第6章 CHAPTER 传统反垃圾邮件技术（下）

在第5章已经介绍了一部分传统的反垃圾邮件技术。本章将继续对反垃圾邮件技术进行介绍。

6.1 数量控制：带宽/连接限制

垃圾邮件的一个基本特征是会在短时间内发送大量的邮件，具体到网络层上来看，这就表现为某个IP地址会在短时间内重复连接收件服务器的SMTP端口，并占用大量的带宽。根据这个特点，可以通过对网络层的限制，达到限制垃圾邮件的目的。具体而言，可以从两个途径下手进行限制，一个是对每个IP地址可用的带宽比例进行限制，这个一般可以通过网络设备的QoS流控管理功能来实现；另一个是对每个IP地址的并发SMTP连接数目进行限制，这个一般可以通过网络层防火墙策略来实现。从这里也可以看出，将传统防火墙技术和反垃圾邮件技术相结合，可以更加有效地防范垃圾邮件。

在对网络层行为进行限制的基础上，还可以进一步结合黑名单技术来提高防范垃圾邮件的效果。例如，在限制并发SMTP连接时，并不只是简单的限制，同时也记录单位时间内的连接请求数量。如果某IP地址的单位累计连接数量大于某个阈值，如1000次/s，则把该IP地址加入黑名单中，拒绝该IP地址的全部连接请求，或者减少对该IP地址允许的并发连接数量，直到其单位累计连接数量小于某个阈值，如1次/15s，再将其从黑名单中删除。通过这样结合动态黑名单，可以更好地限制那些异常的网络流量，从而提高防范垃圾邮件的效果。

带宽/连接限制在对付那些来源单一的垃圾邮件时，有很好的效果。但是，如果垃

圾邮件通过采取将邮件分散到多个 IP 地址发送的手段，减少单一 IP 地址发出邮件数量，那么这种限制技术就很难起到作用。因此，该技术也只能是一种辅助的反垃圾邮件手段。

6.2 数量控制：邮件重复限制

除了在网络层对垃圾邮件流量进行限制外，还可以在 SMTP 应用层对垃圾邮件进行限制。通过商业垃圾邮件制造机制造出来的商业垃圾邮件，同一封邮件的邮件头中的信息往往是不变的，因此可以利用这一点，对那些邮件头中关键信息重复的邮件进行限制，以达到限制垃圾邮件的目的。通常，系统会根据邮件头中的发件 IP 地址、发件人、邮件主题 3 个字段在一段时间内的重复率进行限制。这种限制虽然技术简单，但对从一个 IP 地址发出的大量邮件，特别是通过商业垃圾邮件制造机产生的垃圾邮件非常有效。

在邮件重复限制的基础上，结合动态黑名单技术，也可以达到更好的限制效果。具体说来，就是增加对单位时间内邮件重复数量的计数，如果某 IP 地址的重复邮件数量超过了预先设置的阈值，如 100 封/min，则将该 IP 地址暂时列入黑名单中，拒收该 IP 地址发来的邮件，经过一段预设的时间，如 1 小时之后，再将该 IP 地址移出黑名单。这样既可以达到限制垃圾邮件的目的，也可以减少服务器由于处理垃圾邮件所消耗的资源。

与带宽/连接限制技术一样，邮件重复限制技术能起的作用也是有限的。只要垃圾邮件制造者对邮件进行简单的伪造，经常变换发件人地址和邮件主题等关键字段，就可以很轻易地逃过这种技术的限制。因此，该技术也只能是一种辅助的反垃圾邮件手段。

6.3 新型技术：贝叶斯分析

贝叶斯分析得名于著名数学家托马斯·贝叶斯(1702—1761)，他发展了一个数学领域全新的可能性推论理论，即贝叶斯理论。贝叶斯分析的核心是贝叶斯定理，这是统计学中的一个重要定理，根据该定理，可以根据已知事件发生的概率(该概率可以通过对已发生的事件进行统计获得)来计算下一次该事件发生的可能性。

垃圾邮件的贝叶斯分析技术，就是利用贝叶斯定理，以已知垃圾邮件和非垃圾邮件为样本，通过对样本邮件内容的分析和统计，来计算下一封邮件可能是垃圾邮件的概率。

贝叶斯分析需要以已知的邮件作为样本进行自学习，从理论上说，所用的样本数量越多，贝叶斯分析的准确程度也就越高。

6.3.1 贝叶斯过滤算法的基本步骤

该步骤引自 anti-spam.org.cn[①]。

(1) 收集大量的垃圾邮件和非垃圾邮件,建立垃圾邮件集和非垃圾邮件集。

(2) 提取邮件主题和邮件体中的独立符字符串,如 ABC32,¥234 等作为 TOKEN 串并统计提取出的 TOKEN 串出现的次数,即字频。按照上述的方法分别处理垃圾邮件集和非垃圾邮件集中的所有邮件。

(3) 每一个邮件集对应一个哈希表,hashtable_good 对应非垃圾邮件集而 hashtable_bad 对应垃圾邮件集。表中存储 TOKEN 串到字频的映射关系。

(4) 计算每个哈希表中 TOKEN 串出现的概率:P=某 TOKEN 串的字频/对应哈希表的长度。

(5) 综合考虑 hashtable_good 和 hashtable_bad,推断出当新来的邮件中出现某个 TOKEN 串时,该新邮件为垃圾邮件的概率。数学表达式为:

A 事件——邮件为垃圾邮件,

t1,t2,…,tn 代表 TOKEN 串,

则 P(A|ti)表示在邮件中出现 TOKEN 串 ti 时,该邮件为垃圾邮件的概率。

设

P1(ti)=(ti 在 hashtable_good 中的值)

