网络隐私保护技术综述*

张　军, 熊　枫
(广东商学院 信息学院, 广东 广州 510320)

摘　要: 随着电子事务、电子政务等各种网络应用的不断发展, 网络隐私引起了广泛的关注, 并成为了一个急待解决的问题。介绍了网络隐私的概念及其泄露的方式, 重点讨论了目前的网络隐私保护技术, 最后结合互联网的发展, 给出了网络隐私保护技术的研究方向。

关键词: 网络隐私; 匿名; 隐私偏好平台(P3P)
中图法分类号: TP391　　**文献标识码**: A　　**文章编号**: 1001-3695(2005) 07-0009-03

Survey on Technologies of Internet Privacy Preservation

ZHANG Jun, XIONG Feng
(School of Information, Guangdong Commercial College, Guangzhou Guangdong 510320, *China)*

Abstract: With the rapid development of various applications based on the Internet such as e-business and e-government, the Internet privacy has drawn considerable attention and become an urgent problem. This paper introduces the concept of the Internet privacy, and investigates the main sources of the Internet privacy disclosure. It focuses on surveying the current technologies of the Internet privacy preservation. Finally, taking into account the evolution of Internet, it provides the view on the trends of the technologies in the future.

Key words: Internet Privacy; Anonymous; Platform for Privacy Preferences(P3P)

1　引言

网络隐私问题已成为社会各界关注的焦点, 并严重威胁着网络社会与网络经济的健康发展。据波士顿咨询公司的一项调查显示: 隐私比成本、易用性和安全性等更为用户所关心。另据木星通信公司估计, 2002 年, 这种对网络隐私的担心造成了高达 180 亿美元的经济损失; 到 2006 年, 网络隐私将成为制约电子商务发展的最大障碍。

何为网络隐私? 网络隐私是集社会、法律、技术为一体的综合性概念[1~3]。早在 1890 年, Samuel Warren 等将隐私定义为保持个人独处的权利; 后来, Ferdinand Schoeman 将隐私定义为个人决定和控制自己的信息被他人共享和使用的权利。最早涉及保护个人隐私的立法是 1948 年的《人权世界宣言》, 由此可见, 隐私是一个社会、文化、法律概念。20 世纪 80 年代, 随着 Internet 所带来的信息革命, 以及电子商务、电子政务、网络教育的兴起与蓬勃发展, 人们在网上的活动越来越多, 如网上浏览查询、聊天、购物、学习、收发邮件等。在这些网上活动中, 将涉及到大量的个人隐私信息, 如 IP 地址、网址、E-mail 地址等, 而这些个人信息很容易利用现有的技术手段进行收集和利用, 于是隐私概念又增添了技术的内涵, 从而产生了网络隐私概念。一般地, 网络隐私是指网络用户隐藏个人信息以及控制个人信息被他人共享的权利。其中, 网络用户的个人信息主要包括以下三个方面: ①个人数据, 如姓名、职业、通讯地址、电话、E-mail 地址等; ②数字行为, 如浏览了哪些网站、停留了多长时间、看了哪些网页以及订购了哪些商品等; ③通信内容, 包括电子邮件、公告牌留言、在线选举等。

收稿日期: 2004-04-17; 修返日期: 2004-07-16
基金项目: 广东省自然科学基金资助项目(020199) ; 广东省教育厅自然科学基金资助项目(Z02042)

虽然网络隐私引起了社会各界的广泛关注, 而且人们也采取了许多法规、自律政策和技术等措施来加强对网络隐私的保护, 但这些措施的作用非常有限, 隐私保护仍然是一个亟待解决的挑战性问题。本文将介绍网络隐私泄露的方式和典型的网络隐私保护技术, 并指出网络隐私保护技术的发展趋势。

2　网络隐私的泄露方式

网络隐私泄露主要表现在以下几个方面:

