*Web Privacy with P3P*

# P3P Web 隐私

*Lorrie Faith Cranor* 著

技桥 译

O'REILLY®

清华大学出版社

# P3P Web 隐私

*Lorrie Faith Cranor* 著
技桥 译



O'REILLY®

*Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo*

O'Reilly Media, Inc. 授权清华大学出版社出版

清华大学出版社

**图书在版编目（CIP）数据**

P3P Web隐私/克劳娜（Cranor, F. L.）著；技桥译．－北京：清华大学出版社，2004.5
书名原文：Web Privacy with P3P
ISBN 7-302-07170-5

I. P... II. ①克... ②技... III. 计算机网络－安全技术 IV. TP393.08

中国版本图书馆CIP数据核字（2003）第077265号

北京市版权局著作权合同登记
图字：01-2003-0827号



书　　名/ P3P Web隐私
书　　号/ ISBN 7-302-07170-5/TP · 5227
责任编辑/ 常晓波
封面设计/ Ellie Volckhausen，张健
出版发行/ 清华大学出版社（www.tup.com.cn）
地　　址/ 北京清华大学学研大厦（邮政编码100084）
经　　销/ 各地新华书店
印　　刷/ 北京四季青印刷厂
开　　本/ 178毫米×233毫米　25印张　407千字
版　　次/ 2004年5月第一版　2004年5月第一次印刷
印　　数/ 0001-4000册
定　　价/ 45.00元（册）

# O'Reilly Media, Inc. 介绍

为了满足读者对网络和软件技术知识的迫切需求，世界著名计算机图书出版机构O'Reilly Media, Inc. 授权清华大学出版社，翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly Media, Inc. 是世界上在 UNIX、X、Internet 和其他开放系统图书领域具有领导地位的出版公司，同时是联机出版的先锋。

从最畅销的*The Whole Internet User's Guide & Catalog*（被纽约公共图书馆评为二十世纪最重要的50本书之一）到GNN（最早的Internet门户和商业网站），再到WebSite（第一个桌面PC的Web服务器软件），O'Reilly Media, Inc.一直处于Internet发展的最前沿。

许多书店的反馈表明，O'Reilly Media, Inc. 是最稳定的计算机图书出版商——每一本书都一版再版。与大多数计算机图书出版商相比，O'Reilly Media, Inc. 具有深厚的计算机专业背景，这使得O'Reilly Media, Inc. 形成了一个非常不同于其他出版商的出版方针。O'Reilly Media, Inc. 所有的编辑人员以前都是程序员，或者是顶尖级的技术专家。O'Reilly Media, Inc. 还有许多固定的作者群体——他们本身是相关领域的技术专家、咨询专家，而现在编写著作，O'Reilly Media, Inc. 依靠他们及时地推出图书。因为O'Reilly Media, Inc. 紧密地与计算机业界联系着，所以O'Reilly Media, Inc. 知道市场上真正需要什么图书。

# 目录

# 第三部分 P3P软件与设计

# 序

当前，因特网隐私的问题已经成为了因特网用户最关注的问题之一。对隐私的疑虑使得很多人远离了网络；而垃圾邮件则极大地挫伤了人们对网络的热情，他们开始隐瞒自己的身份以及住处。关于隐私的争论吸引了人们极大的关注，但是到目前为止，在解决隐私方面并未取得任何明显的进展，也没有发生广泛的调整性的变革。

对于这种情况的出现，理由很充分：进展的缓慢在很大程度上归因于“隐私问题”本身的两义性。很显然，有许多种数据是不能进行交易的，例如，在Creditcards.com上应当无法购买我的信用卡号；好事者应当无法从网上找到我在上次检查身体时与医生交谈的内容，诸如此类的数据是不适合在因特网上出现的。通常，对于这些信息的不正当获取是不被允许的。应当制定相应的法律法规以禁止此类事件的发生。

但是无可否认的是，在因特网上还有许许多多其他的数据存在着很多不确定因素。针对此类数据进行商业交易还不能简单地进行立法规范。您可能认为Amazon.com对您在其网站上进行的购物活动实行监控是一种很无礼的行为，这样它常常会建议您购买其他的物品；我也有同感，祝这些公司好运，希望没有太多人与我们持同样的看法。我很讨厌Amazon发来一大堆的电子邮件，告诉我有关该网站近期的商品信息；而您可能会对邮箱中各式各样的信息很感兴趣。人和人是不同的，

