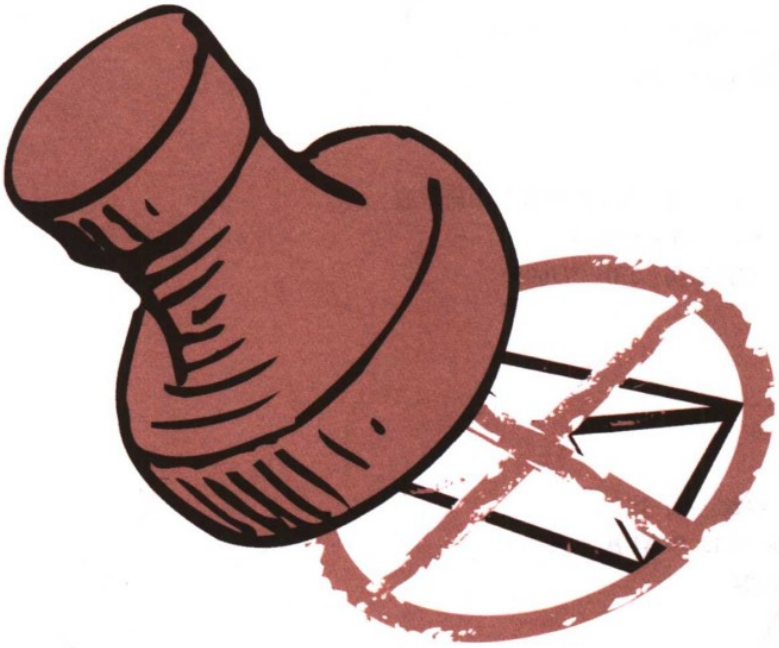
反垃圾@邮件完全手册

陈勇　李卓桓　编著

清华大学出版社

反垃圾@邮件完全手册

陈勇 李卓桓 编著



清华大学出版社
·北京·

内 容 简 介

本书在对垃圾邮件的产生机理进行分析和总结的基础上，对各种反垃圾邮件技术进行了全面介绍，包括反垃圾邮件、反邮件病毒、内容过滤等功能，本书还介绍了邮件溯源和邮件审计、日志及日志分析等功能。

本书对这些相关技术进行了全面分析和介绍，可以作为邮件安全产品用户、邮件安全产品开发人员、邮件安全品测试人员和反不良信息监控机构的重要参考资料。

图书在版编目(CIP)数据

反垃圾邮件完全手册/陈勇等编著. —北京：清华大学出版社，2006.12
ISBN 7-302-14021-9

Ⅰ. 邮…　Ⅱ. 陈…　Ⅲ. 电子邮件－安全技术　Ⅳ. TP393.098

中国版本图书馆 CIP 数据核字(2006)第 123010 号

出 版 者：清华大学出版社　　地　　址：北京清华大学学研大厦
　　http://www.tup.com.cn　　邮　　编：100084
　　社 总 机：010-62770175　　客户服务：010-62776969
组稿编辑：夏兆彦
文稿编辑：刘　霞
印 刷 者：北京鑫丰华彩印有限公司
装 订 者：北京鑫海金澳胶印有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×230　印张：19.75　字数：433 千字
版　　次：2006 年 12 月第 1 版　2006 年 12 月第1 次印刷
书　　号：ISBN 7-302-14021-9/TP·8424
印　　数：1～4000
定　　价：35.00 元

陈勇

(网名OKboy)

毕业于清华大学电子工程系，网际江湖科技有限公司创始人，从事IPTV/网络视频及P2P应用的研究工作。曾任清华紫光、比威网络等公司技术总监、启明星辰高级技术顾问，长期从事网络安全、互联网应用的研发和探索工作。主持国家863关于反垃圾邮件的项目，协助国家相关部门保障邮件安全工作及反垃圾邮件标准制定工作。

李卓桓

（网名zixia，http://zixia.net）：

毕业于清华大学，现任合一网络技术有限公司（YoQoo.com）首席科学家。曾任职于 Tom.com、ChinaRen .com 等多家知名网站及恒信数码、清华紫光比威等网络技术公司，长期致力于互联网研发工作，特别是在反垃圾邮件领域有深入研发经验。