(1) 非法转让隐私信息。例如医疗信息网站 DrKoop. com 在没有征得用户的许可就将他们的健康信息出售给网站 vitacost. com, 结果因其侵害了他人的隐私权而导致破产。

(2) 非法窃取隐私信息。由于 Internet 的弱安全性, 很容易导致黑客利用这一弱点入侵他人的计算机而窃取隐私信息。例如在 2000 年, 一黑客侵入 Seattle Hospital 的计算机网络, 窃取了包含 5 000 病人的健康信息。

(3) 利用在线注册收集隐私信息。用户为了获得各种 Web 服务, 如申请邮箱、注册抽奖或是网上购物等, 网站常常要求用户填写一个登记单, 作为用户利用某项服务的前提条件。这种登记单需要用户输入许多个人信息, 如年龄、性别、出生年月、收入、职业、个人爱好等, 如 263 站点的免费电子邮件服务登记单就有十几项之多。

(4) 利用 IP 地址跟踪用户的位置或行踪。基于 HTTP 的 Web 浏览是 Internet 上最广泛的应用, HTTP 是一个服务器和客户机之间的交互协议, 用于分布式、协作式的超媒体的信息

系统,它运行在一个可靠的传输协议如 TCP 上面。在 Web 服务器和客户机建立会话时,IP 地址、URL 和软件版本等信息都将传送到服务器,因此,对方可利用 IP 地址跟踪用户的位置或在线行为等。特别是,随着 IPv6 的使用,HTTP 的消息头中将包含更直接的位置信息。

(5) 利用 Cookies 文件收集用户的隐私信息。Cookie 是服务器存放在客户机上的一个文件,该文件包含有用户所访问的网页、访问时间、甚至含有电子邮箱密码等。采用 Cookie 机制指定了一种用 HTTP 请求和回复消息创建状态会话的方式,它描述了两个新的标头,Cookie 和 Set-Cookie 头,用于携带参与的服务器和客户机的状态信息。因此,Cookies 具有重构网络用户所从事的网络活动的功能,通过对用户在网络上访问网站、查看产品广告、购买产品等行为的跟踪,结合网络注册系统,就可以得出用户的健康状况、休闲嗜好、政治倾向、宗教信仰等信息,从而生成有关用户的个人档案。

(6) 利用特洛依木马病毒窃取隐私信息。当用户从网站下载免费软件并安装时,如果该软件中含有特洛依木马病毒,则它可窃取用户的隐私信息并上传给网站。

(7) 利用嵌入式软件收集隐私信息。如今,计算机语言如 Java, JavaScript, XML 和 ActiveX 等的功能越来越强,它们允许远端的服务器运行客户机上的应用软件,这些软件可将客户机的计算环境和个人数据传送给服务器。随着高速宽带网在娱乐业中的应用,这种隐私泄露现象将会越来越突出。

(8) 利用 Web Beacons 窃取隐私信息。Web Beacons 也称作 Web 臭虫或电子影像,由于它是被定义在 HTML 脚本文件的 IMG 标记中,且大小为 1 ×1 的图片,因此用户在网页上是无法看见的。它通常是被第三方置于网页中用于监控访问该网页用户的行为,如用户点击的内容或统计访问网页的人次。更为甚者,通过连接远端的服务器,服务器可秘密存放一个 Cookie 文件在客户机上并收集用户输入的关键字等。

(9) 利用窜改网页收集隐私信息。攻击者有两种方法窜改网页:①攻击者首先让用户访问他的网页,如提供一些有趣或好笑的内容。实际上,他的网页是一个陷阱,当用户点击网页上其他的链接时,他就会被带到一个错误的网页上去,因为这个链接已经被攻击者窜改了。②攻击者修改了用户的 URI (如书签中),当用户想要去一个真的服务器网页时,被窜改的 URI 就会带他去攻击者的机器,攻击者可能给他一个错误的网页,或者将原始的 URI 请求传到真的网页服务器上,然后攻击者截取其响应。这样,攻击者通过提供错误的网页,可获取用户访问了哪个网页、在网页表格中输入了哪些数据等隐私信息。

