

多维度 iOS 隐私泄露评估模型研究

邢月秀¹, 胡爱群¹, 王永剑², 赵然¹

(1. 东南大学信息安全研究中心, 江苏 南京 210096;
2. 公安部第三研究所, 上海 200031)

摘要: 针对目前 iOS 平台隐私泄露检测缺乏系统性的评估方法, 提出了一种多维度 iOS 隐私泄露评估模型。该模型分为综合静态分析、动态分析和网络数据分析方法, 从多维度对应用程序隐私泄露行为进行特征抽取和评估。实验测试了 30 款来自苹果 App Store 不同类型的应用程序, 发现超过 50% 的应用程序会读取用户位置信息, 约 40% 的应用程序存在未经用户同意发送数据到服务器的情况。该模型弥补了单一使用静态分析或动态分析方法的局限性, 有效解决了隐私泄露的量化问题。

关键词: iOS; 隐私泄露; 静态分析; 动态分析; 网络数据分析

中图分类号: TN929.53

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00043

Research on multi-dimensional iOS privacy disclosure evaluation model

XING Yue-xiu¹, HU Ai-qun¹, WANG Yong-jian², ZHAO Ran¹

(1. Information Security Research Center, Southeast University, Nanjing 210096, China;
2. The Third Research Institute of Ministry of Public Security, Shanghai 200031, China)

Abstract: The existing iOS platform is lacking of systematic assessment methods of privacy leak detection. To solve this problem, a multi-dimensional iOS privacy disclosure evaluation model was presented. This model combined static analysis, dynamic analysis and network data analysis method to extract the features of the application's privacy disclosure behavior and evaluate it from multiple dimensions way. The model on 30 different types of apps from the iOS App Store and found out that more than 50% of all investigated apps are tracking users' locations were evaluated, almost 40% of all send data to a server without the user's consent. The model makes up for the limitations of single static analysis or dynamic analysis methods, solves the quantization problem of privacy disclosure effectively.

Key words: iOS, privacy disclosure, static analysis, dynamic analysis, network data analysis

1 引言

互联网数据中心 (IDC, Internet data center) 2015 年第二季度智能手机市场调查报告^[1]显示: iOS 智能手机的市场占有率为 13.9%, 总出货量达到了 4 750 万部。智能手机普及的同时, 也带

来了用户隐私泄露等安全问题。但由于 iOS 平台是一个相对封闭的系统, 国内外研究人员对该平台的隐私泄露问题研究相对较少。

2011 年 2 月, Egele 等^[2]提出并实现了一种基于静态分析的 iOS 隐私泄露检测模型 PiOS, 该模型通过数据流分析和切片技术检测 iOS 应用程序

收稿日期: 2016-02-04; 修回日期: 2016-03-08。通信作者: 胡爱群, aqhu@seu.edu.cn

基金项目: 国家重点基础研究发展计划 (“973” 计划) 基金资助项目 (No.2013CB338003)

Foundation Item: The National Basic Research Program of China (973 Program)(No.2013CB338003)

的隐私泄露行为。但是,这种单一的静态分析方法存在无法解析消息目的地等固有缺陷。

2011年6月,Gilbert等^[3]指出PiOS存在较高的误报率,并提出了基于动态分析的iOS隐私泄露检测方法。该方法通过挂钩敏感应用程序编程接口(API, application programming interface)函数,分析iOS应用程序的函数调用行为,判定应用程序是否读取了隐私信息。

2014年,Andreas等^[4]提出并实现了基于动态API调用序列的iOS隐私泄露分析模型DIOS,DIOS主要包括3个部分:backend用作中央数据存储;worker作为backend和iOS设备的数据交互链路;client安装在iOS设备上,用于运行和分析iOS App的行为。该模型通过挂钩iOS API函数监控应用程序访问用户隐私数据的行为,但是应用程序对隐私数据的访问行为并不等同于隐私泄露行为,而且相比于静态分析,动态分析也存在容易遗漏和检测速度较慢的缺点。

综上,目前国内外对iOS隐私泄露的检测大多采用单一的分析方法,缺少对应用程序隐私泄露风险的综合评估。为了更加合理和准确地评估iOS应用程序的隐私泄露,本文在深入研究静态分析、动态分析和网络数据分析模型的基础上,对iOS应用程序隐私泄露行为进行多维度检测,提出并实现了基于多维度的iOS隐私泄露评估模型。

