互联网和网络空间治理专栏系列

互联网用户泛隐私安全热点问题回顾与浅析*

关键词：可信编译　编译器构造方法

肖新光
安天实验室

近期互联网热点事件回顾

在过去的几个月，隐私问题凸显，传统的恶意代码泄露已过渡到“合法”的网络应用，一些著名公司甚至寡头厂商被曝光在聚光灯下。

《环球邮报》近日报道[1]，Twitter（推特）向数据挖掘公司出售用户数据。Boulder、Gnip和DataSift获得Twitter授权分析其存档的tweet（Twitter上的留言）和访问如地理位置等用户基本信息。隐私保护组织担心这对用户的隐私有深远的影响。Privacy Rights Clearinghouse（隐私权情报交换所）组织主席保罗•史蒂芬斯（Paul Stephens）说，收集一个人在一年多时间内说的话具有颠覆性意义，用户的思考过程可能会被挖掘出来。另一位隐私专家说，商家正在了解用户正在想什么。

表1表明了过去的几个月内发生的一些事件。

上述曝光事件引发了人们对合法企业、巨头企业、甚至是寡头企业的行为的担忧，以及对用户隐私受到威胁的担忧。尽管去年媒体通报了恶意代码对用户电话录音等的威胁形式，但在过去几个月内，互联网企业而且是主流的互联网企业、互联网巨头甚至是寡头引起了我们对个人信息和隐私保护

表1 被曝光/报道事件

时间	被曝光/报道事件
2011.09	Facebook在用户退出登录后仍跟踪用户的行为[2]
2011.11	CarrierIQ软件搜集回传隐私[3]
2011.12	Facebook记录大量用户的行为以及用户删除的数据[4]
2012.01	谷歌更改其大量网络服务的隐私条款[5]
2012.02	谷歌通过付费方式搜集和分析用户数据[6]
2012.02	微软和谷歌互相指责对方浏览器的隐私设置[7~8]
2012.02	Facebook、Youtube等的Android（安卓操作系统）和iOS（iphone operating system，手持设备操作系统）应用被指责过度搜集用户隐私、读取短信和拦截电话等[9]
2012.03	Twitter向数据挖掘公司出售用户信息和微博内容[10]

* 本文系根据作者在2012年信息安全论坛上的同名报告录音整理的，因涉及到具体的案例事件，故有较多删节

的关注。

在这样的背景下，隐私威胁的边界趋于模糊。在过去传统PC的互联网领域，有一个庞大的地下经济产业。他们主导的工具是其编写和加工出来的样本，如果用HASH（哈希算法）计算的话，可能是一个数以亿计的恶意代码样本集合，这就是PC上的用户价值侵害主题。但目前从手机终端上来看，这种过度的数据收集主要是由合法厂商实施，由于安全边界一下子变得模糊了，因此区分谁是敌人谁是朋友的问题变得非常复杂。

智能终端的发展进一步带动隐私问题泛化

智能终端设备的快速发展进一步带动了隐私泛化。根据相关统计机构的统计[11]，2011年智能手机的出货量达到480万台，首次超过了PC终端410万台的数量。

表2 智能终端设备出货量统计（数据来源：Canalys. Smart phones overtake client PCs in 2011）

	2011年出货量（百万台）	年增长率
智能手机	487.7	62.7%
全部PC终端	414.6	14.8%
平板电脑	63.2	274.2%
上网本	29.4	-25.3%
笔记本	209.6	7.5%
台式机	2.4	2.3%

通过对比个人电脑和智能手机，我们发现两者有很大不同（如表3所示）。

表3 个人电脑和智能手机对比

	个人电脑	智能终端
与用户关联程度	较低	较高
平台架构和系统	基本统一	异构分化
系统开放性	比较封闭	高度开放
产业链分工	简单	复杂
安全防护方案	比较成熟	尚未成熟

