移动社交应用的用户隐私泄漏问题研究

程 瑶 应凌云 焦四辈 苏璞睿 冯登国

(中国科学院软件研究所可信计算与信息保障实验室 北京 100190)

摘 要 智能移动终端以其强大的处理能力和丰富的功能应用迅速得到普及,成为人们日常生活中存储和处理个人信息必不可少的工具.在众多的移动应用中,社交通信类应用致力于为人们提供便捷的日常通信服务,这类应用相比移动通信运营商提供的传统短消息服务更加经济实用,同时提供多媒体通信方式进一步增强用户的社交体验,从而迅速地被广泛接受.为了进一步巩固自身的用户群体,增加用户黏度,这类应用在其内部增添了一种称为"通讯录匹配"的功能.该功能能够向用户推荐其手机通讯录中已经注册过该应用的线下联系人为好友,从而帮助用户快速地将线下社交圈移植到应用线上.然而,用户在获得便利的同时也面临着潜在的隐私泄露风险.文中首次提出了一种独立于各移动智能平台的、能有效利用移动社交通信类应用的通讯录匹配功能实现大规模收集用户私人数据的方法,该方法能够收集到存储于目标应用服务器的用户个人资料,包括手机号码和虚拟应用账户资料以及两者之间的映射关系;其次,为了获取规模更大,内容更全面、更真实的用户资料,文本提出了基于多款社交通信类应用的跨应用整合分析方法以及针对不同应用来源的用户资料数据一致性与真实性分析;最后,在信息获取和分析方法的指导下,文中建立了利用上述漏洞的原型系统,进行了大规模数据实验,最终验证了上述方法的有效性和良好的可扩展性.

关键词 智能移动终端;社交通信类应用;隐私泄露;移动社交网络;隐私保护;智能手机

中图法分类号 TP309 **DOI 号** 10.3724/SP.J.1016.2014.00087

Research on User Privacy Leakage in Mobile Social Messaging Applications

CHENG Yao YING Ling-Yun JIAO Si-Bei SU Pu-Rui FENG Deng-Guo

(*Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing* 100190)

Abstract Due to their powerful processing capability and diverse equipped applications, smart mobile devices have become the rage to store and manage personal information in people's daily work and lives. This dominant prevalence to a large extent benefits from those various kinds of applications running on the mobile platform. Among them, a staple category of applications have devoted themselves to provide daily social communication service for regular users, which called social messaging applications. It offers users wonderful user experience and various ways of communication via multi-media, such as text, audios, pictures and videos. Comparing to the SMS and MMS, social messaging applications are more widely accepted for their fantastic social experience and economical manner. In order to aggregate user basis and increase their stickiness, social messaging applications incorporate a new functionality component called Address Book Matching which recommends registered user accounts from the address book in one's phone and facilitates the transplantation of users' social circle from offline to online. However, this novel feature brings not only convenience but also potential privacy leakage issues. This paper proposes a novel

收稿日期:2013-04-15;最终修改稿收到日期:2013-11-17.本课题得到国家"九七三"重点基础研究发展规划项目基金(2012CB315804)、国家自然科学基金(61073179)、国家自然科学基金——重大研究计划(91118006)及北京市自然科学基金(4122086)资助.**程 瑶**,女,1987 年生,博士研究生,主要研究方向为移动安全及隐私问题,包括系统、软件和网络中的安全及隐私问题.E-mail:chengyao@tca.iscas.ac.cn.**应凌云**,男,1982 年生,博士,助理研究员,主要研究方向为恶意代码分析与移动智能终端安全.**焦四辈**,男,1986 年生,博士研究生,主要研究方向为 Android 安全和恶意代码分析.**苏璞睿**,男,1976 年生,博士,副研究员,主要研究方向为恶意代码分析与防范.**冯登国**,男,1965 年生,博士,研究员,博士生导师,主要研究领域为密码学与信息安全.

platform-independent method to collect users' personal information in large scale, including their phone numbers and the corresponding application accounts, by means of abusing Address Book Matching. Besides, based on the user information we obtained, two approaches of further analysis are presented, i. e. single application analysis and cross application integration. In order to pursue more authentic user information, we propose the conformity and authentic analysis of user personal information gathered from different social messaging applications. Finally, on the basis of our collection and analysis approaches, we also build up a prototype system to leverage above mentioned vulnerability. The experiment results demonstrate the effectiveness of our method of taking advantage of Address Book Matching to collect user personal information from social messaging applications in large scale.

Keywords smart mobile devices; social messaging application; privacy leakage; mobile social networks; privacy protection; smartphone

1 引 言

近年来,风靡全球的智能移动终端逐渐成为人们日常生活必不可少的组成部分.相比传统移动终端,智能移动终端的优势和核心价值在于用户可按照个人喜好自行安装功能丰富的应用程序.而其中,基于即时通信的社交应用程序(简称"社交通信类应用",下同)充分发挥了智能终端的移动通信优势,极大满足了用户日常通讯需求,成为用户最喜爱的应用类别之一.这类应用主要是利用移动终端的蜂窝移动通信网络或者无线通信网络发送和接收数据,因此相对于传统的 SMS(Short Messaging Service)和 MMS(Multimedia Message Service)来说更加经济适用;同时,这类应用借助包括文字、语音、图片和视频等在内的多媒体信息,充分发挥了移动终端的即时通信特性,极大地丰富了用户的产品体验.相比移动通信运营商提供的传统通信服务,越来越多的用户更倾向于使用社交通信类应用进行日常通讯.在用户使用该类应用的过程中,为了方便现实世界的好友快速而准确地搜索到自己并添加好友,在注册账号时用户通常会提供真实的个人信息,例如姓名、年龄、性别,甚至住址和工作地点.此外,同一个用户可能使用多款社交通信类应用,这意味着不同应用服务器将存储同一用户的多份个人信息拷贝,这潜在地增加了用户个人信息被他人非法获取的风险.

随着社交通信类应用与移动运营商通信服务在移动终端上越来越紧密地结合,移动终端中存储的用户个人通讯录成为了各应用开发商觊觎已久的高价值用户个性化信息.通常情况下,手机通讯录联系人能够相对真实地反映用户的线下社交关系.如果用户需要通过社交通信类应用与自己的朋友进行通信,用户必须在应用中手动搜索目标好友,这种使用方式十分繁琐且不友好.更好的方案是让应用程序自动读取用户通讯录存储的联系人信息,并主动推荐那些正在使用该应用的联系人给用户.事实上,通过实际调研发现目前绝大多数社交通信类应用已经可以利用通讯录向用户推荐合适的联系人作为线上好友.这是用户期望已久的功能,同时也是社交通信类应用能够真正凝聚用户群,提高用户黏度的必经之路.然而,虽然从技术角度来说访问用户通讯录已没有任何困难,但从道德和法律角度来看,鉴于通讯录内容属于用户的私人信息,因此在实际中应用程序在使用用户通讯录之前必须首先获得来自用户的关于其通讯录信息被合法访问的权限许可,否则将会带来严重的后果.譬如,一款热门的国外社交通信类应用 Path,曾在未经用户允许的情况下擅自存储用户通讯录信息.虽然其初衷是为了利用通讯录为用户推荐高质量的好友,然而这种行为显然侵犯了用户的个人隐私,当时这起事件在业界掀起了轩然大波.幸运的是目前大部分的社交通信类应用都意识到合理利用用户信息的重要性,一切涉及访问用户通讯录的行为皆以获得用户的授权许可为前提.

但是,在获得了用户许可的情况下,被访问的隐私内容就安全了吗?正如前文所述,移动应用市场上大部分社交通信类应用已经具备了自动利用通讯录为用户推荐真实好友的功能(称为"通讯录匹配",下同),此功能允许用户上传其移动终端内的通讯录到应用服务器,随后从该服务器返回一组存在于通讯录中且已注册过该应用的联系人列表供用户选择并

添加为好友.例如,腾讯公司旗下的即时移动通信应用“微信”[①]内的“手机通讯录匹配”功能在启用之后会自动向用户推荐上述类型的好友,同时返回好友的基本信息.实际中,只要用户允许应用程序访问其通讯录并将应用程序账号与自己的手机号码绑定,该功能便被激活,开始向用户自动推荐好友.然而,这里存在一种安全隐患,即如果一个攻击者首先伪造自己的通讯录列表,然后激活此功能,这样应用服务器仍然会根据其伪造的通讯录中联系人的信息来向攻击者推荐与伪造信息对应的好友,这样攻击者就获得了这些伪造的联系人的应用账户资料,这些资料往往包含用户真实的个人信息,如用户名、性别和所在区域等.基于这种可能的攻击思路,本文从攻击者的角度提出一种利用通讯录匹配功能进行个人隐私刺探的方法以实现大规模用户个人信息的收集.

