提取代码的特征训练识别器，特征如所使用的 Android 组件、权限、API 等，利用机器学习的方法识别广告库。广告库的特征明显，易于区分。而其他类型的第三方服务库（例如社交网络、游戏、地图等）则难于识别[10]，利用聚类的方法，可以将包含相同第三方服务库的 APP 聚集到相同的簇中，对聚类结果进行分析就能识别宿主应用中的第三方服务库。Crussell 等[8]和 Wang 等[9]通过提取 APP 调用其他类库的代码特征，以大量的 APP 为基础，使用聚类算法识别第三方服务库。Ma 等[10]和 Li 等[11]在前人的基础上[9]，提出了多层级聚类的概念。例如“*a*/*b*/*c*”“*a*/*b*/*d*”和“*a*/*b*/*e*”这 3 个模块，根据模块之间包含关系，将这 3 个模块聚集在一簇中。在基于代码相似性的方案中，Backes 等[12]提出基于相似度匹配的第三库检测方法，该方法将 APP 反编译后的代码按照方法、类和包的层级组织成 Merkle 树，树的叶子节点的值是方法的签名。在大规模 APP 的对比中，匹配到散列值相同或者最相似的模块即认为是同一个第三方服务库。

以上方案能准确识别出 APP 中包含的各种类型的第三方服务库。

2.2 权限管理方案

调查[13]指出，只有 17%的用户会留意权限的警告。目前，已有大量的文献研究如何帮助用户管理权限，这些文献[14~19]可以分为基于上下文环境的权限管理和基于众包的权限管理这两类。

1) 基于上下文环境的权限管理

Fawaz 等[14]提出了 LP-Guardian 方案帮助用户

决策授予/撤销位置权限。Tsai 等[15]使用机器学习的方法，根据用户使用 APP 的反馈和运行的上下文环境对 APP 的权限进行管理。

2) 基于众包的权限管理

Agarwal 等[16]收集用户的权限设置，根据多数人对权限授予/撤销的选择给出推荐结果。Liu 等[17]通过分析用户对权限设置的记录，使用支持同量机算法将用户分为若干类，然后根据分类推荐权限设置。Liu 等[18]通过向用户提问的方式，找出对隐私期望相近的用户，采用协同过滤算法推荐权限配置。Rashidi 等[19]从收集的用户权限设置中寻找“专家用户”，并根据“专家用户”的设置推荐权限配置结果。

然而，上述权限管理方案都没有考虑第三方服务商造成的隐私泄露。

3　预备知识或定义

3.1　Android 系统中权限管理

为了保护用户的隐私数据，Android 系统采用基于权限管理的隐私保护机制，在该机制下，APP 需在配置文件中声明要使用的权限，经用户同意后，才能访问权限所对应的硬件或读写数据。在 Android 6.0（Marshmallow）版本之前，用户安装 APP 软件时，弹出窗口提示 APP 运行时需要用到的权限。如果用户同意，则授予权限继续安装，否则停止安装 APP 的流程。从 Android 6.0 版本开始，Android 系统将权限分为非敏感权限（normal permission）和运行时权限（runtime permission）。非敏感权限不直接涉及用户隐私，比如网络、蓝牙、NFC 等，APP 安装时默认授权，且用户无法取消授权。运行时权限直接涉及用户隐私，比如通讯录、短信、位置等，APP 首次使用时在界面上弹出窗口让用户授权，授权后可以在系统设置界面动态管理该权限。此外，基于 Android 深度定制的系统（如华为公司的 Emotion UI、小米公司的 MIUI 等操作系统）中除了对运行时权限设置外，在其系统设置界面中可以对一般权限进行管理。虽然这种模式能让用户更好地了解和控制权限，但是其将隐私信息的授予/撤销的决定权交给了用户，用户能否正确地理解权限的具体含义是这种模式成功的关键。

3.2　关联规则算法

关联规则学习（association rule learning）是一种在大型数据库中发现变量之间有趣性关系的方法。假设 $I=\{I_1,I_2,\cdots,I_m\}$ 是项的集合，给定一个数据库 $D=\{t_1,t_2,\cdots,t_n\}$，其中每个事物 $t_i(1\leqslant i\leqslant n)$ 是 I 的非空子集，即 $t\subseteq I$。关联规则是形如 $a\Rightarrow b$ 的蕴含式，其中 a、$b\subseteq I$ 且 $a\cap b=\varnothing$，a 和 b 分别称为关联规则的先导（antecedent）和后继（consequent）。关联规则 $a\Rightarrow b$ 在 D 中的支持度（support）是 D 中事物包含 a 和 b 的百分比，即概率 $P(a\cup b)$，其中，$a\cup b$ 表示 t_i 中同时包含 a 和 b。置信度（confidence）是包含 a 的事物且包含 b 的百分比，即条件概率 $P(b\mid a)$，$\mathrm{conf}(a\Rightarrow b)=\dfrac{\mathrm{supp}(a\cup b)}{\mathrm{supp}(a)}$。如果同时满足最小值支持度阈值和最小置信度阈值，则认为关联规则是有趣的，这些阈值由用户或领域专家来设定。为了表示 a 和 b 之间的相关性，使用提升度（lift）来表示，lift 定义为

$$\mathrm{lift}(a\Rightarrow b)=\frac{\mathrm{supp}(a\cup b)}{\mathrm{supp}(a)\mathrm{supp}(b)} \tag{1}$$

当 a 和 b 互相独立时，$P(a\cup b)=P(a)P(b)$，$\mathrm{lift}=1$，说明两者没有任何关联。

Eclat 算法[20]是一种基于深度优先的关联规则挖掘算法，将数据库 D 中数据由 $\{t_i:I_1,I_2,\cdots\}$ 倒排为 $\{I_j:t_1,t_2,\cdots\}$。在此基础上，通过 k–频繁项集之间的交集，迭代计算 $(k+1)$–频繁项集。