2 多维度iOS隐私泄露评估模型

从多维的角度来看,应用程序主要是由代码、行为和数据3个维度组成,多维度iOS隐私泄露评估模型也分为静态分析、动态分析和网络数据分析3个维度,如图1所示。根据已有的研究可知,静态分析速度快、效率高,可以有效地提高检测率^[5],但是误报率较高;动态分析速度相对较慢,但是可以监控应用程序对隐私数据的实际访问行为,有效地解决了程序加壳、隐藏API调用等静态分析无法解决的问题;网络数据分析则解决了静态分析和动态分析只能检测应用程序对隐私数据的访问,但不能确定应用程序是否通过网络途径泄露隐私数据的问题。3个维度相辅相成,有效地弥补了单一分析方法自身的局限性,可有效提高iOS隐私泄露评估模型的科学性和严谨性。

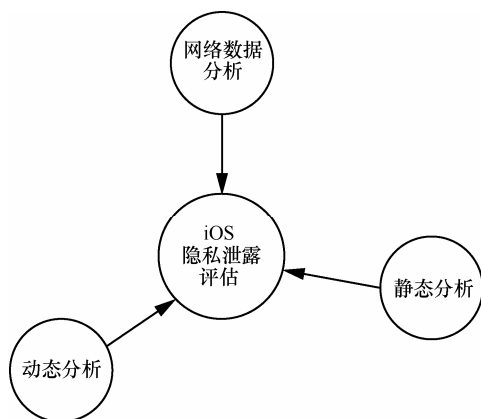


图1 多维度iOS隐私泄露评估模型

2.1 应用程序行为特征的定义

多维度iOS隐私泄露评估模型分为iOS应用程序静态分析、动态分析和网络数据分析3个维度,每个维度对应应用程序隐私泄露行为的一个特征向量。

定义1 静态分析特征向量定义为 S ,内含 $(S_1, S_2, S_3, \dots, S_m)$, m 为静态分析特征向量的维数。其中, $S_i (i=1, 2, \dots, m)$ 表示静态分析中各种隐私泄露行为抽象特征。

定义2 动态分析特征向量定义为 D ,内含 $(D_1, D_2, D_3, \dots, D_n)$, n 为动态分析特征向量的维数。其中, $D_i (i=1, 2, \dots, n)$ 为动态分析中各种隐私泄露行为的抽象特征。

定义3 网络数据分析特征向量定义为 N ,内含 $(N_1, N_2, N_3, \dots, N_t)$, t 为网络数据分析特征向量的维数。其中, $N_i (i=1, 2, \dots, t)$ 为网络数据分析中各种隐私泄露行为的抽象特征。

综合静态分析特征向量、动态分析特征向量和网络数据分析特征向量,iOS应用程序隐私泄露行为特征定义如下。

定义4 应用程序隐私泄露行为特征向量集为 PL , PL 包括静态分析特征向量、动态分析特征向量和网络数据分析特征向量3个元素。

$$PL = [(S_1, S_2, S_3, \dots, S_m), (D_1, D_2, D_3, \dots, D_n), (N_1, N_2, N_3, \dots, N_t)] \quad (1)$$

2.2 隐私泄露风险评估

定义iOS应用程序隐私泄露行为之后,需要对其隐私泄露风险进行量化和评估。根据 PL 中3个元素的隐私泄露相关性大小,设置对应的权重值($Weight$)进行隐私泄露风险量化,量化

完成后,通过权重值加权的方式计算检测结果的总权重,最后归一化计算该结果占模型总权重值的百分比,得到应用程序隐私泄露风险的综合评估值。

定义 5 S_i_w 表示静态分析特征值 S_i 对应的权重值, D_i_w 表示动态分析特征值 D_i 对应的权重值, N_i_w 表示网络数据分析特征值 N_i 对应的权重值。模型中的隐私泄露总权重值为静态分析、动态分析和网络数据分析行为特征向量的权重值之和,计算表达式为

$$Weight_{All} = \sum_{i=1}^m S_i_w + \sum_{i=1}^n D_i_w + \sum_{i=1}^l N_i_w \quad (2)$$