在需要体现出差异的地方，我们不希望使用简单而绝对的规则，而是利用简单的技术来达到求同存异的效果。

在早期的Internet消费领域，人们通过对Web站点上所用隐私策略的分析，找到了这种一致。在（美）联邦贸易委员会（FTC）和许多其他倡议者的共同努力之下，Web站点陈述了它们会对收集到的数据进行哪些处理。显示了隐私策略声明的Web站点数量让FTC引以为豪，但同时它自己有关的声明显然对此有所夸大。

而隐私页面的问题，主要在于没有人阅读这些页面，或者说是不能阅读，因为这些页面的内容是以英语（以英文网站为例）和律师使用的语言拼凑而成的。但更重要的是，隐私页面似乎与人们使用因特网的方式毫不相干。没有人会先访问Web页面，然后突然想起去读取相关的隐私策略，这种方式行不通。即使有人这样做了，他们用得到的隐私策略信息做些什么操作呢？在终端保留这些图表以显示不同站点上的不同隐私策略吗？Amazon许诺不会转售电子邮件地址，但Yahoo有过这样的许诺吗？将这种“方案”当作解决所有问题的法宝（大概只有经验不足的律师会这样想）无疑是很可笑的。

上述这些隐私策略声明的“悠久历史”使人们了解到：在Web上，仅在需要个人进行个性化选择的区域出现更多的文字没有任何帮助。真正需要的是能够使机器理解隐私策略的更好的代码，而不是多多地提供给人阅读的文字。也就是说，P3P这种技术使Web站点以一种机器能够理解的方式表达它们的隐私策略，这样用户就可以根据自己的需要对经常访问的站点设置参数。这样一来，计算机代码就替代了“律师”的代码，这样会越来越方便理解。

P3P本身并不能解决计算机领域所有隐私方面的问题。理解只是第一步。即使P3P已被普遍采用，也还是无法彻底消除对于“隐私是否被妥善保护”这一问题的质疑。在市场范围内进行选择并不是一种自我开脱，这些选择需要有法律作为它们的后盾。另外，如果我们逐渐理解到了人们进行选择的方式，我们有可能可以获知，从本质上来说，我们选择的方式都是一样的。如果是这样，那么就再也无需针对某项选择来精心地制定机制了。

但是，无论是理解人们到底想要些什么，还是理解市场会迎合这些需要生产怎样的产品，这都是一条漫漫长路。所以，尽管个人是否会收集数据、个人如何处理这些数据等等问题仍存在不确定因素，但用于降低解除这些不确定因素的成本的技术也得以不断发展。P3P就是这样一种技术，它的出现使我们向理解这些问题的方向迈出了重要的一步。

在本书中，P3P的原创者之一为我们讲述了这项技术。我们有理由相信，对因特网隐私问题感兴趣的读者会从她那清晰生动的陈述中获益不少；对于隐私技术倡导者来说，本书论证了该技术的承诺以及局限性；而对于技术者而言，本书则为他们介绍了该技术实施方式的一个清晰而完整的画面；最后，对于策略制订者来说，本书则论证了一个独特的技术，列举了如何通过该技术实施策略的示例。

最后的教训往往也是最重要的，这不仅仅适用于我们所讨论的颇有价值的隐私技术，同样也适用于其他有价值的诸多事物。在未来的计算机领域，用于构建该领域的技术将会制订自己的策略。代码即法律（code is law）。如果策略的制订者不希望其地位被技术取代，那么应该好好审视一下代码和法律之间的这种交互关系。在数千年的社会生活中，形成了一些颇为重要的价值，隐私就是其中之一。总地来看这些价值也罢，单看隐私这个价值也罢，从根本上讲，它们是否将继续成为计算机领域中不可分割的一部分，都要根据计算机领域的构建方式而定。人们需要制订一些选择方案，而这些选择方案应该由那些了解选择方案本质的人来制订。

