电子邮件作为目前在Internet上使用最广泛的服务，近年来日益受到垃圾邮件的威胁。众多的垃圾邮件最终将流向我们每一个电子邮件用户，极大地影响了电子邮件的使用效率。垃圾邮件之所以泛滥，原因在电子邮件系统本身在安全方面存在一些漏洞。当前已经有许多比较成熟的反垃圾邮件技术，其中的一些已经投入实用。本文将解释垃圾邮件的成因，并介绍一些主要的反垃圾邮件的技术，最后介绍一种新的类似于垃圾邮件的有害信息传播方式。

反垃圾邮件技术

解放军理工大学通信工程学院　陶卓彬　邓元庆

垃圾邮件的历史

垃圾邮件(Junk Mail，Spam)是指未经收件者同意即大量散发的邮件，信件内容多半以促销商品为意图。垃圾邮件也称作"不请自来的商业电子邮件"(Unsolicited Commercial Email，UCE)或"不请自来的大量电子邮件"(Unsolicited Bulk Email，UBE)。

1975年，Jon Postel提出了垃圾邮件(Junk Mail)的概念，他在名为《关于垃圾邮件问题》(On the Junk Mail Problem)的文章中指出"选择性拒绝消息机制(selectively-refuse-message mechanism)"的缺乏将成为网络上信息传播的一种安全隐患。首次关于垃圾邮件的记录是1985年8月一封通过电子邮件发送的链锁信。历史上比较著名的事件是1994年4月12日，美国亚利桑那州两位从事移民签证咨询服务的律师劳伦斯·坎特(Laurence Canter)和玛撒·西格尔(Martha Siegel)把一封宣传"绿卡抽奖"活动的广告信发到6000多个新闻组，这是因特网上第一次有人大规模地滥发广告邮件。垃圾邮件开始引起了人们的注意和反感的同时，一些触觉敏锐的商人意识到了电子邮件带来的商机，许多人开始利用电子邮件作商业广告，与发送垃圾邮件相关的一些产业也开始出现。1995年5月有人写出了第一个专门的大批量发送电子邮件的程序Floodgate。紧接着在8月份，有人拿两百万个邮件地址出售。垃圾邮件越来越多与商业联系起来。

电子邮件系统的安全缺陷

开放中转(Open Relay)问题

80年代之前，由于网络不健全，机器之间很难以直接连接收发邮件，人们必须先找出一条由多个站点组成的有效通路，该通路上的每一个站点都是邮件服务器。在邮件发出之前，由发送者拟定要经过的站点的列表，邮件将沿着这些站点通过转发传送到目的地。基于这样的背景，SMTP中提供了用于邮件在不同的网络间传送的"邮件中转"(Mail Relay)服务。在该协议中，规定了两种站点列表，一种称为前向路径(forward-path)，这是邮件将要经过的站点的名称，另一种是反向路径(reverse-path)，是邮件已经路过的站点的名称，当邮件到达一个站点时，服务器根据前向路径得知邮件下一站要去的地址，当该邮件已由服务器转发出之后，该站点的地址将成为反向路径列表的第一项，这样邮件的去路与来路就清晰地被记录下来了。

邮件中转的原理如图1所示：

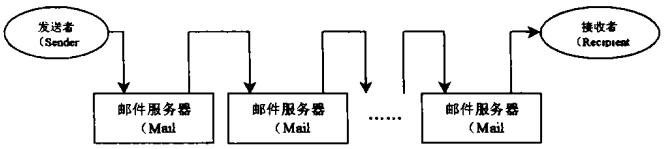

图1 邮件中转的原理图

上图中，邮件在发送者和接收者之间经过了若干个指定的第三方服务器。

SMTP虽然提供了邮件中转服务，但邮件服务器可以根据前向路径列表和反向路径列表对请求中转的邮件进行限制，如只中转来自或去向某些域或IP的邮件。如果对中转没有任何限制，则被称为"开放中转"(Open Relay)或"第三方中转"(Third Party Relay)。

目前，对普通邮件而言大多已经不再需要中转，但SMTP提供的邮件中转服务并由此引出的开放中转方式，却常常被垃圾邮件发送者利用，用来隐藏其邮件真实的来源，以使针对源地址的垃圾邮件过滤技术失效。

这种利用开放中转发出的垃圾邮件，由于无法确定其源地址，某些反垃圾邮件软件和组织会把报复的目标转移到中转该垃圾邮件的邮件服务器(即那些开放中转的服务器)上。2002年3月1日，国内许多媒体报道中国电子邮件将面临欧美的全面封杀，其原因就是