个人电脑往往与用户的关联较低。例如，网民能使用网吧的电脑，但手机一般都是用自己的，因为手机是一种与个人身份关联的设备；基本上是IBM确定了三架马车体系（IBM-PC标准、Win-tel联盟）之后，PC和主流操作系统基本趋同并保持向前的兼容。而智能终端无论从硬件体系结构，还是从几个主流智能机操作系统来看，都呈现出异构分化的角度；个人电脑是一个相对封闭的简单产业链（在高度的标准化的情况下，以PC出品商的品牌和供应量为主导），而智能终端是一个开放而复杂的产业链；从安全防护解决方案的角度来看，虽然基于PC的恶意代码经过几十年，积累了大量与安全体系对抗的经验，但同时安全体系也经历了长期发展，其中既包括安全厂商与操作系统厂商的磨合，也包括操作系统厂商对自身安全的改进等等，因此也比较成熟。而智能终端在这方面尚十分稚嫩。

正因为手机产业链比较复杂，链条上的每个环节都有话语权，但又不承担主要的安全责任，因此导致了手机产业的威胁源头即手机制造商、手机操作系统的生产厂商、移动的运营商以及手机应用软件的开发商的分离。每个环节都能从安全模式上对整个手机安全给予深刻影响，并且在过去的一段时间真实地引入了大量安全威胁。这种威胁产生的类型可分为三种。

1. 失误型 一些厂商的软硬件缺陷导致用户隐私可能被泄露。

2. 故意型 合法产业的角色在法律所不及之处，因利益因素故意侵害用户隐私。

3. 复合型 厂商本身有故意搜集用户信息的意图，由于其程序存在bug（漏洞），导致了程序又被攻击者利用。

智能终端厂商的案例

2011年，在HTC自己开发的Sense界面中，由于一个名为HTCLoggers的组件存在漏洞[12]，所有其他应用程序只要拥有访问互联网的权限，都可以利用这个漏洞获得手机中大量的隐私数据。

2012年1月31日，HTC又被发现另一个安全漏洞。在一些型号的手机上，攻击者可以通过这个漏洞拿到所有已经保存的wi-fi（wirelessfidelity，无线

保真）证书[13]。

这是由于厂商失误导致隐私受到威胁的例子，目前HTC已经通过发布补丁对上述问题予以解决。

智能终端OS的案例

从Android 4.0开始，它的默认配置是将用户的SSID（service set identifier，服务集标识）和wi-fi的用户名和密码备份到谷歌的服务器上。备份的还原凭据就是用户的谷歌账号。在把单账号和密码作为凭据进行还原的情况下，我们尚无法证实谷歌方是不是可见备份数据。这个服务是一个默认开启的服务。不可否认，这是一个非常方便的功能，因为用户可能在单位、家里、咖啡厅配置了很多密码，如果缺乏有效的备份，这些密码可能会丢失。

我们可以联想起沸沸扬扬的谷歌无线街景车的监听事件。2010年，谷歌街景车被发现有缓存和上传所经过地区的wi-fi数据的行为[14]。为此，谷歌遭到欧盟、澳大利亚等地的调查。谷歌在韩国的办公室甚至遭到了警方的搜查[15]。如果把这两点结合起来，实际上谷歌的这种云备份，延展了其监听全球互联网的能力。而这种能力的完成不是依赖大量的CPU +GPU（graphic processing unit，图形处理器）和大量的云计算资源，而是通过用户的体验，让用户自动提交备份数据。

运营商的案例

Carrier IQ是一个为美国运营商提供用户体验和服务支持来改进工具的公司。2011年的“Carrier IQ事件”，媒体是10月份报道出来的，但据事后追溯，从2011年6月份开始，国外主流ROM定制的民间组织就开始逐步删除这个应用[3]，说明相关质疑由来已久。事后它宣称没有重度采集用户信息的行为。但是从它注册的相关专利以及从它网站上删除的文档来看，它具有全面的数据采集能力。根据我们的分析，它的采集行为非常隐蔽，是写入某一个内存区域，然后通过触发条件上传。对这种触发条件我们进行了成功的模拟并发布了分析报告。