Lawrence Lessig

2002年5月

但是，无论是理解人们到底想要些什么，还是理解市场会迎合这些需要走产怎样的产品，这都是一条漫漫长路。所以，与第一个人是否会收集数据、个人如何处理这些数据等等问题仍存在不确定因素，但用于降低解除这些不确定因素的成本的技术也将以不断发展。P3P就是这样一种技术，它的出现使我们向理解这些问题的方向迈出了重要的一步。

在本书中，P3P的原创者之一为我们讲述了这项技术。我们有理由相信，对因特网隐私问题感兴趣的读者会从他那精辟生动的陈述中获益不少；对于隐私技术倡导者来说，本书论证了该技术的来源以及局限性；而对于技术者而言，本书则为他们介绍了该技术实施方式的一个清晰而完整的画面；最后，对于策略制订者来说，本书则论证了一个独特的技术，列举了如何通过该技术实施策略的示例。

最后的数则注往也是最重要的，这不仅仅适用于我们所讨论的颇有价值的隐私技术，同样也适用于其他有价值的许多事物。在未来的计算机领域，用于构建该领域的技术将会制订自己的策略。代码即法律（code is law）。如果策略的制订者不希望其地位被技术取代，那么应该好好地——与代码和法律之间的这种交互关系。在数千年的社会生活中，形成了一些颇为重要的价值，隐私就是其中之一。总拥来看这些价值也罢，单看隐私这个价值也罢，从根本上讲，它们是否将继续成为计算机领域中不可分割的一部分，都要根据计算机领域的构建方式而定。人们需要制订一些选择方案，而这些选择方案应该由那些了解选择方案本质的人来制订。

Lawrence Lessig

2002 年 5 月

# 前言

隐私偏好设定平台（Platform for Privacy Preferences，P3P）是Web站点以及与站点相连的客户隐私策略的通信纽带。Web客户可以使用P3P从Web服务器获取可机读的隐私策略并作适当的响应。本书内容涵盖了P3P的发展历程、让Web站点支持P3P的方式以及如何设计P3P应用程序。

我第一次接触可机读的隐私标签的思想始于1996年10月，那大概是在我完成学位论文并开始在AT&T Labs-Research工作之后的两个星期。那时我还没有意识到历时五年半的历程就是从那时开始的。

我从事的这项工作，始于一个促进个人隐私状况发展的个人愿望，并最终发展成为众人瞩目的P3P。显然，对于许多非常重要的在线或离线隐私问题，P3P未必会产生多大影响。然而，我一直都持有乐观的态度，相信在各方的努力下，P3P必将获得数据隐私在线保护的一个重大突破。该目标的实现是必然的，因为现在仍处于采纳和使用P3P的初期。目前，虽然P3P 1.0规范幕后的技术工作已经完成，但是还有很多工作需要进行：敦促Web站点广泛采用P3P、为用户研制和部署有用的P3P工具，以及推广有关数据隐私的客户教育。毫无疑问，本书将会作为所有这些努力的有用的指导。

我在本书编入了大量的P3P历史，并尝试着提供一个精确的说明。同时，我还在P3P邮件列表收集、个人笔录整理以及与早期从事P3P的工作人员交谈中耗费了

大量的时间。一些读者可能会问，是否有关那些会议的组织者和筹划人的细节真地那么举足轻重。这可能看起来有些不合理，但是我认为，在此讲述他们是因为：我相信理解科技技术标准的创建过程可以开拓眼界。

## 哪些人应该阅读本书

本书是为广泛的读者群所著。我尽力去回答各种各样的有关P3P的问题，这些问题是由那些关注如何保护在线隐私的Web站点运营人员、软件开发人员、策略制作人员和个人提出的。如果您对在Web站点上或软件产品中部署P3P感兴趣，或者仅对P3P的概念和工作方式感到好奇，本书都很适合您。

读者在本书阅读前应具备一定的Web浏览器的使用经验。本书的某些部分比较适合具有一定编程基础的读者阅读，但是，书中的大部分内容都适合于仅有计算机基础知识的读者阅读。

## 本书的组织结构

本书由4部分组成。

第一部分“隐私和P3P”，回顾了在线隐私关注以及用于解决这些关注的各种技术上的、规范以及自我规范的倡议。本部分包含了P3P简介，以及P3P规范和规范制作动机的历史描述。共分为4章：