中国的邮件服务器被利用来中转垃圾邮件，因而遭到了欧美国家邮件服务器的报复。

缺乏"选择性拒绝"机制

1975年，Jon Postel在他的名为《关于垃圾邮件问题》(On the Junk Mail Problem)的文章中指出："在ARPA网络Host/IMP接口协议中，没有主机选择性拒绝消息的机制，这就意味着想要接收某些特定消息的主机不得不阅读发给它的所有消息。这样的主机可能收到出故障的主机发来的大量消息，会使该主机的正常用户遭受'拒绝服务(DoS)'。本地用户和网络通信都会受到损失。"他在这篇文章中使用的"邮件(Mail)"一词并不是后来的"电子邮件(Email)"，而是"消息(Message)"的代称，但他所提出的"选择性拒绝消息机制"却是当前的电子邮件系统所存在的问题。

SMTP并没有将"选择性拒绝消息机制"纳入其考虑的范围，它只是规定了怎样接收邮件，而没有规定该接收哪些邮件，这样垃圾邮件只要在传送过程中不出意外便能到达目的邮箱。

反垃圾邮件技术

改造邮件服务器传送功能

为了切断垃圾邮件的传播途径，首先要根除"开放中转"(Open Relay)问题。由于历史原因，现在大部分邮件服务器软件均将开放中转作为默认配置，这就需要在服务器上进行相应的配置修改。修改方式有两种，一种是完全关闭邮件中转功能，另一种是针对前向路径或反向路径进行限制，只转发特定来路或去路的电子邮件。在配置修改后可以Telnet该邮件服务器的25端口(SMTP默认的端口)来验证修改是否有效，具体过程如下：

C:Telnet smtp.xxx.com 25

S:250 xxx.com

C:HELO remote.system.domainname

S:250 xxx.com

C:MAIL FROM:user@somewherer.net

S:250 OK

C:RCPT TO: user1@elsewhere.net

邮件接收者user1@elsewhere.net中的域名不是本地域名，这时候本地系统可能有两种回答，接受它：

250 OK

或者拒绝接受它：

553 sorry,.that domain is not in my domain list of allowed recphosts

第一种情况说明该服务器是允许中转的，它接收并同意中转一个目的地址不是本地的邮件；而第二种情况则说明不中转非本地邮件。

为邮箱增加"选择性拒绝"机制

定义：

误承认(False positives)，将正常邮件错误地识别为垃圾邮件。

误否认(False negatives)，将垃圾邮件错误地识别为正常邮件。

以上两种错误的发生率是对反垃圾邮件技术有效性进行评估的重要依据。

• 黑名单(Blacklist)

这是最早出现的一种反垃圾邮件技术，一般的邮件服务器都具有该功能。作法是由用户或系统管理员将已知的垃圾邮件发送地址填入一张表，称为黑名单。服务器将拒绝所有来自黑名单地址的邮件。但这种方法太过被动，只能防止已知的固定地址的垃圾邮件发送者，对于新地址或任意伪造源地址的垃圾邮件，这种方法无效。因此在实际中，这种邮箱的误否认率很高。后来又发展出针对"域"的黑名单，如Hotmail就提供这个功能，但其原有缺陷仍然存在，而且会带来误承认率的大大增加。

• 白名单(Whitelist)

白名单方法与黑名单方法恰恰相反，它需要用户定制出一张允许接收的地址列表，来自这些地址之外的邮件将一律被拒绝。这种方法的优点是可以绝对排除垃圾邮件的进入，因为用户不会将垃圾邮件发送者的地址列入白名单，但这样的邮箱使用起来却很麻烦，因为如果用户想要接收来自新地址的邮件就需要更改列表。但白名单邮箱因为其对垃圾邮件的"绝对免疫"能力，将有可能成为每个用户必备的邮箱之一。

• "三次握手"认证

这是在白名单方法的基础上发展出一种"标记消息交付代理"(Tagged Message Delivery Agent，TMDA)的技术，并于2001年4月实现了一个实用性的产品。TMDA原理类似TCP的三次握手过程。使用这种技术的邮箱以白名单为基础，将只允许合法地址的邮件进入，如果有其它地址的邮件到达(第一次握手)，它会将该邮件暂时保存起来，但并不让用户看到该邮件；接着它会根据邮件的源地址发送一封要求认证的邮件，通知对方必须在一定的期限之前回复该邮件(第二次握手)；如果对方的源地址是伪造的或者没有在指定期限之前回复(假定这种技术被大量使用，垃圾邮件发送者不可能回复等同自己发出的垃圾邮件数量的要求认证邮件)，邮箱会将暂存的邮件当作垃圾邮件丢弃，如果对方回复(第三次握手)，则说明该邮件地址不是垃圾邮件，将进入白名单，上次暂存的邮件也将交付用户查看，下一次来自该地址