定义 6 iOS 应用程序隐私泄露风险综合评估值为 $App[i]_{Risk}$, 根据静态分析、动态分析和网络数据分析中实际检测到的各种隐私泄露行为,采用权重值加权和归一化方式计算得到评估值。

如一个应用程序的静态分析结果为 $S = (S_2, S_6, S_8)$, 动态分析结果为 $D = (D_1, D_4)$, 网络数据分析结果为 $N = (N_1, N_2)$, 则该应用程序的总权重值 ($App[i]_{Weight}$) 和隐私泄露风险评估值计算如式(3)和式(4)。

$$App[i]_{Weight} = S_2_w + S_6_w + S_8_w + D_1_w + D_4_w + N_1_w + N_2_w \quad (3)$$

$$App[i]_{Risk} = \frac{App[i]_{Weight}}{Weight_{All}} \times 100\% \quad (4)$$

应用程序隐私泄露评估值越大,说明被检测应用程序涉及的隐私泄露行为数量占系统定义的

全部隐私泄露行为的比重越大,可能泄露的隐私信息越多,隐私泄露的可能性就越大。

3 具体实现

3.1 模型实现

多维度 iOS 隐私数据泄露评估模型实现如图 2 所示。

在图 2 中,左边虚线框内为静态分析特征向量抽取流程,主要操作包括:App 存储路径定位、解除应用程序版权保护、反汇编处理和敏感 API 函数扫描;中间虚线框内为动态分析特征向量抽取流程,主要操作包括:编写 Tweak 程序挂钩和隐私数据相关的敏感 API 函数、遍历 App 功能按钮以及分析系统日志获取应用程序调用敏感 API 的行为记录;右边虚线框内为网络数据分析特征向量抽取流程,主要操作包括:通过流量转发,在 PC 端使用 Tshark 对应用程序进行抓包处理,对抓取的数据分组进行协议,地址分析,通过挂钩 SSL 读写函数,在数据分组 SSL 加密之前捕获数据内容,进行数据分组内容分析。

3.2 隐私泄露行为特征抽取

隐私泄露行为特征中静态分析特征向量和动态分析特征向量的抽取主要是基于 API 调用序列的检测^[6]。iOS API^[7]包括 Published API、UnPublished API 和 Private API^[8]。参考 iOS 系统的隐私分类^[9],本文定义了 9 类隐私 API,每类隐私 API 中包括可能访问这类隐私信息的多个关键 API 函数,本文共检测了 20 个 API 函数^[7]。表