封面设计：装帧设计 010)86390064

谨以此书献给我的家人、朋友、同事及所有支持我们完成此书的人们。

业界推荐

1. “电子邮件已经成为我们日常工作、生活的重要交流工具。面对日益增多的垃圾邮件，如果没有对应有效的技术进行治理，将会对我们的工作、生活造成很大的不便。所以，本书中所讨论的反垃圾邮件技术是当前互联网中一个重要的话题，应该被我们重视起来。”

——古永锵（YOQOO.COM，曾任搜狐公司总裁，现任 YoQoo.com CEO）

2. “垃圾邮件不仅仅是邮件骚扰的事情，已经关系到信息安全的诸多方面，甚至成为导致网络安全和机密窃取的一个重要途径，成为很大的安全隐患和网络公害。本书是我见到的难得的兼顾技术原创性和可操作性的反垃圾邮件技术书籍。”

——严力（VENUSTECH.COM.CN，启明星辰公司总裁）

3. “陈勇是对技术有着深刻理解和执著追求的人。透过他的图书，我们不仅能够了解到反垃圾邮件的相关技术，更能够感受到他对技术精益求精的工作态度和研究方法。”

——王小川（SOHU.COM，搜狐副总裁）

4. “本书全面介绍了反垃圾邮件的各种技术，对国内互联网相关企业的技术管理和日常运营有非常大的参考价值，具有较高的可读价值。”

——李旸（KONG.NET，空中网副总裁）

5. “这本书全面、深入浅出地介绍了反垃圾邮件相关技术，有很高的阅读和参考价值。”

——章文嵩　副教授（LinuxVirtualServer.ORG，已收录 Kernel 的国际知名开源项目 LVS 创始人）

6. “This book is an amazing valuable addition to anybody's collection that is interested in email fraud, spams and internet abuse. It is a must have for people hope to expand their horizons in anti-spam techniques.”

——Halen Heng Ji（NYU.EDU，CS Ph.D）

7. “电子邮件是现代人日常生活中不可或缺的沟通手段，但垃圾邮件的出现给每一个电子邮件的日常使用者带来了很多的麻烦。对于普通用户来说情况还稍好一些，但对于一些跟电子邮件相关的应用，例如电子邮件转手机短信，那么用户受到的就不只是困扰，而是强烈的骚扰了，而一般来说受到这样的骚扰用户还需要付费。不幸的是由于工作的关系，本人必须考虑所设计的电信级设备采用何种方案在保证系统最大性能、最少的误判、漏判的情况下解决垃圾邮件的问题。《反垃圾邮件完全手册》的出现向我提供了很好的参考资料，此书对于垃圾邮件的论述相当全面、分析相当独到，在此书的帮助下，很快确定了特定电信应用环境下的垃圾邮件解决方案，为系统在世界各地的稳定运行提供了很好的技术保证。感谢作者为我们带来了这本好书，相信它同样能给你带来帮助。

——陈新宇（中兴通讯业务产品总工，www.zte.com.cn）

8. “今天，每一个接触 Internet 的人都要受到垃圾邮件的骚扰；大多数人消极防备，为分辨和删除垃圾邮件而浪费了很多时间。其实，我们可以选择许多方式来主动地处理垃圾邮件，从而节省时间和网络带宽。本书详细介绍了反垃圾邮件的各种有效手段。这些手段是将我们从被动防守转向主动出击的法宝。如果你正在为垃圾邮件的骚扰而头疼，本书绝对值得一读！”

——魏永明（MINIGUI.COM，国际著名嵌入式开源软件项目 MiniGUI 创始人，飞漫软件创始人）

9. “本书对反垃圾邮件技术进行了全面的介绍，既有深入的理论分析，又有生动的实例，并以 Apache 软件基金会的著名开源产品 SpamAssassin 和国内的商业反垃圾邮件系统为例进行了实战讲解，是邮件系统管理员和开发者必备的参考手册。”