的邮件将不会受到阻碍。这种方法力图降低邮箱使用者配置白名单的工作量，而将这种工作量分散转移到邮件发送者身上，应该说是可行的。但是在一些特殊情况下可能会遇到麻烦，比如说，哈佛给这种邮箱发来了一封录取通知书，但发送方不一定有耐心去回复证明自己合法的认证邮件，这样就引起邮箱对该信件误承认，从而造成邮箱使用者的损失。

• 基于内容的过滤技术

Paul Graham 于 2002 年 8 月写了一篇极具争议的文章《针对垃圾邮件的计划》(A Plan for Spam)，在该文中 Graham 提议建立区分垃圾邮件和非垃圾邮件单词的贝叶斯概率模型。他的这篇文章代表了基于内容的垃圾邮件过滤技术的思想。这种思想认为尽管垃圾邮件可以采用伪造地址等欺骗方式，但其邮件的内容无法进行伪装，发送者为了传达一定的信息，必定会在内容上带上一些垃圾邮件才有的特征，比如大量出现的某些特定词汇、为了醒目而采用的特殊文字格式、大量使用图片等等。将这些特征用算法进行归纳整理，总结其在垃圾邮件中出现的概率，以这些特征出现在邮件中的频率并设定某一门限值来判断该邮件是否为垃圾邮件。从实践来看基于内容的过滤技术能达到对垃圾邮件较高的识别率，但也存在一定的误承认率与误否认率，对于那些宁肯收到垃圾邮件也不愿丢失邮件的用户而言，这种技术显然还需要加以改进。

主动的反垃圾邮件技术

• 垃圾邮件陷阱

不管是关闭邮件服务器的中转功能，还是为邮箱加装"选择性拒绝消息"模块，这都是在垃圾邮件发动攻击之后采取的被动性措施，作者认为与垃圾邮件的斗争应该可以更主动一些。比如，可以在网络中设置类似陷阱的邮件服务器，它声称可以为任何邮件提供中转服务，引诱垃圾邮件传送到该服务器，在该服务器上有垃圾邮件识别技术，如果判断该邮件是一封垃圾邮件，则将其丢失，并告知发送者该邮件已经到达目的地。通过这种欺骗技术，垃圾邮件发送者不会知道自己发出的邮件已经被消灭，必将继续地试图利用该服务器发送垃圾邮件，这样就达到了"诱捕垃圾邮件"的目的。当然这种方法应该有相应的规范来加以限制，以避免被滥用而影响正常邮件的传送。

垃圾邮件的个人预防措施

除了使各种措施来防御垃圾邮件，用户在使用电子邮件时也要遵守一些原则来预防垃圾邮件的侵扰。因为垃圾邮件的攻击对象是电子邮箱，我们就有必要避免让自己的电子邮件地址落到垃圾邮件发送者手中。为此需要遵守的原则是：

• 不随意公开真实 E-mail 地址

不要在 BBS、论坛、新闻组等网上公开场合公布自己的真实 E-mail 地址。如果你必须留下 E-mail 和别人交流，最好留下你的免费 E-mail 地址。

• 使用特殊的方法书写邮址

上面提到，网上有 MailFinder 等电子邮件地址自动收集软件，而这些软件主要是对电子邮件地址特有的字符"@"敏感，并将该字符前后的其它一些字符识别为一个邮件地址。针对这一点，在不得不留下邮件地址的情况下，要对 @ 进行改头换面，可以将它写成全角形式的"＠"或者用汉字进行说明。

• 不要回复已经收到的垃圾邮件

如果已经有垃圾邮件进入自己的邮箱，不要回复该邮件，也要按上面"取消订阅"之类的链接，这样做只会让垃圾邮件发送者知道这个邮件地址是有效的，更多的垃圾邮件将发到该地址的邮箱。

垃圾邮件的新发展

从目前来看，垃圾邮件的内容大多是商业广告，邮件只是传播商业广告的一种载体，如果反垃圾邮件的发展使这种广告传播方式的效率大大下降，可能会使广告商对电子邮件失去兴趣。现在，商业广告已经开始出现脱离邮件这种载体的趋势，作者在网上发现了这样一个网站，

这种传播商业信息的方式比垃圾邮件更方便，因为它并不要求用户打开邮箱查看，而是主动送到你的桌面，可以判断，如果对这种发送商业信息的方式不加以有效地阻止，将会带给人们比垃圾邮件更大的危害。

垃圾邮件作为一种网络上的有用信息，严重地影响了网络的使用效率。许多行之有效的反垃圾邮件技术已经投入使用并取得了较好的效果。但是垃圾邮件存在的基础，即邮件传输系统在安全上的漏洞依然存在，这需要对邮件系统的安全性做一些标准化的工作，这是今后值得努力的方向。另外，网络有害信息正在寻求新的传播途径，这也是值得我们关注的问题。