3.3　权限撤销概率参数估计

将用户设置权限的事件看作随机变量 χ 且相互独立，随机变量 χ 服从伯努利分布 $F(S_n,k)=B(k;n,p)$。撤销权限 P_i 为事件 1，授予权限 P_i 为事件 0。若用户以概率 p 撤销权限，则 $\Pr(X=1)=p$，$\Pr(X=0)=1-p$，其中，概率 p 为未知参数，需要通过收集样本估算得出。收集样本容量为 n 的用户权限配置集合 X，$X=\{X_1,\cdots,X_n\}$，则统计量为

$$S_n=\sum_j X_j,\ 1\leqslant j\leqslant n$$

用户撤销权限比例的极大似然估计量为

$$\hat{p}=\overline{X}=\frac{S_n}{n} \tag{2}$$

当样本空间较大时，S_n 近似服从均值 $\mu=np$、方差 $\sigma^2=np(1-p)$ 的正态分布，其中，σ^2 包含未知参数 p，可用 $\hat{\sigma}^2=n\hat{p}\ (1-\hat{p})$ 来近似，于是 $\dfrac{S_n-np}{\hat{\sigma}}$ 近似服从标准正态分布。参数 p 的置信度为

$\gamma = 1-\alpha$，置信区间$(\theta_l, \theta u)$为

$$-z_{\frac{\alpha}{2}} < \frac{S_n - np}{\hat{\sigma}} < z_{\frac{\alpha}{2}}$$

上式可计算为

$$\frac{S_n}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{S_n(1-\frac{S_n}{n})}{n^2}} < p < \frac{S_n}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{S_n(1-\frac{S_n}{n})}{n^2}} \quad (3)$$

4 隐私风险评估与管理方案

4.1 系统模型

PRAS 方案由客户端和服务器组成，如图 1 中虚线框所示，运行时涉及智能设备用户和 APP 中包含的服务提供商。客户端读取智能设备中安装的 APP 列表和权限设置，并发送给服务器。服务器负责评估隐私泄露风险并计算权限配置方案。计算完成后，将数据反馈给客户端，客户端接收数据并向用户展示权限配置结果。从图 1 中可以看出，在 PRAS 方案执行前，服务提供商可能获取用户全部的隐私信息。而在执行该方案后，服务提供商只能获取用户部分的隐私信息。

PRAS 方案客户端工作流程为：1) 用户选择个性化隐私保护参数，读取 APP 列表、APP 的版本号和授予的权限列表，将读取的信息发送给服务器，并等待回复；2) 收到回复后，向用户展示风险评估结果和权限管理方案。

PRAS 方案服务器工作流程分为 2 个阶段：预备阶段和服务阶段。预备阶段为风险评估做好准备：1) 利用现有技术[6-12]，识别出 APP 中包含的所有服务提供商；2) 收集用户权限设置样本，量化权限管理的隐私信息的敏感度；3) 收集 APP 集合，计算权限组合对敏感度带来的非线性影响。在服务阶段，工作流程为：1) 识别 APP 列表中服务提供商，并统计出每个服务商拥有的权限，评估系统的隐私泄露风险；2) 根据用户权限设置样本，计算 APP 服务质量损失；3) 计算权限配置方案，将结果反馈给客户端。

4.2 PRAS 预备阶段

4.2.1 服务提供商识别

服务提供商包括第三方服务的提供商和宿主 APP。宿主 APP 采集用户的隐私信息，因此其开发者也被认为是单独的服务提供商。

现有研究成果[6-12]可以大规模地识别出 APP 中包含的第三方服务商。首先基于相似度匹配的方法[12]检测宿主 APP 中包含的第三方服务库。然后在所找出的第三方服务库中，通过提取特征（例如 URL 字符串、使用了网络通信的接口等特征），能够准确地将服务提供商开发的动态库识别出来。Liu 等[7]利用这些特征训练分类器识别广告库，可以达到 98%以上的准确率。在识别过程中，同一服务商可能同时发布开源的动态库和二进制的动态库，虽然这些库的使用目的有所不同，但模块名的前缀相同，例如检测到的“com.google.ads”和“com.google.protobuf”这 2 个模块，前者是广告库，以二进制的形式被打包到宿主 APP 中；后者是开源项目，不会主动向其服务器发送数据，所以该模块不是服务提供商。另外，有些服务商既发布 APP，又发布二进制的第三方服务库，例如 Google、Facebook 等。如果用户安装了某服务商开发的 APP，又安装了包含其第三方服务库的 APP，则认为该服务商获得了 2 个 APP 所有的权限。

