垃圾邮件过滤技术综述

张尼[1,2]，方滨兴[1,2]

（中国科学院计算技术研究所，北京，100080；中国科学院研究生院，北京，100039）

摘　要： 易用、开放和基于信任的设计理念使电子邮件成为当今最重要的通信方式，也造成 Internet 上的垃圾信息泛滥成灾，垃圾邮件过滤逐渐成为国际信息安全研究领域的热点。本文首先对垃圾邮件过滤技术的研究历程作简要回顾；然后从内容过滤，接入过滤，行为过滤这三方面对垃圾邮件过滤技术的研究现状进行全面综述；最后总结现有过滤技术，同时论述目前研究中遇到的新的问题和挑战，并对未来的研究方向进行展望。

关键词： 垃圾邮件；内容过滤；接入过滤；行为过滤

A Survey of Anti-Spam Technology

ZhANG Ni[1,2], FANG Bin-xing[1,2]

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, P.R.China;
Graduate School of the Chinese Academy of Sciences, 100039, Beijing, P.R.China)

Abstract: Easiness, openness and trust which make email popular also make spam out of control. Anti-Spam technology has become one of the hot topics among the information security fields all over the world. Firstly, this paper reviews the history of Anti-Spam technology. Secondly, all kinds of Anti-Spam technology including content filtering access-filtering behavior-filtering are introduced. Finally, the new problems and challenges which arise from current research are discussed and some suggestions for future research work are put forward.

key words: spam; content-filtering; access-filtering; behavior-filtering

1 引言

电子邮件系统是互联网的一个成功典范，它诞生时间不长却给人们的工作和生活带来了深刻的变化。然而，电子邮件系统在给人们提供便捷通信手段的同时，也导致互联网上的垃圾信息泛滥成灾。根据美国 Bright Mail 公司[1]的调查显示：到 2004 年 7 月，全球垃圾邮件占邮件总数的 65%以上，且仍呈增长趋势；全世界的企业每年大概要花费 80 亿～100 亿美元来解决垃圾邮件问题。这些不请自来的信息大肆污染网络环境，占用大量传输、存储和运算资源，对互联网的发展以及广大用户的利益构成严峻的威胁。

迄今为止，垃圾邮件在国际上并没有一个标准的定义，造成这一情况的原因是：定义是为了满足团体利

作者简介：张尼（1979-），男，吉林双辽人，博士研究生。Email:zhsmart@software.ict.ac.cn；
方滨兴（1960-），男，江西万年人，博士生导师，博士。

益和研究的需要[2]，不同的团体对于垃圾信息有不同的评判标准。不过，研究者们普遍认为垃圾邮件具有如下基本特征：(1)信件未经接收人允许，(2)不请自来。信件数量巨大且内容相似。(3) 信件中隐藏或伪造发件人身份、地址、标题信息，且内容以商业性质居多。

目前，已有多篇文献[3、4、5]从不同角度介绍了垃圾邮件过滤技术的研究进展。文献[3]侧重于内容过滤技术。文献[4、5]等虽对多种垃圾邮件过滤方法进行综述，但普遍缺乏清晰的分类，系统化的总结，同时未能涵盖最新的研究成果。本文试图克服以上文献的不足，全面、系统地介绍垃圾邮件过滤技术。本文第 1 节简要回顾垃圾邮件过滤的研究历程。第 2、3、4 节分别从内容过滤，接入过滤，行为过滤这三方面对垃圾邮件过滤技术的研究现状进行综述。最后，对现有技术进行总结，同时论述目前研究中遇到的新的问题和挑战，并指出未来发展方向。

2 垃圾邮件过滤技术的发展历程

垃圾邮件几乎是和电子邮件系统一同成长起来的，其历史可以追朔到 1975 年，John Postel 在文献[6]中首次提到邮件协议中缺乏选择性拒绝服务器机制。1994 年 4 月 12 日美国的 Arizona 州两位从事移民签证咨询服务的律师将广告信发到 6000 多个新闻组，为垃圾邮件的盛行揭开序幕。国际上对于垃圾邮件的研究是从 90 年代中期开始的，当时主要是采用黑-白名单技术[7]。