——邵辉（MOP.COM，千像公司猫扑网网络运营总监）

10. “垃圾邮件是困扰互联网应用的重大难题。本书从垃圾邮件发生的基本原理出发，透彻分析了垃圾邮件攻防的基本思想，并系统介绍了反垃圾邮件的主要技术原理及相关的产品。不同于一般流行的介绍反垃圾邮件的图书，本书特别强调根据垃圾邮件的成因有针对性地选择可能的技术，或拦截、或追踪，而不是盲目地选择软件产品。这一点使得本书不仅对于反垃圾邮件的工程实践大有裨益，也适合于正在进行垃圾邮件有关技术研究工作的专业人士参考。同时，对于希望了解垃圾邮件和反垃圾邮件技术发展全貌的人士，本书也是一本相当全面和周到的入门读物。”

——陈茂科　博士（http://www.net-glyph.org/～cmk，清华大学网络中心/BBS 水木清华站 smth.edu.cn 技术组组长）

11. “本书全面系统地介绍了反垃圾邮件的相关理论和可行性技术实现，无论对终端产品用户、相关软件测试开发人员还是信息安全监控机构都具有很高的参考价值。”

——姜太文 博士后（XOOPS.ORG.CN，国际著名开源软件项目XOOPS项目管理员）

12. “该书较全面系统地介绍了反垃圾邮件技术，结合作者多年的‘实战’经验，提出了比较实用的反垃圾邮件方法。”

——张鸿 博士（WWW.CERT.ORG.CN，国家计算机网络应急技术处理协调中心高级工程师）

13. “作者凭借着其丰富的反垃圾工作经验，在书中全面介绍了反垃圾邮件的各项技术。如果你希望了解互联网邮件安全，那么一定要阅读本书，它可以使你了解到丰富的理论和实践知识。”

——池宇峰（WORLD2.CN，共恩软件董事长，完美世界董事长）

14. “该书对邮件的众多反垃圾技术进行了分析，分析了各种技术的思想，实现和优缺点，从事电子邮件相关工作的人可以从中看到反垃圾邮件技术方向的较全面的信息。”

——周霖（SOHU.COM/NewSMTH.Net，搜狐网络运营总监，水木社区运营总监）

15 “如果你用电子邮件而且关心邮件的安全，那么这是一本不可不读的好书。”

——张本宇（MICROSOFT.COM/BBS.PKU.EDU.CN，微软亚洲研究院研究员，原北大未名BBS站长）

16. “本书详细介绍了反垃圾邮件的背景知识及相关技术，提出并实现了一套基于可追查性检查的反垃圾邮件方案，具有极强的实践指导意义。”

——俞欢 博士（WWW.MATH.PKU.EDU.CN，北京大学数学学院）

17. “本书较为全面、详尽介绍了当前反垃圾邮件的关键技术，是网络运营工作中不可多得的一本好书。”

——姚键（YOQOO.COM，优酷网首席架构师）

18. “这本书较为全面深入地介绍了反垃圾邮件相关技术，具有较高的参考价值和广泛的现实指导意义。”

——程宇（TOM.COM，TOM在线网络支持中心安全部门经理）

19. “如果你想从实践的角度快速了解反垃圾邮件技术的方方面面，构筑自己的反垃圾邮件系统，本书是你不错的选择。”

——吴寿（YTHT.NET，北大一塌糊涂BBS创始人）

PREFACE

序 一

电子邮件是互联网主要的应用之一，极大地方便了人们通信与交流。然而，垃圾邮件的产生，影响了正常的电子邮件通信，占用了传输带宽，大量宝贵的网络带宽被无效、甚至有害的垃圾邮件所拥塞，当垃圾邮件或病毒邮件爆发时，经常收到用户无法连接互联网的报告，这严重浪费了网民的时间，甚至许多用户的邮箱中所收到的正常邮件占所有邮件的比例不足10%，也就是说有些用户处理邮件的90%以上的时间被浪费了。给网民带来了烦恼。电子邮件服务企业，为治理垃圾邮件投入了大量的人力、物力、资金和技术，增大了网络企业运营的成本。垃圾邮件的外延也在扩展，包含了病毒、蠕虫、特洛伊木马的邮件也进入了垃圾邮件的范畴，垃圾邮件发送技术、病毒蠕虫的传染技术、各种入侵技术在相互配合、攻城略地。事实上，随着网络防范技术的提高，直接入侵系统已越来越困难，特洛伊木马成为机密泄露的主要原因，而垃圾邮件恰恰是这类恶意程序传播的一个重要渠道和帮凶。近日闹得沸沸扬扬的工商银行账户被盗事件都不是直接入侵了工商银行的系统，而是特洛伊木马盗取了用户的用户名和口令发送给了犯罪嫌疑人。中了蠕虫病毒后可以自动发送垃圾邮件，垃圾邮件可以携带恶意程序，特洛伊木马、基于邮件的钓鱼网站成为机密泄露的主要原因，一个个不知情的受害者成了犯罪嫌疑人的帮凶……