本文做出了以下贡献:

(1)本文首次提出通过移动终端社交通信类应用自动收集用户个人信息的方法,这种方法在用户的手机号码(现实世界身份)和应用程序账号及资料(网络虚拟身份)之间建立了映射关系,实验结果表明该方法将会导致上述映射信息及内容无条件地泄露给任意第三方;

(2)本文提出了一种基于多款社交通信类应用的跨应用整合分析方法,该方法通过横向整合和纵向渗透两种模式,能够获得数量更庞大、信息更全面的用户资料.此外,基于整合的用户数据,本文还提出了用户个人信息的一致性和真实性分析方法,以获取更真实的用户资料数据库;

(3)本文实现了基于上述漏洞利用方法的原型系统,并通过单应用分析充分验证了该方法的有效性,同时还通过一个具体攻击实例展示了跨应用整合的数据分析过程及攻击效果,实验结果表明,相比单应用分析,跨应用整合方法能够获取更全面、更真实的用户个人信息;

(4)最后,在分析该攻击实施过程的基础上确定了实施攻击所必备的前提条件,并具有针对性地提出相应的防御策略供应用开发者参考.这些策略能有效避免用户虚拟网络身份与真实世界身份映射关系的建立,进而防止攻击者通过用户的网络身份获取其真实身份的个人信息.

本文第 2 节介绍拟解决的问题及涉及到的相关概念;第 3 节对本文提出的漏洞利用方法及原形系统进行深入地阐述,同时展现具体的实现细节;实验和数据分析部分在第 4 节展示;第 5 节分析攻击形成的基本要素和前提,进而为应用开发者提出一系列可行的防御策略;第 6 节讨论与本文相关的其它研究工作;最后第 7 节总结全文.

2　问题描述

本节主要介绍当下社交通信类应用中广泛运用的通讯录匹配功能,并就其潜在的安全问题进行讨论.

2.1　通讯录匹配功能

为了更快速地获取用户群,提升用户的黏度,如何将用户真实生活中的社交关系迁移到应用平台上是社交通信类应用所要考虑的一个重要环节.通讯录匹配功能极大地满足了开发者的这一需求,该功能逐渐成为了社交通信类应用的标配.

通讯录匹配功能的使用前提有两点:首先,需要将用户的手机号码和在应用中注册的用户账户进行绑定;其次,需要用户显式地允许应用程序上传通讯录或者通讯录特征摘要信息.在以上两个前提都满足后,应用程序将该用户的通讯录上传至应用服务端,随即返回已经注册该应用并绑定了应用账户与手机号码的联系人给该用户供其浏览或添加为好友.这种机制一方面极大地方便了用户与好友之间的找寻与交流,另一方面迅速扩大了用户在新应用中的社交圈,提高了用户对于新应用的依赖度.该功能的使用基于一个潜在的默认假设,即通讯录中的联系人信息能够真实反映用户现实世界中的社交关系.应用程序默认假设用户与其联系人之间存在一定的信任度,认为向用户推送其联系人账户信息将不会构成潜在的安全风险.然而,事实往往不是如此.本文经过分析发现通讯录匹配功能在给用户带来便利的同时,也带来了潜在的用户隐私泄露隐患.

2.2　恶意利用通讯录匹配功能形式化描述

首先,用户 u 的通讯录可以表示为联系人的集合:

$C_u=\{c_i \mid i\in[1,N];N$ 为 u 的联系人总数$\}$,

其中,c_i 表示用户 u 的通讯录中第 i 个联系人,c_i 是一个二元组$(name_c, pn_c)$,$name_c$ 和 pn_c 分别表示联系人 c 的姓名和手机号码,pn_c 通常以明文的形式存储在本地通讯录中,且能唯一确定 c_i.

应用服务器的数据库中存储了所有的用户账户信息,构成整个用户账户集合 ADB (Account Database):

① 微信. http://weixin.qq.com

$$ADB = \{Account_u^{app}\}.$$

其中，$Account_u^{app}$ 表示已注册应用程序 app 的用户 u 的帐户：

$$Account_u^{app} = \{ID_u^{app}, PN_u, Profile_u^{app}\}.$$

ID_u^{app} 是应用程序 app 中用户 u 的唯一标识符，PN_u 表示与该账户绑定的手机号码的特征信息，$Profile_u^{app}$ 表示应用程序 app 中用户 u 的所有个人信息.

当用户激活通讯录匹配功能后，形成如下的实体绑定：

$$Binding(Account_u^{app}, Hash(pn_u)).$$

该绑定涉及两类实体，即用户的账户 $Account_u^{app}$ 以及用户手机号码的摘要值. 其中，$Hash$ 表示摘要函数，如 MD5. 对手机号码的摘要处理是应用程序出于隐私考虑，避免手机号码直接存储在服务器上的一种保护手段. 实际中，该绑定将 $Hash(pn_u)$ 的值赋予 $Account_u^{app}$ 的 PN_u.

已绑定手机号码到应用程序 app 的用户构成了如下集合：

$$BIND_{app} = \{Account_u^{app} \mid Binding(Account_u^{app}, Hash(pn_u))\}.$$

基于对上述问题的描述，接下来将对滥用通讯录匹配功能进行用户资料获取的方法进行阐述. 值得注意的是，该方法不针对任何特定的移动终端系统，对任何移动平台，只要存在配备通讯录匹配功能的社交通信类应用，本方法都将适用.

首先，攻击者 f 伪造一个通讯录 C_f，其中每条联系人都由姓名和手机号码组成. 由于手机号码的号码段具有固定的格式，伪造的通讯录中手机号码可以自由选择，攻击者可以选择感兴趣的号码集合、随机号码集合或者特定的号码段. 姓名可以与手机号码无关，选择可读性强且唯一的标识即可. 接下来，将 C_f 上传到应用程序服务器，随后会收到从服务器返回的已绑定该应用的联系人账户信息及其资料，从而构成了目标联系人集合 R_{app}，完成了目标用户信息的收集：

$$R_{app} = \{Account_u^{app} \mid Account_u^{app} \in BIND_{app}, Account_u^{app}.PN_u \in \{Hash(c_i.pn_u) \mid c_i \in C_f\}\}.$$

此外，如上文所述，出于隐私保护的目的，部分应用程序在服务器端存储手机号码的特征信息作为匹配标识，这种隐私保护机制并不能影响本文所述方法的实施. 因为本文提出的方法利用了应用程序中的正常功能，完全从正常用户的使用角度进行攻击，在这种场景中，本地通讯录中的手机号码是明文存储，而不需要分析应用在服务器是否以特征信息形式存储手机号码.

2.3 威胁分析

账户是用户在网络空间中的唯一可见标识. 用户通常只能通过他人的账户和相应的资料来判断对方的身份. 多数情况下，用户习惯于相信那些资料详细、完整且合理的账户. 虽然这些账户看似可信，但当攻击者能够通过某种方法大规模获取用户资料并利用窃取的用户资料进行伪装诈骗时，这种常识将会导致用户严重的错误判断，从而引发网络犯罪，特别是网络欺诈.