识别出服务提供商后，PRAS 需要统计其获得

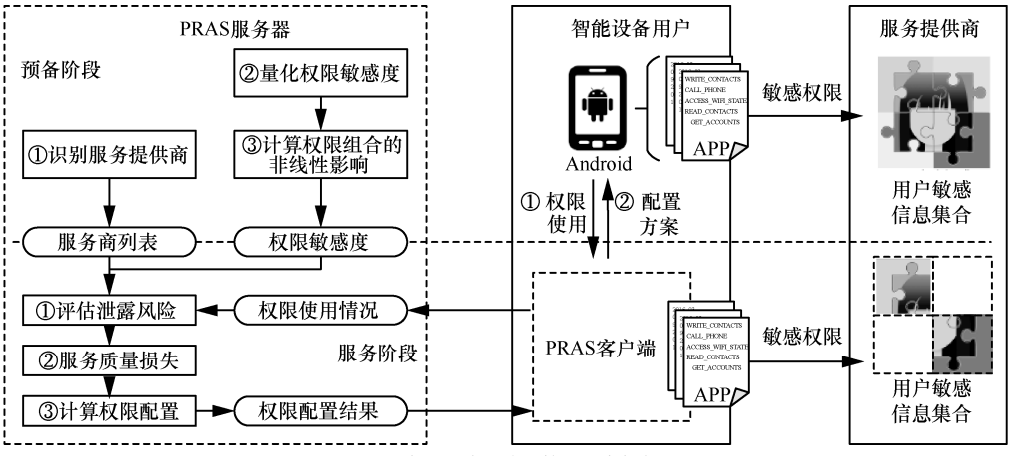


图 1 隐私风险评估及管理方案框架

用户的隐私信息，可以使用静态分析[21]和动态分析[22]结合的方法分析库文件，判断用户的隐私数据是否泄露到外部服务器上。静态分析通过分析 APP 的二进制代码，判断是否有侵犯用户隐私信息的行为发生，但由于动态加载技术的存在，APP 可能在运行时下载侵犯用户隐私的代码并执行，这种情况下会发生漏判隐私泄露事件的情况，因此，静态分析评估的风险值偏低。动态分析（如 TaintDroid 等[22]方案）在 APP 使用的过程中记录敏感信息的调用情况，对隐私泄露事件监测较准确，但需要性能开销较大，且需要 root 权限，只能在代码安全检测的场景中适用，应用市场上并没有类似应用。PRAS 方案的目的是让普通用户实时掌握隐私泄露情况，方便其管理权限，防止在用户一无所知的情况下，隐私信息流向不可信的第三方服务商，因此，本文假设服务商获得权限后，就获得了权限对应的隐私信息。在该假设下，PRAS 评估的隐私泄露风险相对偏高，但优点是可以方便地部署在各种类型的 Android 系统中，且不会带来额外性能开销，偏高的风险评估结果也可以更好地给对用户起警示作用。

4.2.2 权限隐私敏感度量化

基于用户授予/撤销权限的情况可以计算权限的敏感度[23]。与用户隐私相关的敏感权限集合 P 中，$P=\{P_1,P_2,\cdots\}$，权限 P_i 的敏感度 β_i 体现用户对 P_i 所管理隐私信息的重视程度，敏感度越高，用户越重视，则该权限被撤销的可能性越大。使用第 3.3 节中介绍的参数估计方法，可由式(2)估算出用户撤销权限 P_i 的概率 $\hat{p}$。

本文中权限的隐私敏感度 β_i 定义为将 $\hat{p}$ 从 $[0,1]$ 映射到 $[1,+\infty)$，以方便下文计算权限组合的非线性影响。敏感度 β_i 为

$$\beta_i = 1-\log(1-\hat{p}) \tag{4}$$

其中，若 $\hat{p}$=0，表示 Y 中 P_i 没有被撤销过，则 β_i=1；若 $\hat{p}=0.9$，表示 Y 中 90%的情况下 P_i 被撤销，则 β_i=2；若 $\hat{p}$=1，表示 Y 中 P_i 没有被授予过，则 $\beta_i=+\infty$。

4.2.3 权限组合对敏感度的非线性影响

收集标准 APP 集合 D_n 和恶意侵犯隐私的集合 D_m，其中，D_n 选自 APP 市场中每种类型排名前若干位的 APP，因为其下载量大，在采集隐私信息方面受到关注度较多，对申请的权限与实际功能较相符，所以选为标准 APP 集合；D_m 选自 APP 市场中申请敏感权限最多的若干个 APP，其申请的权限越多，且 APP 排名不高，则认为对用户隐私侵犯越严重，因此选入恶意集合中。

权限的非线性影响由 D_n 和 D_m 之间对权限申请次数的差异产生。权限 P_i 在 D_m 和 D_n 中被申请的次数相差越大，该权限被滥用的情况越严重，则对其敏感度的非线性影响越大。对于 D_n 和 D_m 中 APP 申请的敏感权限，分别使用关联规则算法找出 $k-$频繁项集，频繁项集是由权限组成的集合 A。关联规则参数 k 表示 A 中权限数目，$k\in[1,K]$，K 是权限组合 A 中包含权限数目的最大值。权限组合 A 的权限敏感度由式(2)累加得出，即

$$\beta_A=\sum_{i=1}^{k}\beta_i \tag{5}$$

其中，k 是组合 A 中权限的数目。

对权限组合 A 而言，关联规则支持度 $\mathrm{supp}(A)$ 的含义是该权限组合在 APP 集合中出现的比例，对 APP 集合 D 来说，其支持度为

$$\mathrm{supp}_D(A)=\frac{\mathrm{count}_D(A)}{|D|}$$

其中，$\mathrm{supp}_D(A)$ 表示权限组合 A 在数据集 D 的支持度，$|D|$ 表示数据集中包含 APP 的数目，$\mathrm{count}_D(A)$ 表示 D 中使用权限组合 A 的 APP 次数。在关联规则 Eclat 算法执行之前先设定最低的支持度阈值。如果权限组合 A 的支持度 $\mathrm{supp}(A)$ 小于该阈值，则忽略该组合。

在 2 个数据集上分别执行关联规则算法后，得到 2 个权限组合的集合 set，$\mathrm{set}\in\{A_1,A_2,\cdots\}$，$D_n$ 对应集合 set_n，D_m 对应集合 set_m。对 set_m 中的权限组合 A_j，在 D_n 和 D_m 之间的差异 $\mathrm{diff}(A_j)$ 表示权限组合 A_j 是否被滥用。$\mathrm{diff}(A_j)$ 表示为

$$\mathrm{diff}(A_j)=\mathrm{supp}_m(A_j)-\mathrm{supp}_n(A_j)$$

当 $\mathrm{diff}(A_j)>0$ 时，表示 A_j 在 D_n 中的比例比在 D_m 上低，权限组合在恶意集合 D_m 中被滥用。差异值越大，A_j 被滥用的越严重。当 $\mathrm{diff}(A_j)\leqslant 0$ 时，该权限组合没有被滥用。根据权限组合 A_j 的差异 $\mathrm{diff}(A_j)$，A_j 的非线性差异值 $\mathrm{diff}'(A_j)$ 为

$$\mathrm{diff}'(A_j)=\begin{cases}\beta_{A_j}\left(1-\log\left(1-\mathrm{diff}(A_j)\right)\right), & \mathrm{diff}(A_j)\geqslant 0\\ 0, & \text{其他}\end{cases}$$

当 $\mathrm{diff}\left(A_j\right)>0$ 时，$\mathrm{diff}^{'}\left(A_j\right)\in\left[\beta_A,+\infty\right)$。