90 年代末，在巨大的商业前景和重要的学术价值的驱动下，垃圾邮件问题吸引了许多从事交叉学科研究的技术人员的关注。机器学习、模式识别等先进的研究经验都被引入到这一领域，这一阶段的研究成果成为近几年国内外开发反垃圾邮件产品的主要技术依据。

但是随着问题的深入，研究人员发现在服务器或客户端实施内容过滤并非解决问题的唯一之道。因此，近年来一些研究转而专注其他新颖的方法。哥伦比亚大学的 Salvatore 提出了 MET 模型[8]，通过统计帐户活动规律来检测用户异常和错误的发送行为。Kenichi 使用 DMC 技术，通过文档密度的显著差别来识别垃圾邮件[9]。Boykin 提出了基于社会网络垃圾邮件过滤方法[10]，并取得了 53%的召回率。这些方法为垃圾过滤研究开拓了思路，但现有成果距离全面解决垃圾问题还有一定距离。

垃圾邮件过滤技术可大致分为内容过滤，接入过滤和行为过滤三类，下面我们就这三类方案逐一展开论述，探讨其研究进展及关键技术。

3 内容过滤技术

基于内容的过滤技术与文本分类和信息过滤密切相关，它相当于在动态的邮件流中，根据邮件内容将邮件分成合法与垃圾两类，多种分类方法和机器学习理论已应用于垃圾邮件的过滤。内容过滤大致分为基于规则的方法和基于统计的方法两类。规则方法从训练集中得出人们可以理解的显式规则；统计方法往往通过某种计算表达式推出结果。常见的规则方法为关键词过滤，决策树[11]，boosting[12]，ripper[13]等，常见的统计方法为 SVM[14]，NB[15]方法等。详细介绍参见文献[3]。近年来，一些研究者将遗传学、免疫学理论引入内容过滤领域，为垃圾邮件过滤技术注入了新的动力。

3.1 免疫学方法

Terri和Andrew分别提出了基于人工免疫系统的反垃圾邮件方法[16、17]。其主要思想为，生物免疫系统是一个动态自稳的自适应学习系统，而垃圾邮件过滤与生物体的免疫功能极其相似。借鉴其记忆和联想能力可以完成识别和分类任务；借鉴其连续学习能力可以解决用户的正常邮件兴趣漂移问题；借鉴其信息遗忘机制可以解决知识冗余和冲突问题，以维护知识库的动态自稳。

3.2 遗传学方法

IBM研究中心的生物信息组提出了一种基于模式发现的反垃圾邮件方法—钟馗算法[18]。该算法原本为

DNA序列分析而设计，它在训练集中使用Teiresias算法[18]提取垃圾邮件的特征模式，并根据模式匹配的数量以及邮件内容的序列相似性来判定垃圾邮件，试验结果表明，该方法具有较高的准确率。目前，上述研究仍处于起步阶段，研究成果距实用化还有较大距离。

4 接入过滤技术

接入过滤技术在邮件服务器处对邮件信头部分进行检查，因此可以提前发现非法信息，且不至于侵犯个人隐私。接入过滤技术可以分为打补丁的方法和修改协议的方法两类。前者针对现有邮件协议缺点提出补救措施，以增强邮件系统的安全性；后者多为企业界提出的方案，其实现需要改变电子邮件系统的工作方式。

4.1 打补丁的方法

常见的打补丁的方法有黑白名单、反向域名查询、延迟技术、灰名单技术、邮资策略等

黑-白名单是一种简单、有效的身份核实方法[7]。地址在黑名单列表中的邮件直接被判定为垃圾邮件，地址在白名单列表中的邮件直接被判定为合法邮件。DNS反向查询的原理是接收方查询发送者IP地址所对应的真实域名，如果查询结果和邮件宣称的域名不符，则认为该邮件为垃圾信息。这两类技术的缺点是无法区分发送者和中继者，同时提供查询服务的站点易受到DDOS攻击。