本文提出的利用社交通信类应用通讯录匹配功能获取用户资料的方法具有以下被攻击者利用的潜在威胁：

(1) 攻击者可以通过该方法大量获取用户的个人资料和手机号码. 由于手机号码通常能唯一确定用户的真实身份，在结合用户资料后攻击者对手机号码的拥有者具有更全面的了解，因此本方法可以作为进一步攻击的前期准备工作，例如作为社会工程学攻击的前期信息收集阶段；

(2) 通过本方法，网络中的恶意用户，特别是垃圾短信制作者通过向社交通信类应用服务器提交一组指定的目标手机号码，就可以确定选定的手机号码是否处于使用状态. 这样恶意用户能够收集大量的活跃手机号码，进而统计收集到的用户资料信息，最后实现根据性别、地域和兴趣等不同偏好进行广告的定向投放，从而扩大了广告投放的有效受众规模；

(3) 当用户注册多个应用程序后，攻击者可以通过本方法从不同的应用程序中获取多份同一用户的不同资料，结合本文的整合方法，攻击者可以获取该用户更加全面的资料信息，例如包括学校、院系和工作单位等，基于这些资料攻击者能够开展进一步的网络诈骗攻击；

(4) 攻击者也可以通过获取的用户资料信息实施身份盗窃，以克隆的方式创建虚假的用户身份. 克隆身份攻击是指在相同或者不同的应用中攻击者根据已有的目标用户的资料信息克隆出一个和目标用户完全一致的虚假账户来开展进一步的攻击. 对于攻击者来说，获得的信息越全面，创建的克隆身份就越真实，攻击效果就越明显.

3 方法与实现

本节主要介绍利用移动社交通信类应用获取用

户资料的方法及流程，然后基于此流程进一步阐述原形系统的整体架构和具体实现细节.

3.1　信息获取与整合分析

利用社交通信类应用的通讯录匹配功能获取用户手机号码及账户信息的整体流程如图 1 所示. 首先是伪造候选通讯录，候选通讯录的每个条目包括目标联系人的唯一标识符和手机号码；然后，将伪造的通讯录通过目标应用的通讯录匹配功能提交，这时，应用将通过网络将通讯录信息以原文或者摘要信息的形式上传到应用服务器中；服务器随后会返回所有与该通讯录中的手机号码绑定的用户账户及相应的个人资料，这些返回的信息将形成目标用户的初步资料；最后，基于返回的资料作进一步的数据整合与分析，形成了更完整、更真实的目标用户信息，称为用户资料数据库. 其中，获取目标用户应用账户资料的具体实现方法将在 3.2 节中介绍.

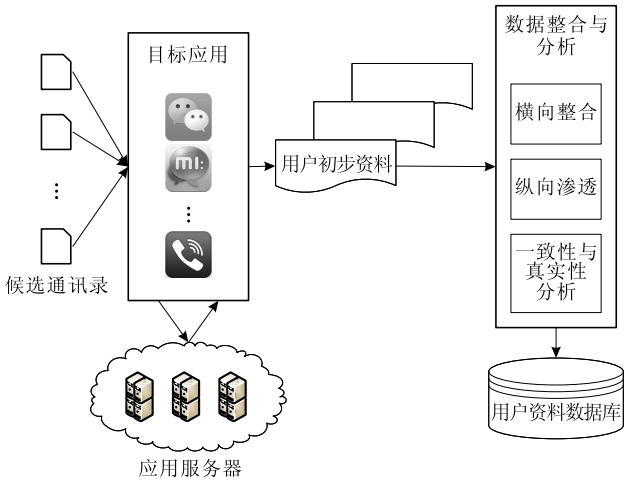

图 1　逻辑流程图

上述流程的最后一步涉及到用户个人资料的整合和分析. 其中，整合主要是针对已获得同一用户在多个不同应用中多份个人资料的情况. 实际中，每种应用平台要求用户填写的账户注册信息或多或少地存在差异和补充，因此攻击者通过对来自多个应用平台的用户资料进行有效整合，往往能够获得关于目标用户更加全面的个人信息. 用户个人资料的整合方法主要分为两类：即横向整合和纵向渗透. 横向整合主要针对从多个非同源应用平台（出自不同开发商）获取用户资料的情况，而纵向渗透关注从多个同源应用平台（出自相同开发商）获取到用户资料的情况. 此外，出于防范意识或者其它原因，用户在应用中提交的个人注册信息不一定总是反映了用户的真实情况，因此本文也提出了一种用于确定用户资料真实度的一致性与真实性分析方法.

3.1.1　横向整合

目前移动应用市场中不同的社交通信类应用在完成通讯录匹配后返回的用户资料的内容和格式都不尽相同，这取决于各种应用平台的用户定位和对用户资料的需求. 因此，如果攻击者通过通讯录匹配功能获得了同一用户在不同应用平台注册的个人账户信息，再加以比对和聚合，就能得到更加完整的用户资料. 这便是横向整合的基本思想. 横向整合按照目的不同可以分为两类：广并集合和深交集合.

由于不同的应用程序拥有各自不同的用户群体，广并集合通过结合从多个应用平台获取的不同用户账户来获得更多数量的用户账户，这些账户构成了集合 $BroadUnion$. 为简单起见，本文以两个应用程序为例进行说明：

$$BroadUnion_{app_1 \cup app_2} = \{Account_u^{app} \mid Account_u^{app} \in R_{app_1} \cup R_{app_2}\}.$$

深交集合旨在挖掘同一用户更完整和全面的个人信息，因此深交数据处理主要针对同一个用户可能活跃在多个应用程序的情况. 这里 PN_u 作为用户 u 在不同应用平台上的唯一标识符，每一个 PN_u 将确定真实世界中的一个用户身份，所以通过 PN_u 能够连接同一用户身份在不同应用平台中的多份用户账号及附带资料. 深交集合 $DeepIntersection$ 表示为

$$DeepIntersection_{app_1 \cap app_2} = \{Account_u^{app_1} \cup Account_u^{app_2} \mid Account_u^{app_1} \in R_{app_1}, Account_u^{app_2} \in R_{app_2}, Account_u^{app_1}.PN_u = Account_u^{app_2}.PN_u\}.$$

其中，

$$Account_u^{app_1} \cup Account_u^{app_2} = \{ID_u^{app_1}, Profile_u^{app_1}, ID_u^{app_2}, Profile_u^{app_2}, PN_u\}.$$

3.1.2　纵向渗透

目前，许多大型应用开发商旗下都拥有多款应用产品，并在这些应用之间共享用户登录 ID. 这样做不但方便用户管理自己的各种应用账号，更重要的是方便开发商整合用户资源从而扩大其每款应用的用户覆盖率. 另外，现在诸多类似于 Facebook 和腾讯的在线社交服务提供商逐步对第 3 方站点开放了 ID 登录验证服务功能. 这意味着大量的应用程序将会共享相同的账户登录 ID 数据. 这里，将针对共享相同账户登录 ID 的多个应用程序间实施的渗透叫做纵向渗透，纵向渗透往往需要人工参与. 相对于横向整合，纵向渗透可以提供更为全面、准确的个人信息. 本文将在 4.2.2 节呈现一个完整的纵向渗透的案例，并对该案例进行详细的阐述和分析.

3.1.3 一致性与真实性分析

市场上不同款社交通信类应用各自侧重于满足用户不同的社交需求,有的应用定位于促进用户与朋友间的通信和交流,例如推荐通讯录联系人为待添加好友;有的应用则定位于帮助用户发掘新朋友,例如基于相似的兴趣爱好或相近的地理位置为用户推荐新朋友.不同的用户定位决定了用户在填写账户注册信息时呈现出不同的行为习惯.在使用定位于促进朋友间通信的应用时,为了方便被朋友找到并添加自己的应用账户为好友,用户通常倾向于填写自己真实的个人信息,同时用户也会潜意识地认为这些信息会被自己的联系人朋友看到.而在另一类定位于推荐和挖掘新朋友的应用中,用户会出于自身防卫意识而填写不完全真实的信息.这种情况下,在分析来自于不同社交通信类应用的数据时,用户个人信息数据的一致性和真实性是需要关注的重点之一,本文针对该问题提出了一致性与真实性分析方法.

这里的一致性和真实性是指来自多个应用平台的用户资料中相同字段的内容是否统一,如果不统一,则意味着该字段的某些版本可能不完全为真实情况.一致性与真实性分析方法主要针对用户来自不同应用平台的个人信息中共有的字段进行分析,其核心思想是比较同一用户的多份个人信息的差异,然后根据不同的信息差别对字段真实情况进行不同的判断.具体来说,对于非类别信息字段(如姓名、住址等),如果这些字段的内容趋于一致(内容完全匹配)或相似(内容部分匹配),则认为该字段的内容很有可能反映了用户的真实信息.注意,对于用户在任何应用平台内都使用一致的虚假信息来注册账户的情况,则任何一种使用用户注册信息进行一致性和真实性分析的方法都将无法判断信息真实与否.另一方面,对于类别信息(如性别),如果出现不一致的情况,则字段采用"胜者全取"机制,即选择在不同应用平台中出现频率最多的值作为最终确定的真实值.特别地,针对用户"用户名"字段的处理,由于中文姓名具有固定的组成模式,在处理中文姓名时本文将运用模式匹配技术对中文姓名的真实性进行分析.