这个事件把美国三大运营商Sprint、AT&T（American Telephone & Telegraph Company，美国电话电报公司）、T-Mobile都牵扯其中。相关报道称受影响的设备数量达1.41亿台。这个事件表明了通信运营商在手机价值链条上拥有巨大的话语权。手机这种生存形态与PC不一样，PC使用的固网运营者多数只起到一个提供通路的作用，没有更多的附加业务，也没有更多的话语权；而移动运行商由于把控接入的因素，能够影响终端厂商向里边进行相应的预装。这个情况在美国更严重，因为他们是以合约用户为主导的。

软件开发者的案例

根据某研究机构对App Store（application store，应用商店）的调研，68%存在回传UUID（universally unique identifier，通用唯一识别码）的行为[16]。《华尔街日报》的调查数据与此接近，App Store应用中，56%回传UUID，47%回传个人位置，5%回传个人数据[17]。显然，这些都是采集用户隐私的故意行为，而以免费为主导的安卓市场（Android Market）相关过度采集问题严重得多。

操作系统厂商通常扮演着软件提供者的角色，由于具有先天预装和用户信任的优势。它如果出现失误，则出现的问题比一般开发者的严重得多。

谷歌钱包是安卓手机上第一款基于NFC（near field communication，近距离无线通信）技术的支付软件，可以把手机当成信用卡直接刷卡。前段时间就暴露出谷歌钱包用明文存储了用户的相关信息，其中包括用户信用卡的后四位等等[18]。这样相关实名信息就可能被恶意的应用获取。

免费模式与传统安全价值的冲突

过去我们探讨隐私时，往往用微技术化观点或泛道德观点来考虑这个问题，缺少对这种隐私威胁风起云涌的原因上的价值链的探讨。一种状态的迅

速发展或者恶化，必然有其经济链条支持。过去，安全研究者更关注以恶意代码为主要工具来进行活动的地下经济产业链。但从前面的案例来看，聚合和挖掘用户隐私同样是合法产业的动力之源，因此我们需要思考其内在规律和崛起的原因。

作为互联网行业“必读本”的《免费——商业的未来》一书介绍了这种新兴“免费”商业模式是一种建立在电脑字节基础上的经济学。这种趋势正在催生一个庞大的新经济。这本书列举了较多的商业现象，但对价值链的规则挖掘并不深入。书里只是介绍了一种“引导性促成销售模式”，如广告商获得被免费内容吸引的消费者的姓名和电子邮件地址或与这些消费者相关的信息等等，但并没有提到安全工作者对这种模式产生安全隐忧，更不用说如何打消这种隐忧了。

我们可以看一下这种免费模式的演进构成。一般来看，不管是免费的应用还是免费的客户端，若被用户采纳并非是由于价格因素，而是来自用户体验。例如Gmail（谷歌提供的电子邮件服务），大家的体验实际上比用任何其他收费的邮件服务感受都好。互联网的开发方法就是以用户的体验为引导快速开发改进的模式，而这种用户体验中的改进亦可产生一些革命性的工程方法。例如，原来输入法的人工智能问题很难解决，但搜狐输入法基于互联模式和用户输入频度统计，就形成了高质量的输入建议。因此用户体验的聚合与联合改进，推动了产品形成巨大的用户群。在用户群的基础上，挖掘它的产品价值，产品价值形成之后，免费将去何方呢？它最终要把自身做成一个开放的平台，当平台的效应形成时，仿佛就是免费模式最为开放的时候。但当你完全接受这个架构之后，你会发现，它基本上成为了你在这个领域上的唯一入口。因此我们回顾一下，当免费模式到来的时候，我们下意识地从情感上以为它与开源是同类，代表一种驱动自由竞争和繁荣的倾向。但后来可以发现，在几乎有主流免费的领域，其产品的竞争性比原有以收费为主导的情况少了很多，最终会推动垄断和寡头的形成。

我们需要探讨一下免费价值在这个过程中是否侵害和干扰了用户权益。互联网用户的基本权益可以概括为以下三点：（1）身份的自我保护；（2）操作的自主选择；（3）信息的自由获取。