$\mathrm{diff}^{'}\left(A\right)$ 计算完成后，对权限 P_i 所管理隐私信息的敏感度进行量化。在 set_m 中挑选出包含 P_i 的权限的组合组成集合 $\mathrm{set}^{'}_i$，$\mathrm{set}^{'}_i=\{\mathrm{diff}^{'}\left(A_1\right),\mathrm{diff}^{'}\left(A_2\right),\cdots\}$，权限 P_i 的非线性敏感度 $\beta^{'}_i$ 定义为

$$\beta^{'}_i=\frac{1}{\left|\mathrm{set}^{'}_i\right|}\sum_{j=1}^{\left|\mathrm{set}^{'}_i\right|}\left(\omega_{ij}\mathrm{diff}^{'}\left(A_j\right)\right) \tag{6}$$

其中，$\left|\mathrm{set}^{'}_i\right|$ 表示 $\mathrm{set}^{'}_i$ 中权限组合的数目，ω_{ij} 表示权限 P_i 在 A_j 的权重，$\omega_{ij}\in\left[0,1\right]$。

对 $\mathrm{set}^{'}_i$ 中的 A_j，ω_{ij} 由 P_i 对 A_j 的重要性计算得出。A_j-P_i 表示 A_j 中删除权限 P_i 的权限组合。D_m 中包含 A_j 的 APP 数目为 N_{A_j}，包含 P_i 的 APP 数目为 N_{P_i}，包含 A_j-P_i 的 APP 数目为 $N_{A_j-P_i}$，其中 $N_{A_j-P_i}>N_{A_j}$。权限 P_i 对 A_j-P_i 的提升度视为 P_i 在 $\mathrm{diff}^{'}\left(A_j\right)$ 中做出的贡献。如式 (1) 所示，$\mathrm{lift}\left(A_j-P_i\Rightarrow P_i\right)$ 表示 A_j-P_i 与 P_i 之间的相关性，P_i 与 A_j-P_i 的相关性越大，表示 P_i 对 A_j 的贡献越大，则 P_i 在 A_j 中的权重越大。假设 A_j 中包含 k 条权限，则 P_i 对 A_j 的贡献度 contri_{ij} 为

$$\mathrm{contri}_{ij}=\frac{\mathrm{supp}\left(P_i\cup\left(A_j-P_i\right)\right)}{\mathrm{supp}\left(P_i\right)\mathrm{supp}\left(\left(A_j-P_i\right)\right)}=\frac{\frac{N_{A_j}}{\left|D_m\right|}}{\frac{N_{P_i}}{\left|D_m\right|}\frac{N_{A_j-P_i}}{\left|D_m\right|}}=\frac{N_{A_j}\left|D_m\right|}{N_{P_i}N_{A_j-P_i}}$$

根据 contri_{ij}，则权限 P_i 的权重 ω_{ij} 定义为权限 P_i 对 A_j 贡献度，如式(1)所示。

$$\omega_{ij}=\frac{\mathrm{contri}_{ij}}{\sum_{i=1}^{k}\mathrm{contri}_{ij}}$$

与式(5)相类似，权限组合 A_j 的非线性敏感度 $\beta^{'}_{A_j}$ 量化为

$$\beta^{'}_{A_j}=\sum_{i=1}^{k}\beta^{'}_i \tag{7}$$

其中，$\beta^{'}_i$ 为权限 P_i 的非线性敏感度，由式(6)计算得出；k 为 A_j 中包含的权限数目。

4.3 PRAS 服务阶段

4.3.1 隐私泄露风险评估

权限 P_i 泄露的隐私信息风险定义[23]为

$$\mathrm{risk}\left(P_i\right)=\sum_{i=1}^{N}\left(L\left(P_i\right)I\left(P_i\right)\right) \tag{8}$$

其中，$L\left(P_i\right)$ 表示 P_i 权限被授予 APP 的概率。如果 P_i 被授予给 SP，则 $L\left(P_i\right)=1$，否则 $L\left(P_i\right)=0$。$I\left(P_i\right)$ 表示 P_i 权限所管理的隐私信息泄露后造成的影响。权限敏感度越大，则表示用户对一条权限所管理的信息越重视。那么信息泄露后，用户受到的影响也越大。因此，$I\left(P_i\right)=\beta^{'}_i$。

根据预备阶段识别出的服务提供商，PRAS 的服务器输入包含 N 个 APP 的列表 list，$\mathrm{list}=\left\{\mathrm{APP}_1,\mathrm{APP}_2,\cdots,\mathrm{APP}_N\right\}$，其中 APP_i 被授权了 m 条敏感权限，$\mathrm{APP}_i=\left\{P_{i1},P_{i2},\cdots,P_{im}\right\}$。输出从 list 中识别出 M 个第三方服务提供商 $\mathrm{SP}=\left\{\mathrm{SP}_1,\mathrm{SP}_2,\cdots,\mathrm{SP}_M\right\}$，$\mathrm{SP}_j$ 从不同 APP 中共获得 L 条权限 $\mathrm{SP}_j=\left\{P_{j1},P_{j2},\cdots,P_{jL}\right\}$。隐私信息泄露给服务提供商 SP_j 的风险为

$$\mathrm{risk}_j=\sum_{k=1}^{L}\mathrm{risk}\left(P_{jk}\right) \tag{9}$$

其中，$\mathrm{risk}\left(P_{jk}\right)$ 为服务提供商 SP_j 获得权限 P_k 后的风险评估结果。在此基础上，对系统中 M 个服务提供商泄露的风险为

$$\mathrm{risk}=\sum_{j=1}^{M}\sum_{k=1}^{L}\mathrm{risk}\left(P_{jk}\right) \tag{10}$$

4.3.2 APP 服务质量损失

在权限管理时必须考虑 APP 的可用性。对 APP 而言，权限撤销的越多，隐私泄露的风险越低，其服务质量也越低。极端情况下，将所有的权限都撤销，隐私泄露的风险最低，但其正常功能将无法使用。同时，APP 的可用性与用户的主观认识相关，可以从用户对同类 APP 的权限设置体现出趋势，例如若 80%的用户对社交类 APP 撤销了位置权限，则说明撤销该权限后对此类服务质量损失不大；若没有用户对地图类 APP 撤销位置权限，则说明撤销后对地图类 APP 服务质量损失非常大。因此，需要定义撤销权限 P_i 对 c 类 APP 服务质量损失。权限越敏感，撤销的比例越少，则说明对 c 类 APP 越重要，撤销后服务