第一章，P3P简介
第二章，在线隐私概览
第三章，隐私技术
第四章，P3P历史

第二部分“使Web站点支持P3P”，为Web管理员提供了在站点上使用P3P的所有信息。它包括：在Web站点上启用P3P的步骤概述，创建P3P文件的方法，以及支持P3P的Web站点的实例。同时，这部分内容还包含了P3P词汇所有元素的详细解释以及P3P基础数据模式。共分为以下6章：

第五章，概述和选项
第六章，P3P 策略语法
第七章，创建 P3P 策略
第八章，创建并引用策略引用文件
第九章，数据模式
第十章，支持 P3P 的 Web 站点示例

第三部分“P3P软件和设计”，讨论了与P3P词汇和P3P软件相关的设计问题。这些章节介绍了P3P词汇幕后的一些设计原理，以及关于将词汇转化到P3P用户代理界面中的方法。它们描述了各种不同的P3P实现，介绍了APPEL语言，同时也提出了解决用户界面问题的建议。

本书的该部分主要面向那些要在自己产品中内置P3P的软件开发者，以及要开发P3P偏好设置的个人。然而，这些章节也将本书前面部分介绍的概念紧紧地串在了一起，另外，它们还演示了如何在P3P软件实现中组织所有的工作。本部分共分为以下4章：

第十一章，P3P 词汇表设计问题
第十二章，P3P 用户代理和其他工具
第十三章，P3P 偏好交换语言（APPEL）
第十四章，用户界面

第四部分“附录”，提供了来自其他资源的技术细节和材料。共分为5个附录：

附录一，P3P 策略和策略引用文件的语法快速参考
附录二，配置 Web 服务器以包含 P3P 报头
附录三，IE6 中的 P3P
附录四，如何为 IE6 创建自定制隐私导入文件
附录五，P3P 指导准则

既然不可能让所有的读者都喜爱本书，那么就不能针对所有人来编写本书，因此，我希望大多数读者至少能够快速地浏览一下本书的所有章节。例如，软件开发人员可能对第二部分和第三部分感兴趣，而我认为他们将会从第一部分所提供的有

关隐私问题的背景知识中找出一些有价值的重要信息，他们可以依据这些信息考虑如何为自己开发的工具设计用户界面和默认设置。

如果您对在Web站点上配置P3P很感兴趣，那么就可能需要快速查看第一部分的内容，然后着重阅读第二部分内容。对第三部分的快速阅读将会帮助您理解用来查看Web站点上P3P策略的P3P用户代理的范围。

如果您对为P3P用户代理部署软件、策略生成器或其他P3P工具很感兴趣，那么您应该阅读整本书，并重点阅读第六章、第八章、第九章、第十二章、第十三章和第十四章。

如果您是一个关注理解P3P工作方式和工作原理的策略制订者，或者是一个仅对P3P好奇的人，那么在第一部分和第十二章以及第十四章中可以找到您感兴趣的内容。第五章和第六章将会使读者对P3P策略有更深层次的理解，第十一章则介绍了关于P3P词汇表的更多深入性的知识。如果您想一展身手，想尝试使用APPEL语言手动创建P3P偏好文件的话，那么还需要仔细阅读第十三章。

## P3P规范一致性

隐私偏好设定平台1.0（P3P 1.0）规范（*http://www.w3.org/TR/P3P*）是关于P3P协议和词汇信息的权威来源。在确保本书的准确性及其与P3P规范的一致性方面，作者已经尽了全力。然而，如果本书与P3P规范之间存在出入的话，那么请以P3P规范为准。（也请您将有出入的地方告诉我们!）

P3P规范使用术语“MUST（必须）”来表明必要条件。违背该要求的P3P实现或支持P3P的Web站点将被认为是与P3P不兼容的。在本书中，当提及该站点或P3P用户代理“必须”做某事时，都是因为规范要求这么做。

P3P规范也使用了“SHOULD（应该）”来表明条件性需求——如果有正当的理由，那么也可以不遵守这些需求。违背一个或多个条件性需求（但满足所有MUST条件）的站点或实现被认为是“条件性兼容”。本书中在描述条件性需求时使用的术语是“应该”。然而有时也使用“应该”来描述作者推荐的处理方式，而P3P规范中并没有做必要的要求。