从各国官方的角度看，这些用户的基本权益是在某些公共权益的约束之内的。例如，这些权益的实行要基于相应的道德、契约和法律，特别是知识产权等要以社会基本准则为基础，也要以尊重政府的合法监听和管理权为边界。

传统的安全产品模式和价值观，甚至包括一些传统应用软件的价值观，是对这种基本权益的体现。比如，个人防火墙就是一种典型的具有身份自主保护的产品，其功能“block all package coming in”，降低了诸如操作系统指纹等节点信息被探测到的可能；传统的反病毒软件能拦截一些恶意代码的非授权的操作，并在发生一些有害数据的传递比如恶意代码的复制时，和用户产生询问交互，或者按照用户的定义去处理，实际上是保障了用户的自主选择权。

传统的浏览器默认首页是空白的（当然，发布者也可以定制）。但是，大多数手机上的商业浏览器给用户订制好了内容，用户就会下意识地去阅读这些内容。传统的网络信息获取是用户基于大的门户站点，或者自己去用搜索引擎选择自己去看什么，这就是信息的自由获取，而且是平衡的获取。而现在的信息，不仅是定制好的，而且是用弹窗推送给用户。对传统的用户体验模式的颠覆，导致了用户逐渐对自我权益的放弃。所以，站在安全的角度，我们对免费的思索是它的价值链基础是聚合海量的弱隐私信息。但它们的价值交换方式是用户放弃部分权益换取免费。在这种换取过程中用户得到了更舒适的体验，互联网厂商和用户之间是各得其所的。这个价值链达成了对用户的信息和行为的控制。

互联网以广告投放、流量重定向和信息封装为主导的免费模式，一定程度上依赖于获取用户的个人信息来实现广告的精确投放；通过让用户感受更加方便的方法来削减用户的选择范围；通过信息推送的方式，影响用户对信息的平衡获取。但是免费模式确实和用户的基本权益存在着某些冲突。我认为免费模式的三个行为支撑点分别是采集、投放和接管。采集是对用户身份、用户习惯、用户存在的位置和相关的社会关系的采集；投放就是向用户投放聚类和再加工的信息，投放自己推荐的软件以及评级和建议，包括投放软硬件方面的广告等；接管是代替用户的所有操作，从而形成一个替用户管理的局面。采集行为伤害的是自我保护的权利；投放行为伤害的是自由获取的权利；接管行为伤害的是自主选择的权利。所以，这种免费模式造就的隐私隐患，依赖于隐私聚合并具备滥用隐私的可能性。

这里，我要对在大规模互联网场景下进行生物识别提出疑议。很多大的厂商如Facebook都提到人脸识别。人脸识别能提升安全性吗？当然不能。比如拖库事件，过去是丢失了用户名和可以废止的密码，现在连不可改变的脸部签名信息都丢了。但为什么这些厂商要这么积极地推进人脸识别呢？实际上就是为了实现更为精确和更有消费价值信息的获取。例如，对圆脸用户，可以推送减肥广告；对有雀斑的用户，可以推送祛斑广告。我们并非敌视这个商业模式，而是反对那种假以安全语汇，通过制造用户恐慌，来推广并掩盖商业图谋的行为。

不乐观的未来

当用户对免费模式的依赖已经达到一定程度时，我们看到一幅名为“免费（Free）”的巨大、诱人的画卷在缓缓展开，在展开的最后，会不会是一把匕首呢？

套用国外一些学者比较常用的判断逻辑：“Threat can be seen as a combination of capability and intention（威胁是能力和意图的结合）”，即在界定是否存在某些威胁时，需要对对象的capability和intention进行考核。当前互联网寡头们的相关能力是毋庸置疑的。那么意图就变成更重要的判断对象了。

我们可以理解为，分析寡头们是否有主动对抗用户信息自我保护意愿的行为，是判断啊它是否有采集扩大化或者滥用意图的一个重要风向标。

我们先来看谷歌的动向，谷歌通过一系列自身的研发和创新，包括一系列的兼并和整合后，形成了一个庞大的应用体系。谷歌原来的各种应用，只能分析自身所采集的数据，而不能交叉使用，而在其调整隐私信息后，就可以交叉使用了。隐私保护维权人士抨击了谷歌的最新举措。华盛顿独立隐私保护和安全研究员索霍安（Christopher Soghoian）说[19]：“现在用户只要上网就处在谷歌的监视之中。没有哪一个独立的实体应该被委以如此多的敏