近年来我国为治理垃圾邮件做出了不懈的努力，中国互联网协会2002年11月成立了民间的反垃圾邮件协调小组。推出了《中国互联网协会反垃圾邮件工作规范》，明确了垃圾邮件的定义和《拒收垃圾邮件指南》。从定期公布黑名单发展到时时公布黑名单。自2004年1月国家四部委联合发文要求整治垃圾邮件和不良信息到2006年3月信息产业部出台的《电子邮件服务管理办法》。应该说，从民间的行业自律到政府的行政法规的规范，还有网络企业采取的大量技术手段。社会各界对垃圾邮件的治理是全方位的。但据最近的统计，全球发送垃圾邮件的第一名是美国、第二名是中国。由于巨大的利益驱动和对垃圾邮件定义理解的不同，给治理垃圾邮件造成了一定的困难，在技术上对垃圾邮件的治理呈现了“道高一尺，魔高一丈”的情况，给垃圾邮件的治理带来了挑战，因此垃圾邮件

的治理是一项长期的工作，任重道远。

特别是反垃圾邮件技术和产品，是反垃圾邮件的基础。了解垃圾邮件产生的机理，系统掌握垃圾邮件的技术，对治理垃圾邮件是很重要的环节。本书从邮件系统的原理和漏洞、垃圾邮件产生的原因和传播方法、对抗垃圾邮件传播的原理和方法、甚至垃圾邮件发送者对抗反垃圾技术等诸多方面整体地论述了反垃圾邮件的技术，是目前对反垃圾邮件技术介绍比较全面、比较深入的一本好书。

本书作者陈勇先生1997年开始研究网络安全技术，2000年开始研究反垃圾邮件技术，是资深的网络安全专家，是国内最早研究反垃圾邮件技术的专业人员，2005年中国互联网大会上陈勇先生的“可追查性检查技术”和2006年中国互联网大会上陈勇先生的“增强社会责任感 防止垃圾邮件外发”的演讲，都是很有独到见解的话题，对整个反垃圾邮件的工作具有很好的指导作用。陈勇先生能够从网络安全的各个层面、各个角度考虑问题，而不仅仅限于内容或邮件本身。

网络安全的各个方面需要互相配合，客户端、服务器、网关设备、路由设备、内容、行为、网络协议、安全厂商、用户等等，这样的立体布局比较合理和全面，希望广大读者通过阅读这本书，对反垃圾邮件工作能够有所帮助。对中国的反垃圾邮件事业有所帮助。也感谢陈勇先生多年来为中国的反垃圾邮件事业所做出的贡献。希望我们大家能携手共同努力促进互联网事业的健康发展。

黄澄清（WWW.ISC.ORG.CN，中国互联网协会秘书长）

2006年10月8日

PREFACE
序 二

我相信每个长期使用电子邮件的人都为近年来大肆泛滥的垃圾邮件所不胜烦扰。这些垃圾邮件中不仅有大量商业广告，还夹杂了很多违法和对青少年有害的内容，不但受到广大用户的痛恨，也引起各国政府的重视。然而，由于界定和分辨垃圾邮件的复杂性，以及垃圾邮件发送手段的提高及其与其他不良软件（malware）的合流，使得反垃圾邮件技术一直面临着“道高一尺，魔高一丈”的压力。

陈勇先生毕业于清华大学电子系，大学期间就积极从事科技创新活动，多年来一直保持着对产业技术的高度敏感，并较早就开始关注和研究垃圾邮件问题。他所发明的专利技术先后被国内多家安全厂商作为核心技术应用于反垃圾邮件产品，为推动我国在相关领域的自主创新起到了一定作用。本书就是陈勇先生多年钻研反垃圾邮件技术中对资料的分类整理、对技术的系统分析，以及他个人心得和构想的阐述。从垃圾邮件的起源和手段，到反垃圾邮件的主要技术和常见工具，以及相关的标准和法律动向，本书对于系统管理人员、产品研发人员和其他与反垃圾邮件有关联或有兴趣的人，都应有很好的参考价值。