在 SMTP 连接中注入延迟使垃圾发送者陷入两难境地，保证成功发送就必须减少垃圾邮件的发送量；追求高吞吐率就必须放弃当前连接。常见的延迟技术有 Tarpit[19]，Teergrubing[20]Throttle 模型[21]，Greylist 技术[22]等。延迟技术并不主动过滤垃圾邮件，它是用增加发信成本的方法抑制垃圾邮件的发送或迫使垃圾产生者因为收益降低而主动放弃发送。

邮资策略的原理是发送者要为每封信件付出一定的代价，这个代价与发信规模成正比。常见的邮资策略如 Yahoo 和 Hotmail 采用了反向图灵机测试技术[23]，慢发送方案[24]，N 次验证[25]等。

4.2 修改协议的方法

AOL 提出了 SPF（Sender Permitted From）方案[26]。它相当于网域邮件服务器上的反向邮件交换记录解析。一般的邮件交换记录所存储的是该网域中有哪些合法的收件 MTA（邮件传输代理），而 SPF 则可以让网域管理者指定哪些地址才是该网域合法的发送者。当收到时，邮件服务器通过 DNS 查询即可得知该信件是否由信件中所宣称的网域寄出。类似的技术还有 Yahoo 提出了 Domain Keys 方案[27]，Microsoft 的 Sender ID 方案[28]等。目前 Sender ID 方案已经被 SENDMAIL 采用，初步测试结果为：服务器对数据包输入、输出速度分别减慢了 15%和 8%。由于需要新的邮件协议支持，这类技术很难在短时间内推广使用。

5 行为过滤技术

文献[29]指出，合法邮件是在社会关系驱动下，以交换信息为目的，双向通信的结果；而垃圾邮件是在发送者利益驱动下，以大范围扩散为目的，单向通信的产物：两者本质上的不同必然导致其行为的显著差异，因此垃圾邮件和合法邮件从行为特征上看是可以区分的。目前国外在该领域的研究主要集中于流量行为、相似性行为这两方面。

5.1 流量行为

从 2003 年起，学术界开始从流量特征入手，对邮件协议行为进行深入研究。流量行为研究可分为基于统计的方法和基于拓扑方法两种。

5.1.1 基于统计的方法

统计方法主要在小范围内对邮件流量进行监测并统计其变化规律，进而为过滤垃圾邮件提供依据。根据监测点所处的位置，可以将研究分为两类。

第一类工作在邮件发送阶段对流量进行统计。文献[30]对 SoBig、MyDoom 两种邮件蠕虫进行深入研究，发现上述病毒爆发时，局域网内 DNS 流量和失败的 SMTP 连接数目急剧增加，因此上述异常事件都是垃圾邮件出现的征兆。文献[22]发现，垃圾发送者为了提高其吞吐率，往往采取 fire and go 方式进行发送，即对发送失败的邮件不做重传尝试，并以此特征实现了灰名单技术。

第二类工作在邮件接收阶段对流量进行统计。[29]对比合法、垃圾邮件流量在到达率，邮件长度分布，发送者、接收者的分布，统计它们在流量行为上的差异，分别为垃圾和合法邮件建立流量模型，其研究成果可以为模拟试验和有效区分合法与垃圾邮件提供依据。

5.1.2 基于拓扑的方法

研究者将邮件视为流数据，以收发信人为顶点，双方通信关系为边构建邮件网络来表示邮件流量，并在此基础上进行统计和分析。根据建立邮件网络所采用的数据源，可以将研究分成两类。

第一类研究使用邮件服务器上的日志信息建立网络。文献[31]偏重于建立邮件网络模型，并通过试验证明邮件世界同样具有 scale-free 和 small world 属性。文献[32]通过分析发送者和接受者产生的边界流图，使用 HIS 算法来分析流量图的演化结构。文献[33]使用二分图来发现恶意发送者。

第二类研究使用用户邮箱中信件构建网络。文献[34]通过对节点出入度，网络聚合性，互通性进行建模分析，研究邮件病毒的扩散趋势。文献[10]通过计算邮件网络中各个子图是否具有社会属性来区分合法邮件与垃圾邮件，试验结果表明该方法可以对 53%的邮件做正确区分。文献[35]是对文献[9]工作的一个扩展，结合社会网络和白名单识别、推断网络中信誉度高的用户。