质量损失越大。使用第 3.3 节中介绍的参数估计方法，可由式(2)估算出用户对 c 类 APP 撤销权限 P_i 的概率 $\hat{p}$。

撤销权限 P_i 对 c 类 APP 服务质量损失定义为

$$\text{loss}_c\left(P_i\right)=\beta_i^{'}(1-\hat{p}) \tag{11}$$

其中，$\beta_i^{'}$ 与 $\text{loss}_c\left(P_i\right)$ 成正比，撤销权限 P_i 与 $\text{loss}_c\left(P_i\right)$ 成反比。

若 APP_i 属于 c_i 类，撤销权限前授予的权限数为 m，撤销后授予的权限数为 m'，撤销的权限集合为 $m-m'$。则 APP_i 服务质量损失为

$$\text{loss}\left(\text{APP}_i\right)=\sum_{k=1}^{m-m'}\text{loss}_{c_i}\left(P_k\right) \tag{12}$$

在此基础上，整个系统损失的服务质量为 N 个 APP 服务质量损失之和，即

$$\text{loss}=\sum_{i=1}^{N}\text{loss}\left(\text{APP}_i\right) \tag{13}$$

其中，每个 APP 的服务质量损失由式(12)得出。

4.3.3 权限配置方案

减少服务提供商 SP_j 获得的权限数量可以降低隐私泄露的风险。

给定应用列表，N 个 APP 的权限配置表示为 $\text{conf}=\left\{\text{conf}\left(\text{APP}_1\right),\text{conf}\left(\text{APP}_2\right),\cdots\right\}$，$\text{conf}\left(\text{APP}_i\right)$ 表示 APP_i 授予/撤销权限的情况。所有的 conf 可能出现的情况组成权限配置空间 Γ，对任意 conf，有 $\text{conf}\in\Gamma$。在 conf 配置下，系统隐私信息泄露的风险如式(10)所示，表示为 $\text{risk}\left(\text{conf}\right)$。服务质量损失如式(13)所示，表示为 $\text{loss}\left(\text{conf}\right)$。权限管理的最优化目标是，搜索权限配置空间 Γ，找到配置 $\text{conf}^{'}$，在满足一定约束条件下，系统隐私泄露风险值下降最大。

搜索整个 Γ 空间使 PRAS 方案计算量过大，通过减少搜索配置空间 Γ 次数来降低运算量。权限配置空间大小为 $\left|\Gamma\right|=2^{N|P|}$，$\left|P\right|$ 为权限的数量。若配置 $\text{conf}^{''}$ 中出现的权限是 conf 的真子集，则 $\text{conf}^{''}$ 中授予的权限在 conf 中定会被授予，conf 中授予的权限在 $\text{conf}^{''}$ 中不一定被授予，此时 $\text{conf}^{''}<\text{conf}$，且有 $\text{risk}\left(\text{conf}^{''}\right)<\text{risk}\left(\text{conf}\right)$，$\text{loss}\left(\text{conf}^{''}\right)>\text{loss}\left(\text{conf}\right)$。因此，在计算权限配置方案时，剪裁配置空间 Γ，选取所有 $\text{conf}^{''}<\text{conf}$ 的配置。个性化权限配置方案最优化模型为

$$\begin{aligned}&\max\quad z=\text{risk}\left(\text{conf}\right)-\text{risk}\left(\text{conf}^{'}\right)\\&\text{s.t.}\begin{cases}\text{conf}^{'}<\text{conf}\\\dfrac{\max\left(\cdot\right)-\min\left(\cdot\right)}{\max\left(\cdot\right)}<\text{LOSS}_{\text{diff}}\\\dfrac{\text{loss}\left(\text{conf}^{'}\right)-\text{loss}\left(\text{conf}\right)}{\text{loss}\left(\text{conf}\right)}<\text{LOSS}\end{cases}\end{aligned} \tag{14}$$

其中，$\text{conf}^{'}$ 为最优权限配置结果；$\max\left(\cdot\right)$ 与 $\min\left(\cdot\right)$ 表示 $\text{conf}^{'}$ 中单个 APP 服务损失的最大值和最小值；$\text{LOSS}_{\text{diff}}$ 和 LOSS 为个性化隐私保护参数。$\text{LOSS}_{\text{diff}}$ 表示最大值与最小值差值的百分比，该阈值避免计算最优化方案时对单一 APP 撤销过多的权限，$\text{LOSS}_{\text{diff}}\in(0,1)$，LOSS 表示系统整体服务质量损失的百分比，由 $\text{conf}\to\text{conf}^{'}$ 时，系统整体服务质量的损失不能超过 LOSS，$\text{LOSS}\in(0,1)$。

5 实验分析

本文基于 Android 6.0（API level 23）实现 PRAS 方案客户端，因为从此版本开始可在运行时授予/撤销控制权限，PRAS 方案计算出最佳的权限配置方案后，用户可以随时对权限进行设置。Android 低版本的 APP 支持运行在高版本的系统中，因此 PRAS 方案客户端可运行在所有 Android 6.0 以上系统。

5.1 数据集

权限集　Android 系统的权限从 6.0 版本开始分为一般权限和运行时权限，其中，一般权限在安装时默认授予，运行时权限与用户隐私信息相关。Android 的运行时权限分为 10 个权限组，共 26 条权限，本文实验中敏感权限集合 P 是基于运行时权限。另外，由于读取 Wi-Fi 列表的权限与位置隐私直接相关[24]，将该权限加入 P 中。因此，权限集合 P 中权限数量 $\left|P\right|=27$。

用户数据集　本文模拟了 50 位用户的 APP 安装列表和权限配置作为实验数据集 Y。安装的 APP 从每类排名前 20 的 APP 中随机产生，平均每位用户安装了 17 个 APP，其中，最多安装 24 个，最少安装 7 个。