无论是查表法（黑白名单、RBL 等）、溯源法（DomainKey、SenderID 等）、统计法（贝叶斯等）、过滤法或其他方法，虽不能一举完成防堵垃圾邮件的大业，有的甚至不能解决垃圾邮件带来的任何网络带宽和存储空间浪费，却都在不同环境下、不同程度上发挥着很好的作用。正如陈勇先生在本书中指出的，真正要从根本上解决问题，还是要法律（社会力量）与协议（网络机制）双管齐下。协议的认证和溯源机制使得立法真正有威慑力量，执法真正有查证依据。本书的出版，定会对推动这项工作发挥相当的作用。

李军　博士（清华大学信息科学技术学院副院长，清华大学信息技术研究院院长）

2006 年初秋

PREFACE
序 三

我每天都收到垃圾邮件，相信每个使用过一段时间 e-mail 的用户都是这样。毫不夸张地说，垃圾邮件已经成为公害，但是由于界定垃圾邮件的难度和网络协议本身存在一些天生弱点，一直没有简单有效的办法杜绝垃圾邮件的泛滥。我本人大学学习的是计算机科学，毕业后一直在互联网领域工作，对垃圾邮件这个话题特别关注，我发现这本书的内容深入浅出，非常值得大家阅读。作者李卓桓曾经是我的同事，也是我的大学校友，他对计算机底层原理特别是互联网的了解和实践应用涉猎范围的广泛，令我非常惊讶，很高兴看到他能有这个作品。本书不仅谈到了解决垃圾邮件的一些崭新方法和实践，而且比较详尽地介绍了周边的知识和应用内容，使得读者能够在铺垫下逐步深入核心内容。这样的结构安排十分体贴和精致。

这是一本知识与实践兼备，深度与广度并重的优秀作品，我真心地把此书推荐给大家。

曾祎安

猫扑网(MOP.COM)，千像集团联合创始人/执行董事/首席运营官

FOREWORD
前 言

电子邮件大大降低了信息的传播成本，但是垃圾邮件也随之产生。笔者作为国内互联网的较早用户，使用的是自己注册域名的电子邮件地址，从开始使用至今一直没有变过，该邮件地址在国内互联网发展初期曾注册过许多网站服务，因此成为垃圾邮件的目标。

笔者在2000年起每天收到近百封垃圾邮件，因此从那时起开始关注垃圾邮件的控制方法。由于市场一直未有很强劲的需求，因此这项研究一直处于断断续续的状态。

2004年1月底，国家四部委联合发文整治垃圾邮件和不良信息，快速启动了国内反垃圾邮件的市场，反垃圾邮件产品风起云涌，也出现了一些分析垃圾邮件和反垃圾邮件技术的图书。但无论是厂商宣传的技术还是这些图书中分析的技术，都很难从根本上解决垃圾邮件问题，因此笔者分析、汇总了自己多年了解的技术和行业推荐的技术，整理出一本比较全面的反垃圾邮件技术书籍。本书在对垃圾邮件的产生机理进行分析和总结的基础上，对反垃圾邮件的各种技术进行了比较全面的介绍。

笔者认为反垃圾邮件仅仅是这类产品功能的一部分，这类产品应该总称为邮件安全产品，至少应当包括反垃圾邮件、反邮件病毒、内容过滤等功能，还可能包括邮件溯源和邮件审计等功能。当然，作为一个安全产品，日志及日志分析是不可缺少的一部分。本书对这些相关的技术做了比较全面的分析和介绍。

此外，本书还对邮件安全控制中涉及的法律等非技术问题进行了简单的探讨。

本书供如下几种读者参考。

- 邮件安全产品用户：本书为他们提供比较邮件安全产品功能、性能的技术基础。目前许多用户不了解邮件安全产品，常常只是被厂商所左右。通过阅读本书，读者可以对邮件安全产品有一个更加清晰和全面的认识，从而更好地选择和使用邮件安全产品。
- 邮件安全产品开发人员：本书为他们提供改进其邮件安全产品的技术参考。