上述文献表明，在接收服务器和客户端通过流量特征研究是可以部分的解决垃圾邮件问题的，在骨干网等大流量环境下，上述方法是否有效还需进一步的研究。

5.2 相似性行为

垃圾发送者为保证其经济利益，唯一的方法就是短时间内发送大量相似的内容[3]，根据这一本质特征实现了许多垃圾邮件过滤技术。

哥伦比亚大学发起的 MET（Malicious email tracking）项目[7]是一个 C/S 框架：客户端工作在邮件服务器上，提取每封邮件的附件部分，并用 MD5 算法为该邮件计算一个唯一的校验和，并将校验和连同邮件描述信息保存在数据库中。如果该校验和在短期内出现频率超过某一域值，则将相应记录发送给服务器。服务器负责管理和向其他客户端下发高频记录。该框架的问题在于缺乏全局统计功能且不能识别相似邮件。

DCC（Distributed Checksum Clearinghouse）工程[36]始于 2000 年，目前已经发展成拥有数万客户，250 个服务器的大型 C/S 系统，每周处理邮件超过 1.5 亿封。DCC 服务器统计各客户端提交的校验和报告，并回答该校验和的在整个系统中出现的频率。客户端工作在邮件服务器上，向服务器提交每封邮件的校验和，同时查询此校验和在整个系统中出现的频率。如果该频率超过客户端的阈值，当前邮件可判定为垃圾信息。DCC 采用检验和模糊匹配算法，可有效地识别内容发生细微变化的相似邮件。

MET 与 DCC 均运行在接收端邮件服务器上，由于识别相似行为必须在大流量环境下才有效果，因此它们采用了 C/S 框架的多机协同处理模型，通过相互交换校验和达到信息共享，其共同的缺点是通信开销较大；客户可以产生假数据干扰系统运行。

6 结论与展望

前面介绍了三种过滤技术下多种治理垃圾邮件的方法：其中内容过滤技术可以独自归为一类，接入过滤技术主要可分为修改协议和打补丁两种方法，行为过滤技术可分为基于流量行为和基于相似行为两种方法。本节首先通过3个性能指标(见表1）比较上述方法的优劣，然后提出目前研究中的难点问题，并指出未来发展方向。

6.1 现有过滤技术比较

表一 各种过滤方法性能比较

过滤方法	所属类别	技术生命周期	技术适用范围	召回率
内容过滤	内容过滤	短	接收阶段	高
修改协议	接入过滤	长	发送、接收阶段	高
打补丁	接入过滤	长	发送、接收、传输阶段	中
流量行为	行为过滤	长	发送、接收、传输阶段	低
相似行为	行为过滤	长	接收、传输阶段	高

技术生命周期决定过滤技术的维护开销。由于垃圾发送技术不断演化、发展，内容过滤技术需要用户不断更新规则库或训练集，这使得维护工作变得异常困难。而其他技术均属于无监督的过滤方法，针对垃圾邮件稳定、不易掩饰的特征采取相应的治理措施，故有较长生命周期。

目前定位在发送或接收阶段的方案种类繁多，但却很少有支持传输阶段过滤的技术。由于复杂的内容测试会造成较大的延迟，各种内容过滤技术在主要应用于接收阶段。同理，多数接入过滤技术都有查询、验证的过程，因此多数方法只能应用于对实时性能要求较低的发送和接收阶段。行为过滤技术可以应用在任何阶段的。

召回率定义为垃圾邮件的检出率，是衡量过滤技术识别垃圾邮件能力的重要指标。由表1可见，多数方案性能令人满意，尤其是内容过滤技术，甚至达到了 99%以上的召回率。打补丁的方式本身召回率较低，如黑名单只能检出 10%的垃圾邮件，灰名单仅对不遵守 RFC 协议的邮件起作用，因此通常需要与其他过滤技术结合使用。基于流量行为的过滤技术召回率仅为 50%，研究成果离实用有较大差距。