APP 集合　基于非官方的开源项目[25]，本文从 Google 官方的应用商店（Google play store）抓取 2 089 169 个 APP 详情，这些 APP 在商店中被分为 39 类。本文选取每类 APP 排名前 50 的 APP 作为标准 APP 集合 D_n，故 $\left|D_n\right|=1\,950$。相应地，对集合 D_m

而言，选取使用敏感权限最多的 2 000 个 APP 作为恶意侵犯隐私的集合 D_m，$|D_m|=2\ 000$，且 $|D_m|\cap|D_n|=22$。集合 D_n 和 D_m 中权限的统计信息如表 1 所示。

表 1　　APP 使用权限数目统计

数据集	最大值/个	最小值/个	平均值/个	中位数/个
D_n	25	3	4.18	4
D_m	24	15	16.45	16

5.2　权限敏感度量化实验

使用关联规则 Eclat 算法，在 D_m 上挖掘频繁出现的权限组合。根据 D_n 和 D_m 数据集上权限组合出现的差异，计算权限非线性敏感度。

频繁权限组合的挖掘过程中，最小支持度对算法产生的权限组合的影响如图 2 所示。图 2 中左纵坐标轴为最小支持度与 set_m 中申请敏感权限数目的关系。虽然 P 中有些权限与用户隐私相关，但 APP 申请的次数较少，支持度较低，因此在算法挖掘出的权限组合中不会出现。从图 2 中分析出，当最小支持度低于 0.775 之后，涉及的权限数目保持不变。图 2 中右纵坐标轴表示最小支持度与 set_m 中挖掘出权限组合数 $|\text{set}_m|$ 之间的关系。最小支持度 supp_m 越低，出现的权限组合越多。具体来说，当 $\text{supp}_m=0.90$ 时，$|\text{set}_m|=7$；当 $\text{supp}_m=0.75$，$|\text{set}_m|=414$；当 $\text{supp}_m=0.70$ 时，$|\text{set}_m|=870$。

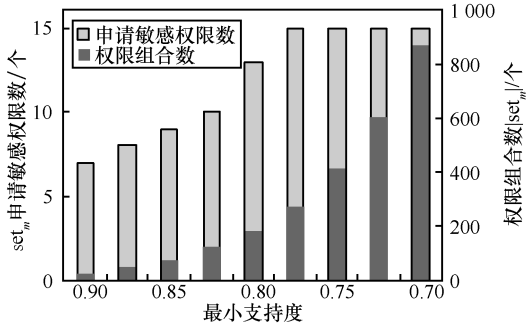


图 2　最小支持度与 set_m 之间变化关系

PRAS 方案计算出的权限量化结果如表 2 所示。量化实验中，用户撤销权限 P_i 的概率 p 的最大似然估计为 $\hat{p}$，当置信度 $\gamma=0.95$ 时，α =0.05，查表得 $z_{\frac{\alpha}{2}}$=1.96，置信度区间为(θ_l,θ_u)，β_i 为权限量化的敏感度，$\beta_i^{'}$ 为考虑权限组合因素后的敏感度量化结果。表 2 中按照 $\beta_i^{'}$ 相对 β_i 的增长率排序，反映 D_n 和 D_m 之间的差异对权限敏感度的提升，其中除了“access_wifi_state”之外其他都可以被撤销。最小值支持度 $\text{supp}_m=0.75$ 时，涉及敏感权限集合 P 中的 15 条权限，其余权限在集合之间的差异不够大，因此忽略。根据收集用户权限设置数据集 Y，β_i 由式(6)计算得出。β_i 与用户对权限的敏感度成正比。权限被撤销越多，则对用户越敏感，β_i 越大。其中有部分权限没有被撤销过，则 β_i =1。权限以组的方式被撤销/授予。例如 APP 申请权限“read contacts”和“write contacts”，系统界面只提示一次，同一权限组内的权限被撤销次数相同，因此 β_i 都相等，β_i =1.933。$\beta_i^{'}$ 表示权限组合在 D_n 和 D_m 之间的差异对敏感度影响的结果。表 2 中排名前五的权限都有超过 100%的增长率，这些权限用户撤销的比例较低，在基于撤销比例的风险评估结果[23]中，这些权限的权重不高，不能引起用户关注。但是，现有研究成果发现恶意 APP 利用“access wifi state”权限[24]和“read call log”权限[26]可获取用户隐私信息。在 PRAS 中，这些权限对隐私泄露风险评估的重要性可以明显地体现出来，达到了方案设计的目的。

当 $\text{supp}_m=0.75$ 时，非线性敏感度 $\beta_A^{'}$ 的统计结果如图 3 所示。其中，横坐标表示 set_m 中组合权限数 k，左纵坐标轴与柱状图表示 k 条权限的组合数。图 3 中大部分组合包含 3 或 4 条，最多包含 6 条敏感权限。图 3 右纵坐标轴与箱型图表示 k 条权限时权限组合 A 非线性敏感度。箱型图上/下沿表示 $k-$权限时，$\beta_A^{'}$ 的最大/小值。图 3 中，2−权限 $\beta_A^{'}$ 的最小值为 2.66，最大值为 4.08，平均值为 3.36，6-权限 $\beta_A^{'}$ 最大值为 9.95，最小值为 9.47，平均值为 9.66，随着 k 增加，敏感度 $\beta_A^{'}$ 也随之增加，说明频繁出现的权限组合 A 中，出现权限数目越多，A 管理的隐私信息也越多，A 就越敏感。