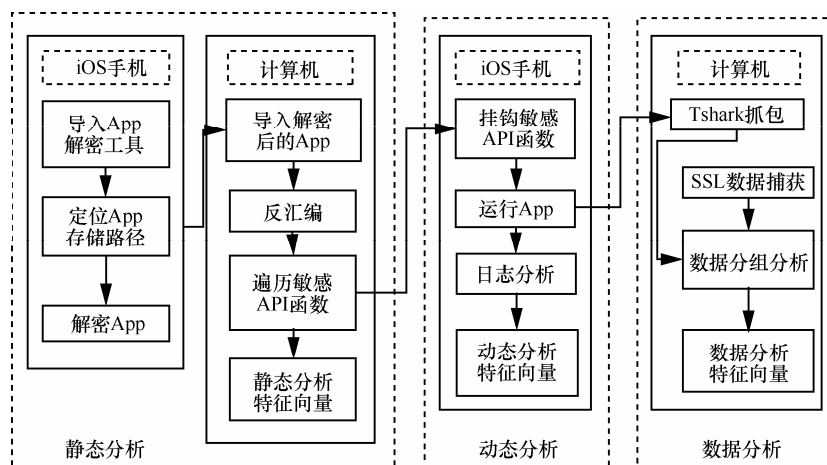


图2 多维度 iOS 隐私数据泄露评估模型实现

表 1

隐私 API 分类

隐私 API 分类	关键 API 函数举例	函数说明
定位服务 API (Location API)	-(void)requestWhenInUseAuthorization	程序前台运行时请求使用定位服务
通讯录 API (Contacts API)	-(void)requestAccessForEntityType:(CNEntityType)entityTypecompletionHandler:(void(^)(BOOL granted, NSError*error)) completionHandler	请求访问通讯录
日历和提醒事项 API (Calendar API)	-(void)requestAccessToEntityType:(EKEntityType)entityType completion: (EKEventStoreRequest AccessCompletionHandler)completion	请求访问日历和提醒事项的用户权限
照片 API (Photo API)	+(void)requestAuthorization:(void(^)(PHAuthorizationStatus status))handler	请求访问照片库的用户权限
麦克风 API (Microphone API)	-(void)requestRecordPermission:(PermissionBlock)response	请求使用麦克风
相机 API (Camera API)	+(instancetype)deviceInputWithDevice:(AVCaptureDevice *)device error:(NSError **)outError	捕获设备输入
健康 API (Health API)	-(void)requestAuthorizationToShareTypes:(NSSet<HKSampleType *> *)typesToSharereadTypes:(NSSet<HKObjectType *> *)typesToRead completion:(void(^)(BOOL success, NSError*error)) completion	请求访问健康数据
社交账户 API (Social API)	-(void)requestAccessToAccountsWithType:(ACAccountType *)accountType options:(NSDictionary *)optionscompletion:(ACAccountStoreRequestAccessCompletionHandler)completion	请求访问社交账户数据
本机信息 API (Nativeinfo API)	extern NSString* CTSettingCopyMyPhoneNumber()	获取本机号码

1 为隐私 API 分类。

静态分析特征向量的抽取主要包括 App 解密、反汇编和敏感 API 遍历 3 个步骤。App Store 发布的应用程序都是经过 FairPlay^[10]加密的二进制文件，所以，静态检测的第一步就是对应用程序进行解密^[11]，获得解密的二进制文件 appname.decrypted 后，调用 hopper Disassembler 软件^[12]对应用程序进行反汇编，得到应用程序的反汇编文件 appname.asm。最后，分析 appname.asm 中的敏感 API 函数，得到静态分析特征向量。

动态分析特征向量的抽取主要是利用 MobileSubstrate^[13]框架编写 Tweak 插件，通过代码注入技术挂钩和监控敏感 API 函数。在应用程序运行期间记录应用程序的函数调用行为。根据上文定义的敏感 API 函数，本文使用 %hook 方法编写对应的挂钩函数对原函数进行覆盖，并在挂钩函数中添加弹框提示和日志记录功能。当应用程序调用相关敏感 API 时，系统会先执行挂钩函数，弹框提示用户，并将其调用行为记录到系统日志中。通过分析系统日志中应用程序的隐私泄露行为，就可得到动态分析特征向量。

网络数据分析特征向量的抽取是使用 Tshark^[14]工具对应用程序所有的数据分组进行网络抓包分析，解析数据分组的协议和目的地

址；编写 Tweak 程序挂钩 SSL 读写函数，捕获未经 SSL 加密的发送数据。综合检测结果进行隐私泄露的网络数据分析，可能检测到 4 种隐私泄露行为。

1) 应用程序发送数据到非法地址，可以解析出发送数据为敏感信息。

2) 应用程序发送数据到非法地址，但无法解析出数据是否为敏感信息。

3) 未经用户同意，应用程序发送数据到应用程序官方服务器，可以解析出发送数据为敏感信息。

4) 未经用户同意，应用程序发送数据到应用程序官方服务器，无法解析出数据是否为敏感信息。

通过检测这 4 种隐私泄露行为，获得网络数据分析特征向量。

3.3 权重设置

静态分析特征向量、动态分析特征向量和网络数据分析特征向量的权重值是根据其对应用程序隐私泄露风险评估的贡献值进行设定的，最后的评估值使用归一化的方式计算。因此，计算结果仅对权重值的相对大小敏感，所以下文权重值根据各个特征向量的相对影响大小确定，权重值的具体数值没有特别含义。静态分析、动态分析和网络数据分析特征向量的

权重值设定依据如下。

1) 静态分析是对应用程序的反编译文件进行隐私 API 扫描,但是扫描到隐私 API 代码并不能证明有实际调用行为的发生,而且静态分析存在较高的误报率^[3],因此,其对应用程序隐私泄露风险评估的贡献值较低。