综上所述，目前反垃圾技术方案和商业产品数量虽多，但还没有一种可以全面的解决垃圾邮件问题，垃圾邮件过滤至今仍然是一个开放性的问题。

6.2 未来的发展方向

由图 1、表 1 可以看出，目前垃圾邮件过滤面临的困难主要来自于两个方面：一是现有邮件体系自身的问题，二是过分依赖内容过滤技术，缺乏对新理论、新技术的深入研究。

现有的邮件体系是建立在易用、开放和信任的基础上的，本善的初衷反而成为信息安全的隐患。发信成本低廉，不核实发送者身份，对大量发送没有惩罚机制，这是垃圾邮件问题愈演愈烈的根本原因。客观地说，只要现有邮件体系不改变，垃圾邮件问题就不可能得到彻底地解决。尽管 AOL、微软、Yahoo 陆续提出了多种修改邮件体系结构的方案，但是从实际运行效果和推广的进程来看，各种方案均存在性能问题且很难就标准问题达成一致。

随着研究的不断深入，人们发现尽管在主流产品中得到广泛使用，准确率和召回率另人满意，但内容过滤技术并非解决垃圾问题的最佳方案。这个论断的产生来自于如下两点：（1）虽然终端用户接收的垃圾信息减少了，但网络有效利用率没有得到改善。（2）内容过滤技术生命周期短（表 1）.由此引发的问题是如何寻

求更为有效的解决方案？对于这个问题，Gomes 等人提出，垃圾邮件与合法邮件本质上的不同必然导致其行为的显著差异，因此建议从邮件的行为模式中寻求答案[29]。近年来，免疫学和遗传算法等新计算理论的引入也为该领域注入了新动力[16、17、18]。此外，Greylist，Throttle 等项目的经验表明，多种技术有机融合和协作式过滤是垃圾过滤领域未来的发展趋势。

综上所述，今后垃圾邮件过滤问题的研究路线主要包括如下几个方面：（1）设计更为安全和完善的邮件体系结构，这是个渐进、协商的过程，需要技术，立法等多层次的合作。（2）积极开拓新的思路，将过滤技术由基于静态规则和统计分类的方法向基于行为模式分析的方法转变。（3）借鉴数据挖掘、图论、对等网络、免疫学等研究成果，为解决垃圾邮件问题提供更有效的依据。

参考文献：

[1] Brightmail Inc[EB/OL]. http://www.brightmail.com.

[2] Anselm L. Analysis of spam. University of Dublin, Trinity College Master thesis[D]. Sep 2003

[3] 王斌，潘文峰。 基于内容的垃圾邮件过滤技术综述[EB/OL]. http://www.nlp.org.cn/docs/docredirect.php?doc_id=1024

[4] Spam Filter Analysis[EB/OL]. www.cs.ru.nl/~flaviog/publications/spam-filter.pdf

[5] Prahanth S. A. Overview of Spam Handling Techniques.

[6] On the junk mail problem[EB/OL]. http://cs.gmu.edu/~huangyih/756/spam-report.pdf

[7] Read-time Blackhole List.[EB/OL]. http://mail-abuse.org/rbl/

[8] Manasi B, Shlomo H, Eleazar E. MET: An Experimental System for Malicious Email Tracking[A]. In Proceedings of the 2002 New Security Paradigms Workshop (NSPW-2002)[C]. Virginia Beach, VA: September 23rd - 26th, 2002.

[9] Kenichi Y, Fuminori A. Density-based spam detecton[A]. KDD2004[C], August 22-25,Seattle,Washington, USA. Page 486~493.

[10] Oscar. B, Vwani R. Personal Email Networks: an effective anti-spam tool[EB/OL]. http://www.arxiv.org/abs/cond-mat/0421 43.

[11] Carreras X and Marquez L, Boosting Trees for Anti-Spam Email Filtering[A], in Proceedings of Euro Conference Recent Advances in NLP[C], pp. 58-64, Sep. 2001.