3.2 系统架构与实现

本文实现了利用上述漏洞实施大规模获取用户信息的原型系统,该系统分为3个模块,即通讯录伪造模块、信息提取模块和数据分析模块.通讯录伪造模块负责自动产生候选手机号码,并以标准通讯录格式输出.通常情况下,攻击者可以枚举若干号码段.通讯录规模的设定值得关注,本文提出的方法在应用服务器没有限定响应通讯录匹配请求数量的情况下具有很好的可扩展性,但在实际实验过程中需要充分考虑移动终端的处理能力,所以选择一个合适的通讯录规模是实验成功的前提.信息提取模块的功能是将应用程序返回的一系列用户账户及其个人资料从移动终端系统中提取出来.该模块是系统的核心,其难点在于需要考虑终端系统的资源保护机制,例如沙盒、权限机制等.最后,数据分析模块负责对从信息提取模块提取出的信息进行分析处理,具体的处理过程和实验结果将在第4节给出.

出于安全考虑,移动终端中的应用数据受系统保护且不允许其它非授权应用或者进程进行访问,例如,iOS系统的沙盒机制和Android系统基于权限的严格访问控制机制.所以,具体到不同的平台,信息提取的方法随着平台特性的不同而有所不同.值得注意的是,本文提出的利用通讯录匹配功能获取用户个人信息的方法将适用于所有具备该功能的社交通信类应用,而与具体的移动系统平台无关.

本文实现的原型系统基于时下流行的智能终端系统Android系统,作此选择主要出于以下3点考虑:首先,近年来搭载Android系统的移动设备的销量增长迅速,市场占有率大幅提升,将Android系统作为目标系统具有更广泛的应用价值;其次,由于Android设备被广泛接纳,各个应用开发商都纷纷推出了各自应用的Android版本,选择Android系统不会出现应用集合覆盖面不足的情况;最后更重要的是,Android系统作为一个开源软件项目,极大方便了本文原型系统的设计和实现.

在Android系统中,每一个应用都运行在单独的Dalvik虚拟机中,这种机制是为了保证系统中运行的各应用之间的隔离性.每一个Android应用的组件,例如Content Provider,都具有被自身或其它应用调用的入口.然而,Android系统实现的权限机制有效地防止了肆意调用其它应用组件的情况,这同时也意味着寄宿在目标应用中的数据将无法从应用程序外部直接访问,从而给本文的用户资料信息提取工作带来巨大的挑战.

为了提供一种可扩展性良好且通用的解决方案,即尽量避免针对不同的社交通信类应用程序需要实现不同用户资料提取方案,本文采用了一种独立于具体应用的解决方案.Android架构拥有一套可重用的应用程序访问接口(Android API),该接

口封装了一系列的 Linux 系统级调用. 由于任何基于 Android 系统的应用在开发时都会涉及到 Android API 的调用，因此本文将采用动态监控 API 的方法来提取寄宿于应用当中的用户资料数据. 如图 2 所示，信息提取模块位于 Android 框架层. 通过通讯录匹配从目标应用服务器返回的用户账户及个人资料是实验所需的目标数据，当目标应用将这些数据显示到屏幕时，需要调用与在设备屏幕上显示文本数据相关的 Android API，如 setText(). 所以，通过修改 Android 系统代码，监控这些与显示数据相关的 Android API 调用及其参数并记录所有经由这些 API 处理的信息，就能够记录下所有被显示到设备屏幕上的用户数据. 最后，通过重放一系列“sendevent”ADB 命令模拟用户的浏览动作，以保证每个返回的用户账户及个人账户信息都被信息提取模块过滤并保存，从而实现了整个用户资料数据提取过程的自动化.

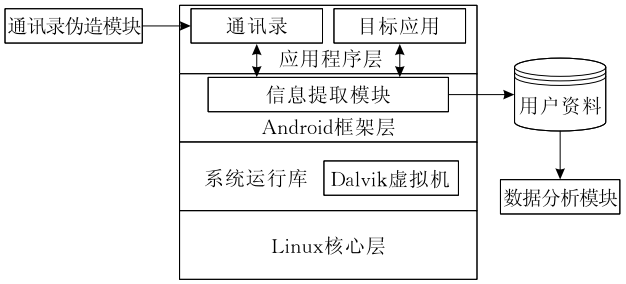

图 2　系统架构图

4　实验与数据分析

本节将详细阐述利用社交通信类应用通讯录匹配功能获取用户应用账户数据的实验过程和用户应用账户数据的分析结果. 由于实验可能会涉及到用户隐私问题，遂在此做出必要的隐私尊重与保护说明.

相对于其它的方法，从真实世界中获得用户数据进行实验能够最有效、最直接地验证本文提出的思想和方法. 然而这些真实的用户数据在一定程度上也触及到了用户隐私，因此如何有效地保护实验获取的真实的用户数据是开展此项工作前必须考虑的问题. 事实上，在此前的研究项目中，不少研究人员也遇到过同样的问题[1-2]，人们普遍认为基于真实世界数据的实验能够提供更具说服力的实验结果. 针对实验过程中涉及到的用户隐私信息，本文将采用必要措施，最大限度地保证这些信息仅被用于本工作的实验过程中，具体来说：第一，禁止用户资料信息被扩散至实验设备以外，同时也不提供任何信息给任何不相关的个人或机构；第二，除本文展示的实验内容和数据外，作者将不会使用获取的用户资料信息实施进一步的渗透攻击.

4.1　单应用分析

单应用分析主要针对单独一款社交通信类应用的通讯录匹配功能利用进行分析. 实验环境部署在 Ubuntu 12.04 系统中，应用运行环境为 Android 模拟器，Android 版本 4.2. 这里选取市场上非常流行的一款社交通信类应用微信(版本 4.2，Android)作为分析对象. 微信是由腾讯公司开发的一款跨平台通信工具，能够支持文字、图片、语音短信和视频等多媒体信息，并具有各种移动终端系统平台下的版本，包括 iOS，Android，Windows Phone，Symbian 和 BlackBerry 等. 据统计，截至 2013 年 1 月微信用户数量已经超过了 3 亿.

根据 3.1 节阐述的内容，单应用分析的第一步是伪造候选通讯录. 候选通讯录中主要包含目标用户的手机号码和姓名，其中姓名仅用于唯一标识每一个目标用户，所以这一步骤的重点在于选取合适的手机号码. 中国大陆的手机号码是由国际电话区位码＋86 和一串 11 位的数字组成. 实验中由通讯录伪造模块生成的手机号码段为：＋861521063 至＋861521070 和＋861521098 至＋861521099.

接下来，整个实验在配置了 3.2 节所述的原型系统和微信(需预先注册一个实验用的微信账号)的模拟器上完成. 首先，在微信中绑定测试手机号码与微信实验账号，并开启通讯录匹配功能，然后导入伪造通讯录文件. 微信的通讯录匹配功能会将伪造通讯录上传到应用服务器，随后便开始向用户推荐一系列微信用户账户以及对应的账户信息，这些账户已经与上传的伪造通讯录中的手机号码进行过绑定. 在这些推荐账号呈现在微信程序界面时，系统中的信息提取模块已记录下所有返回的用户信息，形成了用户初步资料.

用户初步资料包括与伪造通讯录中手机号码绑定的所有微信账户，这些账户与对应的手机号码形成了一一映射. 微信账户个人资料包含 5 个字段，即微信账户 ID、用户名、性别、所在地区和个性签名. 通常用户名字段已经能够让好友确定账户所属者的身份. 用户名通常会由与用户真实姓名有关的字符组成，甚至就是用户真实姓名本身. 用户使用自己真实姓名作为用户名是为了方便被好友查询和识别，而通常不会意识到这些信息是否会被陌生人获取.