3 网络隐私保护技术

一般地,保护隐私有四种措施:①政府制定相关的法律法规;②行业自律,各行业制定隐私保护政策,并在其网站上公布;③加强对消费者和 IT 人员的隐私宣传教育;④研发和利用各种隐私保护技术。本文仅对技术进行分析和讨论,根据实施隐私保护的主体将它们分为三类:基于用户、中间代理、服务商的隐私保护技术。

3.1 基于用户的隐私保护技术

基于用户的隐私保护技术用于控制他人访问存储在客户机上的个人隐私信息,主要有个人防火墙和 Cookies 管理器[4]。个人防火墙是保护个人计算机系统的一个软件,可以防止特洛依木马、黑客程序等窃取客户机上的个人隐私信息,也可屏蔽某些 IP 地址的访问。目前有 ZoneAlarm, NetBoz, Outpost 等软件公司开发的个人防火墙软件。Cookies 管理器允许用户关闭 Cookies 文件,选择性地接收来自某些服务器的 Cookies 文件以及搜索和查看其中的内容。这类软件主要有 Bullet Proof Soft 和 No Trace。以上这些技术或软件可安装在用户计算机上运行,使用简便,但它们只能起到防备性的保护作用,不能控制用户在网络交互过程中的隐私泄露。

3.2 基于中间代理的隐私保护技术

基于中间代理的隐私保护技术利用中间代理来隐匿用户的身份,从而保护用户的隐私。匿名技术主要分为以下三种:

(1) 基于代理服务器的匿名技术(Proxy-based Anonymizers)。用户访问某个网站的请求首先发送到代理服务器,代理服务器将用户的 IP 地址等信息转换为一个匿名信息,然后再将这个请求发送到目的网站,从而使网站不能识别用户的身份。这种代理服务器主要有 Anonymizer, Lucent Personal Web Assistant (朗讯个性化 Web 助手, LPWA), IPrivacy 和 WebSecure。在 LPWA 中,当用户给一个网址发送 HTTP 请求时,这个请求首先到达 LPWA, LPWA 将 HTTP 请求头中的潜在的身份认证信息过滤掉。如果该网站提供个性化服务,需要用户输入一个表格时,用户只需分别输入“\u”,“\p”,“\@”作为用户名、口令和 E-mail 地址就可以了,LPWA 会自动计算出一个假的用户名、假口令和假 E-mail 地址来填写表格,然后递交给服务器。与此类似,也有相应的匿名邮件重发器,称为类型 0 的邮件重发器(Remailers)[5,6]。以上匿名技术的不足就是代理服务器知道用户的真实身份,因而存在代理服务器泄露用户隐私的隐患。

(2) 基于路由的匿名技术(Routing-based Anonymizers)。这种技术将用户的请求通过多个中间主机发送到网站,使得网站和每一个中间主机不能识别用户的 IP 地址,如 AT&T 公司的研究项目 Crowds(群体)。群体是以“混杂于群体”为原理的一个匿名工具,即一个用户的行为隐藏在其他用户的行为里。加入群体的用户首先要在他的电脑上运行一个叫做“jondo”的进程,然后他在一个叫做掺和器的服务器上注册,取得一个账号(用户名和密码)。掺和器将新的 jondo 加入群体,并告诉群体中其他成员的 jondo,同时这个新的 jondo 也会得到其他已注册过的 jondo 清单。当用户发出访问网站的请求时,这个请求交给他的 jondo,jondo 剥去那些可能从请求中确定用户身份的信息,再以概率 $p(p>0.5)$ 随机地从群体中选择一个 jondo 或以概率 $1-p$ 选择终端的服务器,经过一定数量的 jondo 后,请求将会到达终端的服务器,以后的请求使用同样的路径通过网络。最后,网站的响应信息再按相反的路径顺序发送到用户[7,8]。这种技术的不足是:如果黑客监听用户和终端的服务器之间连接的所有通信,就会分析出发送者和接收者。基于这种原理的邮件系统称为类型 1 的邮件重发器。