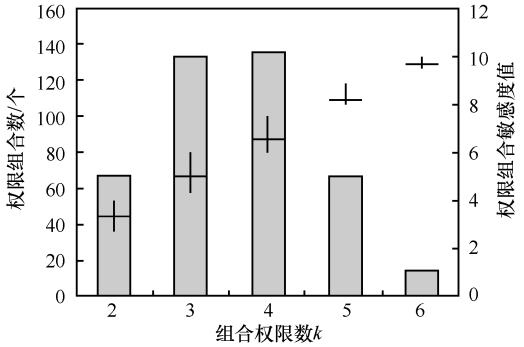


图 3　关联规则算法最小支持度 0.75 时，414 个权限组合 $\beta_A^{'}$ 统计

表 2　　**权限敏感度量化值**

权限	$\hat{p}$	θ_l	θ_u	β_i	$\beta_i^{'}$	增长率
access_Wi-Fi_state	0.0	0.0	0.0	1.0	2.266	126%
read_external_storage	0.03	0.015	0.046	1.030	2.290	122%
write_call_log	0.0	0.0	0.0	1.0	2.150	115%
read_call_log	0.0	0.0	0.0	1.0	2.137	113%
call_phone	0.087	0.039	0.14	1.092	2.307	111%
read_phone_state	0.202	0.155	0.249	1.226	2.350	91.60%
access_fine_location	0.310	0.262	0.357	1.371	1.964	43.21%
receive_SMS	0.577	0.494	0.661	1.859	2.602	39.95%
access_coarse_location	0.356	0.303	0.408	1.440	1.989	38.10%
read_SMS	0.624	0.539	0.709	1.978	2.724	37.73%
get_accounts	0.207 8	0.169	0.244	1.233	1.646	33.51%
write_contacts	0.350	0.263	0.438	1.431	1.905	33.14%
read_contacts	0.469	0.410	0.528	1.633	2.171	32.97%
send_SMS	0.112	0.046	0.177	1.118	1.429	27.85%
camera	0.485 2	0.429	0.537	1.664	2.055	23.45%

5.3　隐私泄露风险评估

在用户数据集 Y 中，不撤销权限的情况下，识别出服务商从用户获取权限的统计如图 4 所示。其中，横坐标为用户索引号，按包含的服务商的数目从大到小排序。左纵坐标轴和柱状图为系统中所有服务提供商的数目，其中 0 号设备中包含 67 个服务商，49 号设备中包含 24 个服务商，平均每个用户包含 48 个服务商。右纵坐标轴和箱型图表示服务提供商从智能设备中获取权限详情，箱型图上/下沿表示获得权限最多/最少的服务商。所有识别出的服务商中，平均每个服务商从用户获得 11 条权限。每部设备中，获得最多权限的服务商平均得到 19 条权限，获得权限最少的服务商平均获得 2 条，使用最多 3 条权限以依次“read phone state”“access wifi state”和“read external storage”。

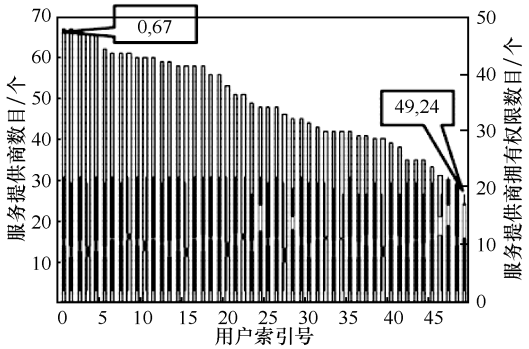


图 4　单台智能手机中包含的服务商统计

以数据集中 0 号用户作为典型示例，分析其隐私信息泄露风险，如图 5 所示。其中，横坐标为包含的服务商索引号，左纵坐标轴表示隐私泄露风险，右纵坐标轴表示服务商获取权限来源 APP 的数目。从图 5 中可以看出，0 号服务提供商从 15 个 APP 中获取权限，隐私信息泄露风险为 31.38；5~13 号服务提供商从 2 个 APP 中获取权限，由于得到的权限数目及敏感度不同，风险值由 13.79 变化到 28.3；14~66 号服务提供商从一个 APP 中获取权限，风险值由 3.29 变化到 29.72。

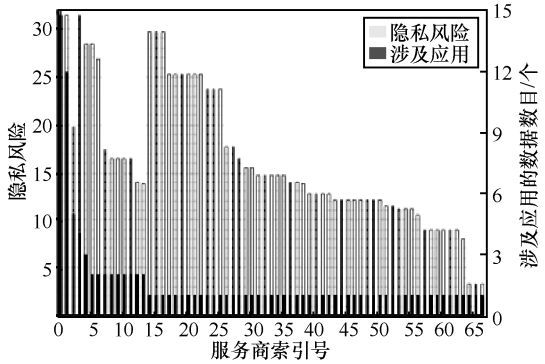


图 5　典型用户的隐私泄露详情

智能设备中隐私泄露的风险与权限决策优化结果如图 6 所示，选取 $LOSS_{diff}$=0.2 和 LOSS=0.3。优化后，单个应用服务质量损失的最大值和最小值的差值不超过 20%，且整体服务质量损失不超过

30%。图中横坐标为用户索引号，按照隐私泄露风险值从大到小对用户排序，纵坐标为隐私泄露的风险值。从图 6 中可以看出，按照式(14)优化后，每位用户隐私信息泄露的风险得到不同程度的降低。优化后，用户隐私泄露风险最大下降 29.4%，最少下降 9.6%，平均下降 18.5%，下降百分比的中位数为 18.2%。

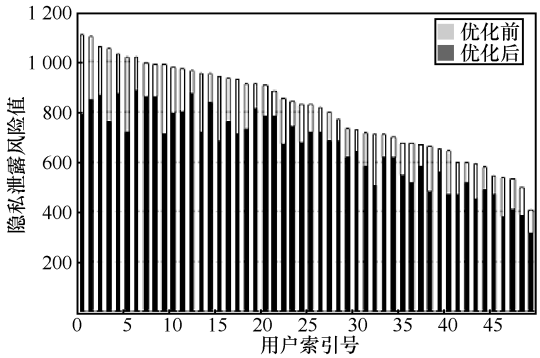


图 6 典型用户隐私泄露风险优化

6 结束语

为缓解移动设备日益严峻的隐私信息泄露风险，本文针对智能设备中 APP 包含第三方服务提供商非法读取用户隐私信息的问题，提出了一种风险评估方案。该方案通过频繁项集 Eclat 算法挖掘常用权限组合，计算恶意侵犯隐私的 APP 集合与正常 APP 集合之间权限组合支持度的差异，量化权限组合对权限敏感度的影响，识别出系统中包含的服务提供商，构建模型评估系统整体的隐私信息泄露风险。在 APP 整体的服务质量与隐私保护之间做出均衡分析，构建个性化最优模型，计算系统整体的权限管理方案。实验结果进一步地验证了所提方案的有效性和高效性。

参考文献：

[1] 李凤华, 李晖, 贾焰, 等. 隐私计算研究范畴及发展趋势[J]. 通信学报, 2016, 37(4): 1-11.
LI F H, LI H, JIA Y, et al. Privacy computing: concept, connotation and its research trend[J]. Journal on Communications, 2016, 37(4): 1-11.