感数据。”

苹果的产品定期回传用户的GPS坐标、wi-fi接入点、基站信息等[20]，每12小时向苹果的服务器发送一次，这些产品包括 iPhone、iPad、iPod Touch、Snow Leopard系统以及Safari浏览器（甚至包括windows版的Safari ）。当被发现之后，它采用的做法不是道歉，不是解释说明，而是强制性地修改用户条款，要求用户必须接受[21]。

当用户对品牌的认知一旦形成，那么巨大的用户群体惯性将使寡头在一段时间内可以为所欲为。而这种群体惯性成为事实之后，就可以藐视传统的法制和安全惯例。这种惯性将成为一种阻挡的力量。

最能说明问题的是近期微软和谷歌关于隐私浏览问题的冲突，微软指责谷歌用相关JS脚本绕过IE隐私浏览模式，来获取用户信息[7]，而谷歌回应IE隐私保护功能不符实际 [8]。首先，隐私浏览的模式，是有行业标准的；其次，隐私浏览模式并不是浏览器的默认模式，而是用户强制启动模式。也就是说，这是一个非常明显的用户主观意愿。因此这种获取是一种明显与用户的意愿对抗的行为。

但是我们能用传统的道德观点或者行业标准去要求谷歌吗？不能。因为它提供了免费的应用，并把这点写入使用免费应用的用户义务，使其具有了无可争议的话语权。所以，在加入移动背景场景的情况下，隐私威胁从源头、目的和方式上呈现分散化的趋势。更大的威胁是从主观的直接威胁过渡到各种间接风险。在免费的模式下，隐私价值已经能够支撑巨大的产业链条，用户的权益被逐渐侵蚀。

值得反思的是，我们一直在探讨恶意代码、漏洞挖掘、各种攻击防御技术以及各种加密算法的合理应用等等。但实际上，我们始终在关注这种具体的威胁，猜测一些并不具备实证性的故意行为，但对产业模式变革带来的安全风险缺少足够的警惕。

这不是一篇反智能终端的战术，我始终认为作为一个信息安全工作者，需要做的是保障用户利益，但不是限制用户的价值；这不是反互联网模式的宣言，我认为需要警醒的不是互联网模式本身，而是互联网模式无节制地膨胀反噬用户的利益。

致谢：感谢我的同事肖梓航、李琦等对本文的贡献

肖新光

安天实验室首席技术架构师。。主要研发方向为反病毒引擎、大规模恶意事件样本的流水线处理、计算机犯罪取证等。seak@antiy.com

参考文献

[1] Twitter sells your feed to Big Data. http://www.theglobeandmail.com/news/technology/digital-culture/social-networking/twitter-sells-your-feed-to-big-data/article2355287/

更多参考文献：www.ccf.org.cn/cccf

北航成立CCF学生分会

2012年3月22日，中国计算机学会（CCF）在北京航空航天大学成立学生分会，这是CCF在北京成立的第一家学生分会，也是CCF在全国的第五家学生分会。CCF秘书长杜子德、副秘书长马殿富、刘雨出席了成立大会。北航计算机学院院长吕卫锋教授，副院长胡春明博士，党委副书记张炯博士,研究生指导主任牛建伟教授及近百名学生参加了成立大会。

成立大会由胡春明主持。杜子德、吕卫锋致辞，阐述了CCF学生分会的成立对于学院计算机学科的发展和对计算机科学技术人才培养的积极意义，鼓励同学们在这个平台上，充分利用资源展示和提升自我。杜子德为CCF北航学生分会第一任执委会颁发了证书。

CCF 北航学生分会首任主席、副主席名单：

主 席：唐晓岚副主席：张宁远、孙铭涛、孙剑文