表 1 微信返回用户帐户的数量分布

手机号码段	返回账户数量	手机号码段	返回账户数量
+861521063	1531	+861521068	1449
+861521064	1469	+861521069	1486
+861521065	1390	+861521070	0
+861521066	1463	+861521098	1936
+861521067	1603	+861521099	1852

本文伪造的通讯录总共包含 100 000 个手机号码，通过微信的通讯录匹配功能成功返回了 14 179 个账户及对应的个人信息. 可见实验中的用户渗透率为 14.18%，这一数字与实验同期(2012 年 11 月)的中国互联网信息中心(CNNIC)公布的网民微信使用调查结果基本一致(13.73%)①，进一步说明了本文的样本具有较好的代表性. 返回的微信账户数量在对应的手机号码段中的分布情况如表 1 所示. 注意到 +861521070 号码段没有任何对应账号返回. 经过后期调查，能够确定该号码段目前处于空闲状态，尚未启用.

下面对返回的微信账户资料的各个字段进行分析. 从返回的用户数据可以看到并不是所有用户都完整地填写了账户个人资料. 对于一个账户，首先能确定的是账户所属者的手机号码，此外由于微信 ID 字段和用户名字段在用户注册时都是必填字段，因此这两个字段的完整率为 100%，另外 3 项可选填字段(性别、所在地区和个性签名)的完整度统计如表 2 所示. 可以看出全部 5 项资料都填写完整的用户比例超过了一半(55.13%)，而不填写任何可选资料的用户占总人数的 12.43%. 需要注意的是，即使用户不填写任何可选资料，攻击者也能够获取其微信 ID、用户名和手机号码等关键信息.

表 2 可选资料字段完整度

可选资料字段数目	填写的用户数量	完整度比例/%
0	1762	12.43
1	437	3.08
2	4163	29.36
3	7817	55.13

对于每个可选字段，本文也进一步确定了这些字段的填写情况(表 3). 可以看到，性别和地区的填写比例都在 85% 左右，而个性签名的完整度为 57.86%，低于其它 2 项比例值.

表 3 可选资料内容完整度

可选资料字段	填写的用户数量	完整度比例/%
性别	12066	85.10
地区	11944	84.24
个性签名	8204	57.86

4.2 跨应用分析

跨应用分析主要针对多款社交通信类应用进行分析，其目的在于收集同一用户在不同应用平台注册的账户资料. 除微信外，实验中选择另一款流行的社交通信类应用“米聊”②作为分析对象. 米聊是小米科技公司出品的一款跨平台的通信工具. 与微信类似，米聊也支持市场上大部分主流手机操作系统，并支持文字、语音和图片等多媒体通信. 据米聊官方公布的数据显示，截至 2013 年 1 月，米聊注册用户已经达到 2300 万，是目前国内最大的社交通信类应用之一. 米聊作为基于手机通讯录的强联系社交入口，同样具有推荐用户通讯录好友功能. 本文使用米聊 Android 版本 5.0.565 进行实验. 米聊的通讯录好友推荐机制与微信的略有不同，米聊实现了应用程序与通讯录间的无缝对接，在通讯录匹配返回推荐好友后，米聊一方面会直接发送好友请求至被推荐的好友账户，另一方面会将这些好友账户直接添加至用户的“已添加好友”列表.

4.2.1 横向整合

根据 4.1 节中表 1 显示的各个手机号码段绑定微信账户的情况，实验中选择返回账户数量最多的 3 个号码段构造候选通讯录，分别为 +861521098，+861521099 和 +861521067，共覆盖 30 000 个手机号码.

随后利用米聊的通讯录匹配漏洞收集米聊用户的账户资料，最终成功返回了 955 个与候选通讯录中手机号码绑定的米聊账户(表 4). 通过观察收集到的米聊账户，发现米聊用户的账户资料比微信更加丰富，包括的字段有名称、性别、生日、照片链接、地区、学校和公司等. 几乎所有返回的用户都至少拥有一个照片 URL 链接. 另外，近 30% 的用户填写了所在学校和公司信息，这些信息对于攻击者来说都具有很好的利用价值.

表 4 米聊返回用户帐户的数量分布

手机号码段	返回用户数量
+861521098	360
+861521099	370
+861521067	225

如 3.1.1 节所述，横向整合可以分为两类，即广并和深交. 其中，跨应用的广并能够收集更多数量的不同应用下的目标用户账户，例如，对比 4.1 节针对微信的单应用分析结果，实验中通过米聊的通讯录

① 微信. 移动互联网时代的新宠儿. http://www.cnnic.cn/hlwfzyj/fxszl/fxswz/201211/t20121112_37173.htm

② 米聊. http://www.miliao.com/

匹配漏洞又收集到 348 个新增的手机号码和对应的米聊账户. 通过跨应用的广并操作，实验一共获得了两个应用中的 5739 个用户账户和对应的手机号码. 另一方面，就深交而言，通过对比从两个应用收集到的账户信息，发现了 607 个在两个应用中使用相同的手机号码绑定的账户，然后通过整合这些用户在两个应用中的用户资料字段，便获得了这些用户更全面的个人信息.

4.2.2　纵向渗透

纵向渗透主要应用于多个应用程序或者服务之间共享相同用户登录数据库的场景. 例如，作为一款广泛流行的即时通讯工具，QQ 软件和微信平台都是腾讯公司旗下的社交应用产品. 鉴于共属同一个开发商，它们共享相同的用户登录认证数据库，表现出用户能够通过自己的 QQ 账号和密码来登录微信平台. 类似的，这种关系同样存在于小米科技公司旗下的小米手机社区官方论坛和米聊平台之间. 跨平台账户共享一方面简化了用户对各种应用账号的管理，另一方面也方便了开发商管理自己的用户，同时还帮助开发商拓展其老用户到新的应用平台. 纵向渗透攻击就利用了跨平台账户共享特性，目的在于通过已经收集的关于某个应用的用户账户信息，来进一步发掘另一款共享相同账户登录数据库的应用中对应的用户账户信息，以获得对用户更丰富的了解.

下面通过一个具体的案例阐述纵向渗透攻击的原理和步骤. 首先，从深交集合中选取一条记录，每条记录包含每个用户在不同应用(微信和米聊)中填写的账户资料，这里选取的用户的手机号码为“152＊＊＊＊＊＊56”，微信 ID 为“q76＊＊＊＊74”，米聊 ID 为“47＊＊＊49”，星号字符表示被隐去的部分数字. 考虑到微信平台和 QQ 软件的同源性，可以推断该用户微信 ID 号的后 8 位数字很可能是该用户的 QQ 号码. 为了验证此推测，将此 8 位数字串作为查询输入，通过 QQ 软件的账号查询功能进行查询. 查询结果成功返回了一位匹配用户，通过“查看个人资料”面板可以了解到该账号的主人是一位 1983 年 6 月 20 日出生的女性，故乡是黑龙江省大庆市，现居北京市，姓名备注一栏填写的内容是“LMX”，由于从米聊获得的用户名拼音首字母缩写也为“LMX”，推测该备注为用户姓名的拼音缩写. 此外，通过访问该用户在个人资料中提供的个人主页网址，发现该用户主页上有 24 篇日志，377 张照片. 从这些信息中进一步了解到该用户的职位、工作的单位和地址，以及个人的兴趣爱好等信息. 最终，成功在相册中发现了与其米聊账户头像相同的照片，从而证实了上述阐述的所有信息都属于同一个用户. 另外，通过 QQ 软件和小米手机社区官方论坛提供的找回账户密码服务，也可以进一步确定该账户绑定的手机号码的部分特征(找回密码服务页面通常会显示目标用户手机号码的部分确定数字，其它部分则用星号字符代替)，因此更加确定了该用户的存在. 这样，通过一次完整的纵向渗透攻击就成功地获取了目标用户在多个应用平台中留下的更加全面的个人资料. 上述纵向渗透的具体流程及获取信息内容如图 3 所示.