2) 动态分析通过挂钩隐私 API 函数,监控和记录应用程序对 API 函数的调用行为,记录结果可以代表应用程序对隐私信息的实际访问操作,因此,其对应用程序隐私泄露风险评估的贡献值较高。

3) 静态分析和动态分析的检测对象都是隐私 API 函数,本文定义了 9 组隐私 API,每组隐私 API 又可以根据函数功能划分为 3 类:读写 API、申请授权 API 和查看授权状态 API,每类函数对应用程序隐私泄露风险评估的贡献值也不一样。其中,读写 API 会直接获取用户隐私数据,风险等级高,贡献值高;申请授权 API 向用户申请对某类隐私数据的访问权限,没有直接访问行为,而且用户可以选择拒绝授权,因此,风险等级和贡献值为中;查看授权状态 API 用于查询用户对此类敏感数据的授权状态,如果状态为未授权,应用程序可能发起授权申请,此类 API 仅说明应用程序有访问敏感数据的意图,风险等级为低,贡献值最小。

4) 网络数据分析中直接检测到行为 1) 和行为 3) 时,可以直接判定应用程序通过手机边界泄露了用户隐私数据,即 $App[i]_{Risk}=100\%$ 。行为 2) 和行为 4) 虽然不能确定发送数据是否为本文定义的隐私数据,但是未经用户同意发送数据,和发送数据到非法地址的行为都存在很大的隐私泄露嫌疑。对判定应用程序是否存在直接的隐私数据泄露行为贡献值很高。

综上所述,静态分析、动态分析和网络数据分析对应用程序隐私泄露风险评估的贡献值排序如下:网络数据分析>动态分析>静态分析。其中,静态分析和动态分析中的不同类 API 函数的贡献值排序如下:读写 API>申请授权 API>查看授权状态 API。根据以上排序规则以及评估值计算仅对权重值相对大小敏感的理论基础,可以得出本文具体权重设置,如表 2 所示,最小权重值设为 1,在同一分析方法中权重值递

增间隔为 2,不同分析方法之间权重值递增间隔为 5。

表 2 API 函数权重设置

分析方法	类别	权重值	备注
静态分析	查看授权状态 API	1	/
	申请授权 API	3	/
	读写 API	5	/
动态分析	查看授权状态 API	6	/
	申请授权 API	8	/
	读写 API	10	/
网络数据分析	泄露行为 1)	/	$App[i]_{Risk}=100\%$
	泄露行为 2)	15	/
	泄露行为 3)	/	$App[i]_{Risk}=100\%$
	泄露行为 4)	15	/

应用程序隐私泄露行为特征抽取和权重值设置完成后,可以计算模型总权重值为

$$Weight_{All} = \sum_{i=1}^m S_i \cdot w + \sum_{i=1}^n D_i \cdot w + \sum_{i=1}^l N_i \cdot w$$

$$= 90 + 230 + 30 = 350$$

如上所述,网络数据分析的 N_1 和 N_3 (行为 1)和行为 3)) 不参与 $Weight_{All}$ 的计算,用于单独判定 $App[i]_{Risk}=100\%$ 。

4 系统测试与分析

4.1 测试环境

本文的测试环境如表 3 所示。

表 3 测试环境

设备	配置描述
计算机 CPU	Intel Core i5 2.4 GHz
计算机内存	8 GB 1 600 MHz DDR3
计算机操作系统	OS X Yosemite 10.10.1
iOS 手机版本	iOS 8.3
iOS 手机内存	1 024 MB

4.2 测试结果与分析

4.2.1 应用程序隐私泄露行为

为了保证样本能够最大程度地反映总体情况,本文根据苹果 App Store 的 10 类最受欢迎 App

(社交、摄影、新闻、儿童、工作、运动、学习、美食、游戏、购物), 选择了来自 App Store 的 30 款软件作为模型测试样本, 每类 App 选择下载量排名前 3 的应用程序进行测试。样本的行为特征检测结果如图 3 所示。

图 3 中 9 类敏感 API 的数据取静态分析和动态分析的合集, N_1 、 N_2 、 N_3 、 N_4 分别表示网络数据分析的 4 种隐私泄露行为。静态分析和动态分析的相同检测项的结果取并集。