[2] 中国消费者协会. APP 个人信息泄露情况调查报告[R]. 中国消费者协会, 2018.
CCA. Survey on personal information leakage by APP[R]. China Consumers Association, 2018.

[3] 奇虎 360. 2018 中国手机安全生态研究报告[R]. 北京奇虎科技有限公司, 2018.
Qihoo 360. China mobile phone safety ecology report[R]. Qihoo 360 Technology Co., Ltd., 2018.

[4] GRACE M C, ZHOU W, JIANG X, et al. Unsafe exposure analysis of mobile in-APP advertisements[C]//The ACM Conference on Security and Privacy in Wireless and Mobile Networks. ACM, 2012: 101-112.

[5] CHEN K, LIU P, ZHANG Y. Achieving accuracy and scalability simultaneously in detecting application clones on Android markets[C]//The ACM International Conference on Software Engineering. ACM, 2014: 175-186.

[6] NARAYANAN A, CHEN L, CHAN C K. Addetect: automated detection of android ad libraries using semantic analysis[C] //The IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing. IEEE, 2014: 1-6.

[7] LIU B, LIU B, JIN H, et al. Efficient privilege de-escalation for ad libraries in mobile APPs[C]//The ACM Annual International Conference on Mobile Systems, APPlications, and Services. ACM, 2015: 89-103.

[8] CRUSSELL J, GIBLER C, CHEN H. Scalable semantics-based detection of similar android applications[C]//The European Symposium on Computer Security. 2013: 1-21.

[9] WANG H, GUO Y, MA Z, et al. WuKong: a scalable and accurate two-phase approach to Android APP clone detection[C]//The ACM International Symposium on Software Testing and Analysis. ACM, 2015: 71-82.

[10] MA Z, WANG H, GUO Y, et al. LibRadar: fast and accurate detection of third-party libraries in Android apps[C]//The ACM International Conference on Software Engineering. ACM, 2016: 653-656.

[11] LI M, WANG W, WANG P, et al. LibD: scalable and precise third-party library detection in android markets[C]//The ACM International Conference on Software Engineering. ACM, 2017: 335-346.

[12] BACKES M, BUGIEL S, DERR E. Reliable third-party library detection in Android and its security applications[C]//The ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016: 356-367.

[13] FELT A, HA E, EGELMAN S, et al. Android permissions: User attention, comprehension, and behavior[C]//The ACM Symposium on Usable Privacy and Security. ACM, 2012: 1-14.

[14] FAWAZ K, SHIN K G. Location privacy protection for smartphone users[C]//The ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014: 239-250.

[15] TSAI L, WIJESEKERA P, REARDON J, et al. Turtle guard: helping android users apply contextual privacy preferences[C]//The ACM Symposium on Usable Privacy and Security.ACM, 2017: 145-162.

[16] AGARWAL Y, HALL M. ProtectMyPrivacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing[C]//The ACM Annual International Conference on Mobile Systems, APPlications, and Services. ACM, 2013: 97-110.

[17] LIU B, LIN J, SADEH N. Reconciling mobile app privacy and usability on smartphones: could user privacy profiles help?[C]//The ACM

International Conference on World Wide Web. ACM, 2014: 201-212.

[18] LIU R, CAO J, YANG L, et al. PriWe: recommendation for privacy settings of mobile APPs based on crowdsourced users[C]//IEEE International Conference on Mobile Services. IEEE, 2015: 150-157.

[19] RASHIDI B, FUNG C, NGUYEN A, et al. Android user privacy preserving through crowdsourcing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(3): 773-787.

[20] ZAKI M J. Scalable algorithms for association mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(3): 372-390.

[21] LU L, LI Z, WU Z, et al. CHEX: statically vetting android apps for component hijacking vulnerabilities[C]//The ACM SIGSAC Conference on Computer and Communications Security. ACM, 2012: 229-240.

[22] ENCK W, GILBERT P, HAN S, et al. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones[J]. ACM Transactions on Computer Systems, 2014, 32(2): 1-29.

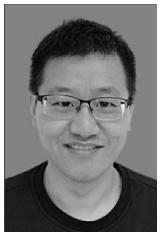
[23] LIU K, TERZI E. A framework for computing the privacy scores of users in online social networks[J]. ACM Transactions on Knowledge Discovery from Data, 2010, 5(1): 1-30.

[24] LI F H, WANG X Y, NIU B, et al. TrackU: exploiting user's mobility behavior via wifi list[C]//IEEE Global Communications Conference (GLOBECOM). IEEE, 2017: 1-6.

[25] EGIRAULT. Google play unofficial python API[Z]. GitHub, 2016.

[26] XING L, PAN X, WANG R, et al. Upgrading your android, elevating my malware: privilege escalation through mobile OS updating[C]//The IEEE Symposium on Security and Privacy. IEEE, 2014: 393-408.

[作者简介]



王新宇（1989− ），男，甘肃平凉人，中国科学院信息工程研究所博士生，主要研究方向为信息保护、隐私计算。

牛犇（1984− ），男，陕西西安人，博士，中国科学院信息工程研究所副研究员，主要研究方向为网络安全、隐私计算。

李凤华（1966− ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。

贺坤（1995− ），男，安徽安庆人，中国科学院信息工程研究所硕士生，主要研究方向为信息保护、隐私计算。