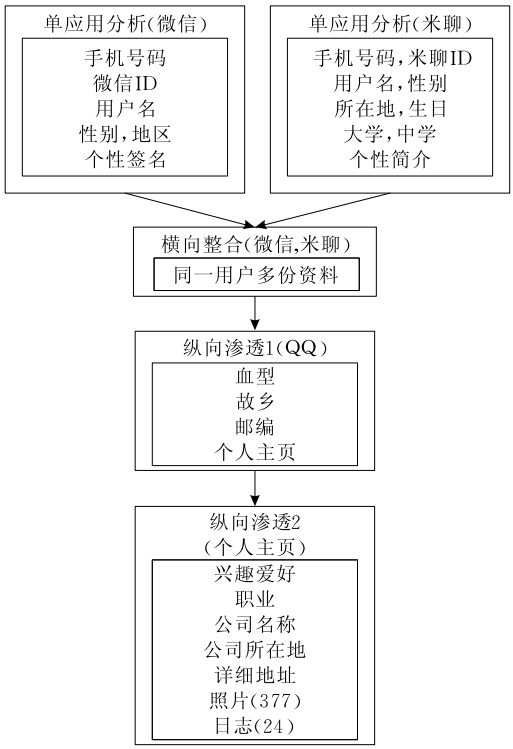

图 3　纵向渗透示例

可以看出，首先针对特定的应用平台利用本文提出的通讯录匹配漏洞，攻击者可以获取目标用户的手机号码与其应用账号间的映射关系，然后通过跨应用的横向整合继而获得了目标用户多个应用账户及资料，最后基于同源应用平台的纵向渗透攻击为攻击者提供了关于目标用户更为详细和全面的个人资料，这种攻击过程呈现出逐步深入递进的特性. 虽然纵向渗透目前主要依靠人工完成，但从纵向渗透的效果来看，人工分析是必要且很有价值的. 此外，人工分析的步骤很容易标准化，这将进一步提高纵向渗透攻击的实施效率.

另外，在本文测试的其它纵向渗透案例中，还遇

到一种情况是在访问目标用户的个人主页时被提示回答用户预设的主页访问控制问题,例如“我的真实姓名是?”,这种问题主要用于确保访问人群为自己相识的人(一般假设与自己相识的人会知道自己的姓名).而这时通过 4.2.1 节所述的横向整合多个应用提供的更加详细和确定的用户信息可以提供极大的帮助,从而使得陌生人能够突破预设主页问答的限制,给用户造成了极大的隐私泄露.

4.3 一致性与真实性分析

在跨应用分析中,不同的应用程序返回的用户账户信息常常会拥有共同的字段,例如用户姓名和性别,而针对同一用户来说,其在不同应用注册的个人信息可能会不一致,甚至出现相互矛盾.这种情况的出现说明用户提供的部分信息不能反映用户的真实情况,因此需要通过某种方法辨别出用户真实的个人资料(即有利用价值的资料),本小节根据 3.1.3 中的方法进行跨应用环境下的用户信息一致性分析与真实性分析方法进行相关实验分析,下面通过两个典型的实例来进行阐述.

第 1 个实例针对用户账户中的非类别字段进行一致性与真实性分析,这里以“用户名”为例.在一致性方面,根据 4.2 节阐述的跨应用分析的结果,可以发现 34.02%的用户在两个应用程序中拥有的用户名是相近的(共同字符至少占较短用户名的 α 以上,这里 α 取 0.5).基于两个应用中存在的用户名不一致情况,本文的用户名真实性分析采用模式匹配的方法来确定用户名是否可能为用户真实的姓名.典型的中文姓名是由一到两个汉字构成“姓”加上一到两个汉字构成的“名”组成.本文借助 2007 年中华人民共和国公安部对全国户籍人口姓名的最新统计来判别应用程序账户姓名的真实程度.该统计列出了全国范围内排名前一百的最常用的姓氏,这一百姓氏总人口占全国人口的 84.77%.实验中假设如果应用程序的用户名是以上述姓氏列表中的特定姓氏开头且后跟有一到两个中文汉字一起组成的,则认为该用户名接近用户的真实姓名.接下来分别针对微信平台收集的数据集、米聊平台收集的数据集和深交数据集进行了用户名真实性分析,结果如图 4 所示.可以看出,米聊数据集中用户名可能是真实姓名的比例为 53.61%,远超过微信平台的 31.24%.这是因为米聊平台的产品定位是为用户提供一个基于通讯录的社交入口.在注册米聊账户时,用户被建议尽可能真实地填写个人信息以方便联系好友.在这种应用定位的引导下,用户也逐渐把米聊当成了传统短信息通讯的替代,因此更多的用户在注册米聊账号的过程中往往会使用自己的真实姓名作为用户名,从而导致填写真实用户名的用户比例大大增加.此外,前文提到米聊致力于在用户的应用线上社交圈和基于通讯录的真实社会关系网络之间建立一个无缝的连接,这种战略定位使得用户更加信任米聊中的好友,同时也自然以为自己的个人资料将会被真实世界中的朋友浏览,这也是米聊数据集中用户名更接近用户的真实姓名的另一个原因.在多个社交应用中注册的用户往往是在线社交的活跃分子,因此他们更愿意把自己的真实信息公布出来,从而导致了深交集合中姓名的真实度高达 60.96%,这个结果表明在同时使用微信和米聊两种平台的用户当中,有超过一半的用户至少在一个应用中使用了接近真实姓名的用户名进行账号的注册.

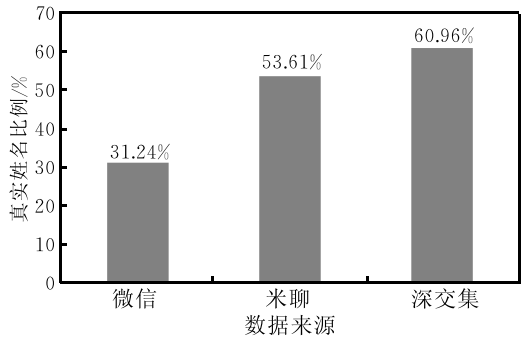

图 4 不同数据集的姓名真实度比较

第 2 个实例针对用户账户中的类别字段进行一致性和真实性分析,这里以“性别”字段为例.实验过程中发现在同时注册了微信账户和米聊账户的用户当中,性别不一致的情况非常严重,这种现象同样也存在于不同类型的社交网络中[1].一个用户的真实性别是唯一的,然而其网络性别可以不同.实验进一步对这种不一致的情况进行了统计分析,如表 5 所示.其中,“米聊男,微信女”的比例达到了不一致情况总数的 66.37%.为了进一步确定用户的真实性别,在考虑多个应用程序时,可以采用“胜者全取”的策略,即选择大多数一致的内容作为该字段的内容,然而本文的实验只涉及了两个应用,故无法采取该方法.但通过观察可以发现相对于米聊来说,微信中的男女用户比例严重失衡,结合前述用户名真实性分析的结果,可以看出用户在米聊上提供个人信息资料更接近于真实情况.同时实验中还发现用户姓名的极性与米聊性别字段的内容基本一致,从而进一步佐证了上述观点.

表 5　微信与米聊用户性别对比统计

微信 \ 米聊	男	女
男	93	114
女	225	141

一致性和真实性分析是信息整合分析中必不可少的一部分，只有分析确定有价值的真实数据才能体现出信息整合的意义. 在跨应用分析中通过收集和整合多个应用账户的个人资料，利用一致性和真实性分析将可以获得更为真实的用户个人信息. 真实的信息具有极高的利用价值，无论对于垃圾短信制造者发放目标广告，还是在线欺诈攻击的实施等行为都是非常宝贵的前期资料.

4.4　其它应用实验

除微信和米聊这两款拥有国内用户基数最大的应用之外，本文还针对其它 6 款社交通信类应用进行了分析，包括网络通话应用 Viber、免费聊天应用 Kakao Talk、Kik Messenger、有信、开心网手机应用和人人网手机应用等. 实验结果如表 6 所示.