- 邮件安全产品测试人员：本书为他们提供了一个相对全面客观的测试方法以及应考虑因素的参考。目前笔者了解的一些测试实验室的测试方法过于片面，不能很好地反映产品的实际效果和性能，因此有必要在了解反垃圾邮件技术原理的基础上，对现有的测试方法进行改进。
- 反不良信息监控机构：本书可以作为良好监控、全面监控、有效监控的有益参考。

本书于2004年9月已基本成稿，但由于工作繁忙一直无法成书，感谢瞿华先生对本书的大量后期整理和补充工作，他的工作才使本书真正完稿得以出版。

由于反垃圾邮件技术处于发展阶段，我们的研究尚在探求过程中，本书内容难免有不足之处，欢迎读者朋友与我们讨论，共同推进我国邮件安全技术的发展。本书的进一步补充和勘误将在 www.tup.tsinghua.edu.cn 和 http://nospam.cn 上随时更新，相关的软件也将在网站上提供下载，读者的任何意见和建议可给我们发邮件：book@nospam.cn

作　者

2006年7月

3.3.2 Foxmail …… 47
3.4 常用免费 Web 邮箱反垃圾设置 …… 53

第 4 章 电子邮件工作原理简介 …… 55

4.1 电子邮件的历史 …… 55
4.2 TCP/IP 协议简介 …… 57
4.2.1 OSI 网络模型 …… 57
4.2.2 TCP/IP 协议概览 …… 60
4.2.3 IP 协议 …… 61
4.2.4 TCP 协议 …… 63
4.2.5 其他协议 …… 63
4.3 DNS …… 64
4.3.1 DNS 的历史 …… 64
4.3.2 DNS 概述 …… 64
4.3.3 常见的域名信息类型 …… 67
4.3.4 手工查找 DNS 域名 …… 67
4.4 电子邮件工作原理 …… 69
4.4.1 电子邮件的格式 …… 69
4.4.2 电子邮件发送的基本过程 …… 72
4.5 SMTP 协议简介 …… 74
4.6 电子邮件的安全缺陷 …… 75
4.6.1 模拟 SMTP 收发过程 …… 76
4.6.2 缺陷分析 …… 77

第 5 章 传统反垃圾邮件技术(上) …… 79

5.1 基础工作：关闭开放式转发和打开发件认证 …… 79
5.1.1 Exchange Server 5.5 服务器的设置 …… 80
5.1.2 Sendmail 服务器的设置 …… 81
5.1.3 qmail 服务器设置 …… 82
5.2 基础工作：关闭匿名代理 …… 82
5.2.1 wingate 软件设置用户身份认证 …… 83
5.2.2 squid 设置用户身份认证 …… 84
5.3 传统技术：静态黑名单和静态白名单 …… 85

5.4 传统技术：静态内容过滤 …… 85
5.5 传统技术：实时黑名单及其改进 …… 86
5.5.1 实时黑名单 …… 86
5.5.2 MAPS …… 87
5.5.3 SpamCop Blocklist …… 89
5.5.4 SURBL …… 89
5.5.5 HABEAS …… 90
5.5.6 NJABL …… 91
5.5.7 SORBS …… 91
5.5.8 OPM …… 92
5.5.9 Spamhaus XBL+SBL …… 93
5.5.10 RFC-Ignorant blacklists …… 93
5.5.11 DSBL …… 95
5.5.12 AHBL …… 96
5.5.13 BSP …… 97
5.5.14 OpenRBL …… 98
5.5.15 SenderBase …… 98
5.5.16 小结 …… 102

第 6 章 传统反垃圾邮件技术(下) …… 103

6.1 数量控制：带宽/连接限制 …… 103
6.2 数量控制：邮件重复限制 …… 104
6.3 新型技术：贝叶斯分析 …… 104
6.3.1 贝叶斯过滤算法的基本步骤 …… 105
6.3.2 贝叶斯过滤算法举例 …… 106
6.4 新型技术：分布协作的内容指纹分析 …… 107
6.4.1 DCC …… 108
6.4.2 Razor …… 109
6.4.3 Pyzor …… 110
6.5 辅助工作 …… 111
6.6 反-反垃圾技术方法 …… 114
6.7 彻底解决反垃圾邮件问题：源头认证 …… 115
6.8 质询-回应技术 …… 116