[12] Using AdaBoost and Decision Stumps to Identify Spam E-mail[EB/OL]. http: //nlp.stanford.edu/ courses/cs224n/ 2003/fp/

[13] Cohen W. Fast effective rule induction[A], in Machine Learning. Proceedings of the Twelfth International Conference[C], Lake Taho, California, Mongan Kanfmann, pp. 115-123, 1995

[14] Drucker H, Wu D, and Vapnik V, Support Vector Machines for Spam Categorization[J], IEEE Transactions on Neural Networks, Vol. 20, No. 5, pp. 1048-1054, Sep. 1999

[15] Sahami M, Dumais S, Heckerman D. "A Bayesian approach to filtering junk e-mail", in Proc. of AAAI Workshop on Learning for Text Categorization, pp. 55-62, 1998[C]

[16] Terri O and Tony W. Developing an Immunity to Spam[A]. In Proceedings of the Genetic and Evolutionary Computation Conference(GECCO 2003)[C], Chicago, July 2003.

[17] Andrew S, Alex A. Freitas Jon Timmis. AISEC: an Artificial Immune System for E-mail Classification[A]. Proceedings of the Congress on Evolutionary Computation (CEC-2003)[C], pp. 131-139, Canberra. Australia, December 2003. IEEE Press..2003

[18] Isidore R and Tien H. Chung-Kwei: a Pattern-discovery-based System for the Automatic Identification of Unsolicited E-mail Message[A]. CEAS 2004[C]. Mountain View, CA. July 30 and 31, 2004

[19] Tarpit[EB/OL]. http://www.palomine.net/qmail/

[20] Lutz D. Teergrubing[EB/OL]. http://www.iks-jena.de/mitarb/lutz/usenet/teergrube.en.html.

[21] Matthew M. Design, implementation and test of an email virus throttle. In Proceedings of ACSAC Security Conference, Las Vegas, Nevada, December 2003. Available from http://www.hpl.hp.com/techreports/2003/HPL-2003-118.html.

[22] Evan H. The next step in the spam control war: Greylisting[EB/OL]. http://projects.puremagic.com/greylisting/

[23] Reverse Turing test[EB/OL]. http://en.wikipedia.org/wiki /Reverse_Turing_test

[24] Hird S. Technical Solution for Controlling Spam[A]. In proceedings of AUUG2002[C], Melbourne September 2002.

[25] Joshua G, Robert R. Stopping Outgoing Spam[A]. EC'04[C], May 17-20, 2004, New York, USA.

[26] Lentczner M, Wong M. Sender Policy Framework[EB/OL], http://www.ietf.org/internetdrafts/draft-mengwong-spf-01.txt.

[27] Delany M. Domain-based Email Authentication Using Public-Keys Advertised in the DNS[EB/OL]. http://www.ietf.org/internet-drafts/ draft-delany-domainkeys-base-00.txt.

[28] Atkinson B. Caller ID for E-Mail[EB/OL]. http://www.ietf.org/internet-drafts/draft-atkinson-callerid-00.txt,

[29] Gomes L, Cazita C, Almeida J. Characterizing a Spam Traffic[A]. IMC'04[C]. Oct, 25-27,2004. Taormina, Sicily, Italy.

[30] Wong C, Bielski S, McCune J. A Study of Mass-mailing Worms[A]. Proceedings of the 2004 ACM workshop on Rapid malcode[C], Oct 29,2004, Washingto, DC, USA.

[31]Holger E, Lutz-Ingo M, and Stefan B. Scale-free topology of e-mail networks[J]. Phys. Rev. E 66, 035103(2002)

[32] Gomes L, Almeida R, Bettencourt L. Comparative graph Theoretical Characterization of Networks of Spam and Regular Email[EB/OL]. http://arxiv.org/abs/cond-mat/0503725.

[33] Prasanna E , Jaideep S. Analyzing Network Traffic to Detect E-Mail Spamming Machines[R]. Technical Report 180, Army High Performance Computing Research Center TECHNICAL REPORT, 2004

[34] Newman M, Stephanie F, and Justin B. Email network and the spread of computer viruses[J]. Phys. Rev. E 66, 035101(2002)

[35] Jennifer G, James H. Reputation Network Analysis for Email Filtering[A]. CEAS 2004[C]. Mountain View, CA. July 30, 2004

[36] DCC[EB/OL]. http://www.rhyolite.com/anti-spam/dcc/.