表 6　其它应用实验结果

应用名称	实验结果
Viber	通讯录匹配返回率 7.9%，返回资料包括手机号码、显示名称和用户头像等
Kakao Talk	通讯录匹配返回 0.3%，返回资料包括头像、手机号码、个人签名等
Kik Messenger	无返回结果
有信	通讯录匹配返回率 7%，自动添加已有账户到有信好友列表，对大通讯录处理敏感，会导致应用崩溃
开心网应用	返回资料有用户头像、姓名、性别、居住地和大学等，存在用户设置访问限制的信息，例如学校等
人人网应用	加密方式上传保存通讯录，通讯录匹配进行了严格的返回数量限制

这其中有一部分应用的固有用户基数较少，因此采用本方法大规模枚举手机号码返回的用户数量有限，例如 Kakao Talk 只有 3‰的返回率. 目前流行的通话应用 Viber 用户资料返回率达到了 7.9%，然而由于该应用的主要功能是拨打免费网络电话，用户需要确定的信息是手机号码，所以 Viber 用户账户信息也只由手机号码、显示名称和用户头像构成，该应用中没有其它对攻击者有利用价值的信息. 与 Viber 具有相同功能的有信的用户资料返回率也达到了 7%，而该应用针对大规模通讯录的处理十分敏感，容易出现应用程序崩溃的情况. 另外，社交网站开心网也推出了自己的移动版本，开心网手机应用也同样具有通讯录匹配功能，在匹配成功后能够返回开心网用户的各项资料，包括用户头像、姓名、性别、居住地和大学等. 然而值得注意的是，由于开心网用户设置了隐私级别访问控制，通常情况下部分信息（如被用户设置了“仅对好友开放”的个人资料）对陌生人是不可见的，而那些通过开心网手机应用的通讯录匹配功能返回的推荐列表中的用户，在常理之中与使用该功能的用户具有潜在好友关系，因此通过通讯录匹配返回的推荐好友信息相对于陌生人直接从这些用户的页面上看到的信息要更加完整. 通过本文的方法获得的用户信息包含了这部分被用户设置了隐私保护的信息，如所在学校、工作单位等，这些信息往往是用户不愿被陌生人见到的，从而产生了用户隐私的泄露. 更令人担忧的是，作为目前国内领先且最具影响力的实名制社交网站之一的开心网，其用户资料的真实度较其它应用更有保障且更具价值，因此一旦泄露将会给用户带来更大的损失. 人人网手机应用的可靠性在所有实验应用中表现较为突出，其通讯录匹配过程以加密方式上传，并对匹配的返回内容及数量进行了严格的控制.

5　防御策略

社交通信类应用的通讯录匹配功能是一个非常有价值的功能，它能够给用户线上交流带来极大的便利，同时也是使这类应用能够迅速融入用户日常生活、提高用户黏度的重要途径. 然而，本文提出的这种漏洞利用方法，让攻击者能够通过通讯录匹配对用户信息进行渗透，进而获取用户手机号码和应用账号之间的关联. 为了防范这类行为的发生，避免通讯录匹配功能遭到滥用，本文提出了多种可行的防御策略供应用程序开发人员参考，以设计出更加鲁棒和安全的社交通信类应用程序.

社交通信类应用的通讯录匹配功能的机理是将用户手机通讯录上传至应用服务端，然后将该通讯录的所有联系人与所有已注册用户进行比较，最后将通讯录中已注册该应用的联系人作为推荐的好友返回给用户. 在此过程中，导致通讯录匹配功能被滥用的前提条件有以下 4 个方面：(1) 手机号码能够唯一标识一个手机用户；(2) 应用服务器存储着用户的手机号码（或手机号码摘要值）与该号码所属的用户账号的映射关系；(3) 应用服务器返回到应用客户端界面上的内容包含了用户账户与其手机号码的映射关系信息；(4) 应用服务器没有限制其响应请求的行为，其返回用户数据的数量仅取决于攻击者设定的输入请求，从而使得攻击者能够进行大规模的自动化攻击. 其中，条件 (1) 是移动通信能够实

现的必要条件;条件(2)是通讯录匹配功能能够实现的必要条件,如果应用没有将手机号码和用户账号进行绑定,应用服务器就无法实现通讯录联系人的自动匹配及推荐.所以,既然应用开发者无法保证用户的通讯录是否真实体现了用户的社交关系,那么就需要有效地限制上述条件(3)和(4)的形成.直观地,限制用户单次查询手机号码的数量是最简单有效的方法.然而,这种对数量的限制虽然降低了攻击者大规模发送查询请求的可能,但仍无法阻止攻击者实施多次小规模攻击,同时这种方法也增加了普通用户查询手机号码的重复操作频率.

为此,本文提出了3种可行的方案:

(1)基于实名制的应用程序中,账户的用户名往往能提供其他用户明确的身份提示,鉴于在应用中查询和添加通讯录联系人为好友时,用户实际上已具备了对目标联系人好友的先验知识(因为被应用推荐的好友都来自于用户自身的通讯录),使其仅通过用户名就能够判断该用户的真实身份.因此,建议实名制社交通信类应用的服务器在返回通讯录联系人对应账户的注册信息时,可以仅返回该账户的用户名,而不返回除此以外的其它个人信息,例如手机号码,这样就能有效防止用户个人信息被攻击者收集;

(2)针对未采取实名制的社交通信类应用,建议仅显示目标用户在通讯录中的信息,包括联系人姓名和手机号码,因为联系人姓名已经能给使用通讯录匹配功能的用户以足够的提示来确定推荐好友的身份.推荐好友的其它信息应该设定为仅对已添加好友可见,这样避免了返回手机号码和应用账户之间的绑定映射,阻止了攻击者获得通讯录联系人对应的应用账户信息;

(3)在使用通讯录匹配功能为用户推荐好友时,建议向用户推荐那些通讯录中拥有该用户的手机号码的用户为好友,这种方法特别适用于通过手机号码推荐好友的情形.与公开的个人E-mail地址不同,手机号码的隐私程度更高,反映出的用户关系也更近,相互拥有对方的手机号码这一事实能够反映出双边的朋友关系,采用这种方法不仅不会影响通讯录匹配这一功能的使用,也最大程度地减小了攻击者利用通讯录匹配获取用户资料的可能性.

作者发现本文所述问题后,立即分别向微信和米聊两款应用的开发商腾讯公司和小米科技反应上述情况.随后,得到腾讯公司的回复,声明已经采取了相应的防御措施.作者在使用较大规模的候选通讯录在当前(2013年4月)最新版本(4.5.1)上再次进行实验后发现,微信通过通讯录匹配推荐好友的数量和频率都大大降低,并且在用户添加了部分推荐好友后才会继续向用户推荐剩下的好友,这使得恶意用户无法在短期内实现大规模收集用户资料,但是应用服务器返回用户手机号码与账号映射关系的问题仍然存在,意味着攻击者可以将较大规模的候选通讯录划分至较小规模实施多次自动化攻击,从而无限量地收集微信用户的账户信息.这种仅对返回推荐用户数量进行限制的方法使得恶意获取用户手机号码及对应账号的问题未得到根本性的解决.米聊方面,作者一直没有得到小米科技的相关回应.随后作者使用米聊当前(2013年4月)最新版本(5.0.700)进行验证.验证结果表明,米聊仍然存在本文描述的问题.实验过程中,在勾选了新版本米聊中的"推荐手机联系人"选项后,程序没有直接返回"已添加好友"列表界面,而实验账户不断地收到好友添加成功提示,并有部分好友主动发起的会话.这表明联系人的米聊帐户已经收到了米聊自动发送的添加好友请求,因此可以看出米聊仍然会在通讯录匹配成功后自动向已注册的通讯录好友账户发送添加好友请求,并提示该添加操作是通过通讯录匹配进行的,从而使得用户放松警惕而接受此次好友请求并随即主动发起会话.同时由于米聊客户端数据库并没有实施加密保护,而手机通讯录中已经注册的联系人账户及其详细资料已经存储在本地数据库,恶意用户可以通过读取数据库直接获取所需信息,用户数据仍然存在泄露的风险.

6 相关工作

近几年随着移动终端设备逐渐深入人们的日常生活,移动终端应用程序隐私问题成为了研究人员重点关注的对象.