(3) 基于洋葱路由的匿名技术(Onion Routing-based Anonymizers)。这种技术的基本原理是将来自不同用户或应用的数据分割成固定大小的数据包,这些数据包分别选择随机的路

径进行传送, 到达目的地以后再重新组合, 采用混淆网络的思想来支持匿名连接, 从而使黑客不能监测到通信双方的身份[6,9,10]。应用软件(如 HTTP) 和洋葱路由代理产生套接层连接, 洋葱路由代理经过好几个洋葱路由器, 建立一个到终点的匿名连接。在传送数据之前, 第一个洋葱路由器基于混淆网络思想随机选取路径, 然后将每个洋葱路由器进行一层加密, 路径上每个洋葱路由器再将加密除去, 直到到达终点。打包在匿名连接路径里的多层数据结构称为洋葱。一旦建立了连接, 数据就能在被选择的路径里传输。基于这种原理的邮件系统称为类型 2的邮件重发器。这种技术的不足是相应的系统结构复杂, 可扩展性、性能与可靠性差, 开发成本高等。

3.3　基于服务商的隐私保护技术

前述的匿名技术其实质是隐藏用户的身份, 从而实现隐私保护目的。然而, 在许多 Web 应用中, 向服务商提供个人信息是必须的, 例如在线购物时, 你必须提供银行账号、联系地址等; 在健康咨询时, 你必须提供病史信息等。由此, 在服务商的网站系统中将收集存储着大量的个人隐私信息, 因而, 服务商有责任和义务采取相应的措施保护其系统中的隐私信息。服务商所采取的措施有虚拟隐私网络(Virtual Private Networks) 和防火墙, 以防止黑客从系统中窃取隐私信息。另外, 服务商在收集用户的隐私信息时, 将其隐私政策公布在网站上, 以提示用户是否同意其收集和使用隐私信息。然而, 众多的网站有着各自不同的隐私政策, 而且很难被用户理解。据权威机构调查显示, 这些隐私政策只有大学文化的用户才能理解。为此, 在 2002 年 4 月, W3C(World Wide Web Consortium) 开发出一个隐私偏好平台 P3P(Platform for Privacy Preferences) 。P3P 使 Web 站点能够以一种标准的机器可读的 XML 格式描述其隐私政策, 包括描述隐私信息收集、存储和使用的词汇的语法和语义。Web 用户可用 APPEL (A P3P Preference Exchange Language) 定义自己的隐私偏好规则, 基于这一规则, 用户 Agent 可自动或半自动地决定是否接受 Web 站点的隐私政策。目前, Netscape 7.0和 Explorer 6.0 均支持 P3P 标准[11~14]。虽然, P3P 使隐私政策向机器可读的方向前进了一大步, 但 Gunter Karjoth 等[14]指出 P3P 仍存在三个方面的不足: ①在语法上, 元素 Recipient, Retention 和 Purpose 未被清楚地分离; ②在语义上, P3P 没有定义是否同一数据元素可被用在多个隐私条款中, 以及 P3P 中的 “anonymizing data ”的语义也不甚明确; ③完全没有对用户 Agent 的指导, 没有指导如何写一个对用户 Agent 友好的隐私政策以及用户 Agent 怎样解释一个隐私政策。除了这些不足外, 本质上, P3P 的作用仅仅是增加了隐私政策的透明性, 使用户可以清楚地理解个人的何种信息被收集、用于何种目的以及存储多长时间等, P3P 本身不能保证 Web 站点是否履行其隐私政策, 因而, P3P 不能完全解决隐私保护问题。