由图 3 可以看出 50.0% 的应用程序都存在访问用户位置信息的行为, 访问通讯录、相册和相机的应用也超过了 30% 以上, 在网络数据分析结果中可以看到 36.7% 的应用程序都存在未经用户同意发送数据到服务器的行为, 但是难以解析出发送内容。

本文的分析结果与 PiOS 模型^[2]相比, 增加了动态监控和网络数据分析, 解决了基于静态分析的 PiOS 模型存在的两个缺陷: 1) 部分应用程序进行了混淆和加密处理, 无法进行反汇编处理, 从而导致静态分析失效; 2) 静态分析无法解析出应用程序通过容器发出的请求数据 (如 Objective-C 的 NSArray 数据), 而导致不能检测出应用程序发送数据的内容和目的地。本文的动态分析对应用程序运行时分析, 绕过了应用程序代码混淆和加密的问题; 网络数据在系统网络边界对应用程序发送的数据分组进行协议、目的地址和数据内容的分析, 可以解析出应用程序是否存在通过系统网

络接口发送用户隐私数据的行为。

4.2.2 应用程序隐私泄露评估值计算

30 款测试软件中有 24 款软件的隐私泄露评估值在 30% 以下, 4 款软件评估值在 30%~40% 之间。有一款软件 X 的评估值为 60.3%, 其具体检测结果如表 4 所示。

由表 4 可以得到应用程序 X 的总权重值 $App[i]_{Weight} = 211$, 根据上文计算的 $Weight_{All}$, 可以得到应用程序 X 的隐私泄露风险评估值为

$$App[i]_{Risk} = \frac{App[i]_{Weight}}{Weight_{All}} \times 100\% \\ = \frac{211}{350} \times 100\% = 60.3\%$$

还有一款游戏软件 Y 在网络数据分析时检测到发送用户位置信息到非法地址, 即行为 1), 因此判定隐私风险评估值为 100%。

5 结束语

多维度 iOS 隐私泄露评估模型综合静态分析、动态分析和网络数据分析方法对应用程序的隐私泄露行为进行检测, 有效地弥补了单一静态分析或动态分析的局限性, 并且能够捕捉和分析应用程序实际泄露的隐私数据分组, 有效证明了应用程序的隐私泄露行为。模型对应用程序的隐私泄露进行了风险评估, 评估值能在一定程度上代表应用程序泄露隐私的数量多少和可能性大小。