Enck等人[3]通过修改Android系统的Dalvik虚拟机实现了动态污点传播分析工具TaintDroid.TaintDroid可以标记敏感数据作为污点源,跟踪污点数据,通过污点数据是否流出手机来判断该应用是否存在隐私侵犯行为.PiOS[4]是一款针对苹果公司iOS系统的隐私泄露分析工具.作者静态分析iOS应用的二进制文件,根据Mach-O文件构造控制流程图(CFG, Control Flow Graph).通过基于CFG图的控制流和数据流分析,检测数据是否到达一个汇点,从而确定应用是否可能存在隐私泄露.此外,不少研究关注于隐私泄露的防范,如

MockDroid[5]、TISSA[6]等. MockDroid 是一款控制 Android 应用程序访问敏感资源的工具，其允许用户授予应用程序虚假的权限，所谓虚假是指系统并没有真正允许其访问敏感资源，而是当应用访问这些资源时，系统根据需求返回空值、不可用信息等. 这样的修改虽然很好地防止了敏感信息外流，但当应用程序正常访问资源时返回不可用信息会导致应用意外崩溃. TISSA 同样也是一款 Android 系统上的用户隐私控制工具，允许用户在应用程序运行时动态调整授予权限. 与上述研究不同的是，本文主要提出一种利用现有社交通信类应用中的通讯录匹配功能自动收集用户手机号码和存储在应用服务器上的个人资料的方法，而隐私信息如何从手机泄露不是本文关注的内容.

此外，研究人员也在用户资料自动收集方面做了很多工作，尤其是面向社交网络(Social Network Service，SNS)方面. Bilge 等人[7]通过 SNS 网络收集用户资料，达到了自动窃取身份的目的，然而该方法要求目标账号与攻击者具备好友关系，也就意味着攻击者在获取用户资料之前需要首先与目标对象相互添加为好友. Barlduzzi 等人[1]利用社交网络中的 E-mail 查询功能收集用户资料. 这个功能允许用户查询并验证其 E-mail 地址列表中的好友是否在该社交网络中已注册了账户. 该团队通过抓取 SNS 用户主页收集了大量的用户信息，并对这些信息运用了关联分析. 对比前人的研究，到目前为止本文是首次提出利用移动终端应用作为媒介来自动收集用户的个人资料. 本文利用社交通信类应用的通讯录匹配功能获取用户的手机号码及其存储在应用中的个人资料，通过监控与处理用户资料信息相关的 Android API 和模拟用户浏览动作，实现了自动收集的过程且不需要添加目标用户为好友. 本文的方法与 Barlduzzi 等人的研究都可以被用来完成社交工程分析的前期准备工作.

目前通讯应用中使用手机号码作为账户唯一标识符的情况越来越多，Schrittwieser 等人[8]针对这种现状做了一系列的分析. 该工作主要针对应用的认证机制. 分析结果显示，使用手机号码作为账户的唯一标识符会带来众多安全问题，例如账户劫持、ID 欺骗等. 同样，本文的实验结果显示，配置了通讯录匹配功能的应用也存在用手机号码作为用户账户的唯一标识符的情况. 针对该问题，本文最后提出了一系列防御策略来保证通讯录匹配功能在正常使用的同时仍能够有效地防范该功能被恶意利用.

7　结　论

本文揭示了移动终端中配备通讯录匹配功能的社交通信应用存在的用户隐私泄露安全隐患，该安全隐患使得攻击者能够通过通讯录匹配机制自动地收集应用注册用户的个人资料信息. 本文具体展示了一种利用该漏洞来自动收集用户资料(包括手机号码)的方法. 这是利用移动终端应用自动收集用户个人资料方向的首次探索. 此外，本文基于实现的原型系统进行大量实验，实验结果充分验证了该方法的有效性. 针对从多个应用获取用户账户个人信息的情况，本文还提出了两种通过跨应用分析获得数量更加庞大、字段更加全面的用户资料的方法，即横向整合和纵向渗透，同时还进行了用户数据的一致性和真实性分析以确定更真实的用户信息. 实验结果表明，用户通常倾向于在移动社交通信类应用中填写真实的个人信息，而对于不同的应用，其不同的用户价值定位和推广策略导致用户对其信任度略有差异，从而使用户在各应用中填写的用户资料的真实性也各不相同. 最后，通过分析通讯录匹配功能漏洞产生的原因，本文提出了一系列防御策略. 建议开发者在设计开发社交通信类应用时应时刻关注用户真实世界身份和虚拟网络身份的界限，当两者出现关联时，应尽量避免显式返回这种关联关系，同时还需要一定程度地限制返回的数据量大小，阻止恶意用户的自动批量攻击.

参　考　文　献

[1] Balduzzi M, Platzer C, Holz T, et al. Abusing social networks for automated user profiling//Proceedings of the Recent Advances in Intrusion Detection. Ottawa, Canada, 2010: 422-441

[2] Jakobsson M, Johnson N, Finn P. Why and how to perform fraud experiments. IEEE Security and Privacy, 2008, 6(2): 66-68

[3] Enck W, Gilbert P, Chun B-G, et al. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones//Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation (OSDI'10). Berkeley, USA, 2010: 1-6

[4] Egele M, Kruegel C, Kirda E, Vigna G. PiOS: Detecting privacy leaks in iOS applications//Proceedings of the 18th Annual Network & Distributed System Security Symposium (NDSS). San Diego, USA, 2011

[5] Beresford A R, Rice A, Skehin N, Sohan R. Mockdroid: Trading privacy for application functionality on smartphones//Proceedings of the 12th Workshop on Mobile Computing Systems and Applications (HotMobile'11). New York, USA, 2011: 49-54

[6] Zhou Y, Zhang X, Jiang X, Freeh V W. Taming information-stealing smartphone applications (on android)//Proceedings of the 4th International Conference on Trust and Trustworthy Computing (TRUST'11). Berlin, Heidelberg, 2011: 93-107

[7] Bilge L, Strufe T, Balzarotti D, Kirda E. All your contacts are belong to us: Automated identity theft attacks on social networks//Proceedings of the 18th International Conference on World Wide Web (WWW'09). New York, USA, 2009: 551-560

[8] Schrittwieser S, Fruehwirt P, Kieseberg P, et al. Guess who is texting you? evaluating the security of smartphone messaging applications//Proceedings of the Network and Distributed System Security Symposium (NDSS 2012). San Diego, USA, 2012

CHENG Yao, born in 1987, Ph. D. candidate. Her research interests are mobile security and privacy, including security and privacy issues in systems, software and network.

YING Ling-Yun, born in 1982, Ph. D., assistant professor. His research interests include malware analysis and mobile security.

JIAO Si-Bei, born in 1986, Ph. D. candidate. His research interests focus on Android security and malware analysis.

SU Pu-Rui, born in 1976, Ph. D., associate professor. His research interests are malware analysis and prevention.

FENG Deng-Guo, born in 1965, Ph. D., professor, Ph. D. supervisor. His research interests are cryptography and information security.

Background

As mobile devices dominate people's work process and daily communication, huge amount of user privacy information has been stored in the phone, such as contacts, agenda, photos and so on. Privacy should never be ignored, especially in mobile devices. Nowadays, there is a category of applications on mobile devices devote themselves to people's daily communication. This kind of applications provides rich user experience and is cheaper than the SMS or MMS which make them gain a quick popularity and catch on around all over the world. A majority of the social messaging applications equip a component called "Address Book Matching", which takes advantage of user's address book and recommends all of the contacts who also have an account in the same application. This component facilitates user a lot to get in touch with their contact friends, however, also brings some privacy leakages problems.

There are a lot of research concerning about mobile privacy, such as using dynamic taint analysis or preventing privacy leakage by forging sensitive information. However, no one has mentioned the malicious usage of a popular component in social messaging applications. The paper proposes a novel way to leverage the "Address Book Matching" component to gather user information in large scale and proofs the availability via a prototype system. Though the method, a stranger can get user's phone number and the corresponding user account in applications. The paper also presents analysis approaches to get more complete and more authenticate information. Through deeper analysis of the core factors that lead to the vulnerability, the paper also gives several alternative countermeasures to avoid the leakage effectively. The research group has done much work in the area of protecting security and privacy of mobile systems and proposed dynamic analysis tools of mobile applications based on the hardware virtualization. They also give vulnerability reports to the relevant companies who immediately give feedback, discussing about the particular amendment.

This work is supported by National Program on Key Basic Research Project (Grant No. 2012CB315804), National Natural Science Foundation of China (Grant No. 61073179 and No. 91118006) and Beijing Municipal Natural Science Foundation (Grant No. 4122086).