4　网络隐私保护技术的发展方向

网络隐私起源于网络, 因此网络的发展将为网络隐私带来新的问题和挑战。目前, 对等网(P2P) 和语义 Web 是网络发展的两个主要趋势, 基于 P2P 和语义 Web 的隐私保护技术将会成为新的研究方向。

(1) 基于 P2P 的隐私保护技术。P2P 是 Peer-to-Peer, 又称为对等计算, 可简单地定义为通过直接交换共享计算机资源和服务。作为网络计算的形式之一, 它近年来引起了学术界和工业界的广泛关注。P2P 计算提供一个全新的方式, P2P 网络认为所有节点在共享信息方面能力平等, 每个用户可提供分布信息仓储, 而且每个人都可加入网络, 由此变成一个包含分布信息仓储的增长异常迅速的网络。P2P 的应用类型主要有对等计算(如 SETI@ home 和 Folding@ home) 、文件交换(如 Napster, Freenet) 和即时通信(如 ICQ, OICQ 和 AOL IM) 等。在 P2P 结构下, 隐私保护技术研究主要表现在两方面: ①在文件交换等应用中, 共享文件的用户可被跟踪和识别, 人们需要一种匿名的共享服务, 从而提出了基于 P2P 的隐私保护新需求。②在前述的基于中间代理的匿名技术中, 往往具有复杂的系统结构和昂贵的开发维护代价, 而 P2P 结构是由众多的个人电脑组成, 因此采用 P2P 网络结构可克服这一不足。在 P2P 结构下, 一些自愿者的单机可充当路由节点, 以取代前面的复杂结构, 并利用相同的原理来实现匿名。因此, 基于 P2P 的隐私保护技术将成为一个主要的研究方向。

(2) 基于语义 Web 的隐私保护技术。Tim Berners-Lee[15] 指出, 语义 Web 是当前 Web 网的下一步发展方向, 通过规范定义和组织信息内容, 使之具有语义信息, 能被计算机 “理解 ”, 从而更好地与人沟通。在语义 Web 环境下有两个实体: Web 服务和 Web Agent。Web 服务是通过网络以标准的 XML 消息格式访问的一组服务接口; Web Agent 可自动地发现、集成、执行和监控 Web 服务。在 Web Agent 与 Web 服务的交互过程中必将涉及大量的个人隐私信息, 这些隐私信息是在无人介入下被 Web Agent 自动处理的。因而, 语义 Web 环境下的隐私保护将面临着更大的挑战[16, 17]。显然, 现有的技术不能解决这一问题。事实上, 由前所述, 近年来隐私保护技术并没有取得实质性的进展, 其中一个重要的原因是人们将隐私问题与信息安全问题等同起来。实质上, 这两者存在着本质的区别: 前者允许他人获得个人信息而控制他人对个人信息的使用, 而后者是控制他人获得秘密信息; 前者不是一个单纯的技术问题, 而后者可用加密或匿名技术予以解决。因而, 语义 Web 环境下新的隐私保护方法与机制将是又一个主要研究方向。

5　结束语

网络隐私是集社会、法律、技术为一体的综合性概念, 因而, 网络隐私保护必须最大化技术的作用, 并为从法律上解决隐私侵权提供有力的技术支持。一个有效的隐私保护系统应该是: 在未经本人的许可下, 他人不能或无权收集和使用个人的信息。在这一系统下, 隐私信息的收集需要与本人协商, 隐私信息的使用需要得到社会的监督, 隐私信息的侵权需要得到法律的制裁。然而, 近年来各种立法并没能阻止对隐私的侵权, 隐私保护技术的作用也非常有限, 因此还需努力探索真正有效的隐私保护技术。

参考文献:

[1]　bdelmounaam Rezgui, Athman Bouguettaya, Mohamed Y Eltoweissy. Privacy on the Web: Facts, Challenges, and Solutions[J] . IEEE Security & Privacy, 2003, 40-49.

(下转第 28 页)