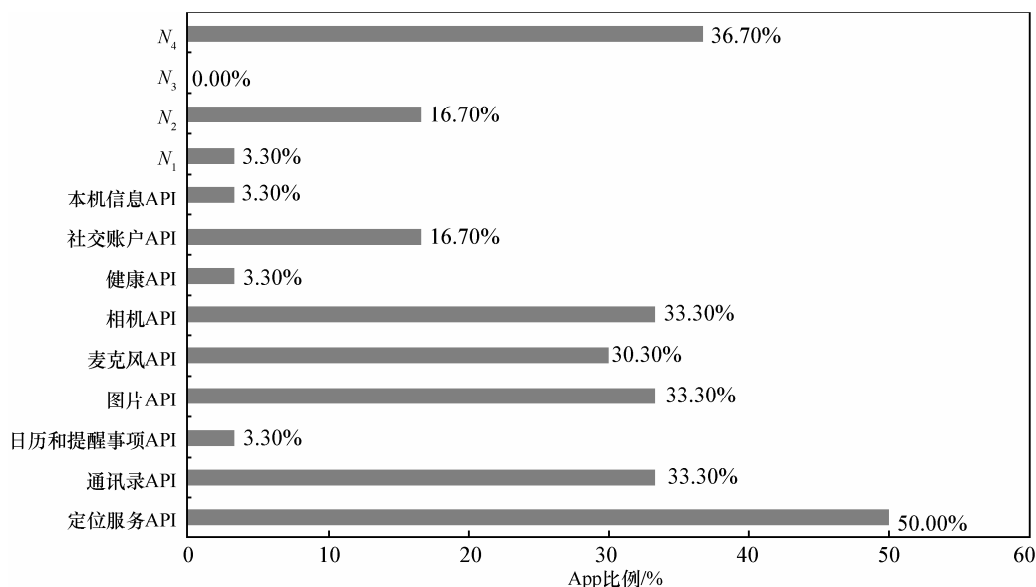


图3 应用程序行为特征检测结果

表 4 应用程序 X 的检测结果

分析方法	隐私 API 分类	API 函数		隐私泄露行为	权重值
		类别	个数		
静态分析	定位服务 API	A,B,C	1,2,1	/	12
	通讯录 API	A,B	1,1	/	8
	日历和提醒事务 API	/	/	/	/
	图片 API	A,B,C	1,1,1,	/	9
	麦克风 API	B	1	/	3
	相机 API	A,B,C	1,1,1	/	9
	健康 API	/	/	/	/
	社交 API	A	1	/	5
	本机 API	/	/	/	/
动态分析	定位服务 API	A,B,C	1,2,1	/	32
	通讯录 API	A,B,C	1,1,1	/	24
	日历和提醒事务 API	/	/	/	/
	图片 API	A,B,C	1,1,1	/	24
	麦克风 API	B	1	/	8
	相机 API	A,B,C	1,1,1	/	24
	健康 API	/	/	/	/
	社交 API	A,B	1,1	/	18
	本机 API	A	2	/	20
网络数据分析	/	/	/	行为 1)不存在	/
	/	/	/	行为 2) 不存在	/
	/	/	/	行为 3) 不存在	/
	/	/	/	行为 4) 存在	15

注: API 函数类别中, A 表示读写类 API, B 表示申请授权类 API, C 表示查看授权状态类 API。

参考文献:

- [1] Smartphone OS market share, 2015 Q2[EB/OL]. <http://www.idc.com/prodserv/smartphone-os-market-share.jsp>.
- [2] EGELE M, KRUEGEL C, KIRDA E, et al. PiOS: detecting privacy leaks in iOS applications[C]//NDSS. c2011:289-291.
- [3] GILBERT P, CHUN B G, LANDON P, et al. Automating privacy testing of smartphone applications: technical report: CS-2011-02[R]. Duke University, 2011.
- [4] KURTZ A, WEINLEIN A, SETTGAST C, et al. DiOS: dynamic privacy analysis of iOS applications: friedrich-alexander-universität erlangen-nürnberg, dept of computer science, CS-2014-03[R]. 2014.
- [5] PENG L, CUI B J. A comparative study on software vulnerability static analysis techniques and tools[C]//Information Theory and Information Security. c2010:521-524.
- [6] EPIFANI M, PASQUALE S. Learning iOS Forensics[M]// Birmingham: Packt Publishing Ltd. 2015: 153-170.
- [7] iOS developer library[EB/OL]. <https://developer.apple.com/library/ios/navigation/#section=Frameworks&topic=AssetsLibrary>.
- [8] NYGARD S. Class-dump[EB/OL]. <http://stevenygard.com/projects/class-dump/>. 2013.
- [9] AGARWAL Y, HALL M. Protect my privacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing[C]//The International Conference on Mobile Systems, Application, and Services. c2013:97-110.
- [10] Dumpdecrypted[EB/OL]. <https://github.com/stefanesser/dumpdecrypted>.
- [11] FairPlay[EB/OL]. <https://en.wikipedia.org/wiki/FairPlay>.
- [12] Hopper[EB/OL]. <http://www.hopperapp.com/>. 2015.
- [13] FREEMAN J. Mobile substrate[EB/OL]. <http://www.cydiasubstrate.com>.
- [14] Tshark[EB/OL]. <https://www.wireshark.org/docs/man-pages/tshark.html>.

作者简介:



邢月秀 (1991-), 女, 安徽宣城人, 东南大学硕士生, 主要研究方向为隐私保护、手机取证。

胡爱群 (1964-), 男, 江苏如皋人, 博士, 东南大学教授、博士生导师, 主要研究方向为网络与信息安全、物理层安全技术。

王永剑 (1981-), 男, 山西长治人, 公安部第三研究所副研究员, 主要研究方向为网络与信息安全。

赵然 (1991-), 男, 江苏南京人, 东南大学硕士生, 主要研究方向为隐私保护、手机取证。