P2(ti)=(ti 在 hashtable_bad 中的值)

则 $P(A|ti)=P1(ti)/[(P1(ti)+P2(ti)]$

(6) 建立新的哈希表 hashtable_probability 存储 TOKEN 串 ti 到 P(A|ti)的映射。

(7) 至此,垃圾邮件集和非垃圾邮件集的学习过程结束。根据建立的哈希表 hashtable_probability 可以估计一封新到的邮件为垃圾邮件的可能性。

当新到一封邮件时,按照步骤(2)生成 TOKEN 串。查询 hashtable_probability 得到该 TOKEN 串的键值。

假设由该邮件共得到 N 个 TOKEN 串,分别为 t1,t2,…,tn,hashtable_probability 中对应的值分别为 P1,P2,…,PN,

P(A|t1,t2,t3…tn)表示在邮件中同时出现多个 TOKEN 串 t1,t2…tn 时,该邮件为垃圾邮件的概率。

由复合概率公式可得

$P(A|t1,t2,t3\cdots tn)=(P1*P2*\cdots*PN)/[P1*P2*\cdots*PN+(1-P1)(1-P2)\cdots(1-PN)]$

① http://anti-spam.org.cn/forums/indwx.php? showt opic=448

当 P(A|t1,t2,t3…tn)超过预定阈值时,就可以判断邮件为垃圾邮件。

6.3.2 贝叶斯过滤算法举例

例如:有一封含有“法轮功”字样的垃圾邮件 A 和 一封含有“法律”字样的非垃圾邮件 B,根据邮件 A 生成 hashtable_ bad,该哈希表中的记录为

法:1 次

轮:1 次

功:1 次

计算得在本表中:

法出现的概率为 0.3

轮出现的概率为 0.3

功出现的概率为 0.3

根据邮件 B 生成 hashtable_good,该哈希表中的记录为

法:1

律:1

计算得在本表中:

法出现的概率为 0.5

律出现的概率为 0.5

综合考虑两个哈希表,共有 4 个 TOKEN 串:法 轮 功 律

当邮件中出现“法”时,该邮件为垃圾邮件的概率为

$$P=0.3/(0.3+0.5)=0.375$$

出现“轮”时:

$$P=0.3/(0.3+0)=1$$

出现“功”时:

$$P=0.3/(0.3+0)=1$$

出现“律”时

$$P=0/(0+0.5)=0;$$

由此可得第 3 个哈希表:hashtable_probability 其数据为

法:0.375

轮:1

功:1

律:0

当新到一封含有“功律”的邮件时,可得到两个 TOKEN 串:功律。

查询哈希表 hashtable_probability 可得

$$P(\text{垃圾邮件}|\text{功})=1$$

$$P(\text{垃圾邮件}|\text{律})=0$$

此时该邮件为垃圾邮件的可能性为：

$$P=(0\times1)/[0\times1+(1-0)\times(1-1)]=0$$

由此可推出该邮件为非垃圾邮件。

贝叶斯技术的出现，为基于内容分析的反垃圾邮件发方法提出了一条新路。它既克服了传统内容分析技术准确性低、误报率高的缺陷，又不需要预先搜集和编制关键词表，特别是结合了其他垃圾邮件分析技术后，可以实现对样本的自动采集和学习，而不需要管理员的额外干预。由于贝叶斯分析技术具有这些优点，因此，它在各种反垃圾邮件系统中都得到了广泛的应用。

但贝叶斯技术也不是无懈可击的。它还是一种基于内容分析的方法，因此，通过对内容进行特殊处理，比如将关键的内容改成图片等，让贝叶斯过滤器无法找到可分析的内容，就可以逃过贝叶斯技术的检查。

6.4 新型技术：分布协作的内容指纹分析

分布协作分析是基于这样的实际情况：垃圾邮件发送者将相同的邮件发送给巨大数量的邮件地址，可能试图从中取得某些商业、政治利益。

这种邮件绝大部分是使用假的邮件地址、伪造了邮件头或利用了开放式转发功能发送的，只有内容是这种垃圾邮件要传递的信息。但每个收件人必须看了内容以后才知道这是垃圾邮件。因此分布协作分析是基于这样的方法：从邮件中提取出可以代表内容的指纹数据(一般是利用加密哈希算法或检查和的算法来产生指纹特征)，不同的内容会产生不同的指纹(为了防止垃圾邮件发送者利用小的变化，如大小写等，来躲避反垃圾系统，一般还做模糊指纹，使相似内容的邮件产生相同的指纹)，用这些指纹代表邮件，全球的兼容用户会提交邮件(或只提交垃圾邮件)的指纹，从服务器得到响应，以知道有多少封相同的邮件在全球传播，这样来识别邮件是否垃圾邮件。

分布式哈希数据库分析邮件内容并根据指纹算法产生指纹。分布式哈希数据库将认为是垃圾邮件(或全部邮件)的指纹特征提交给分析服务器，由分析服务器根据提交总数量、频率等参数认定它是垃圾邮件的可能性。这样的系统可以集合全球的邮件特征，将大范围发送、但局部数量并不多的垃圾邮件甄别出来。所有支持分布式哈希数据库的防垃圾邮件系统都会互相促进，提高准确性。

6.4.1 DCC

DCC(Distributed Checksum Clearinghouse[①])是一个合作的、分布式的反垃圾邮件系统,目的是检测出巨大的邮件或发给大量人的邮件。该系统提供一个机制,使一个收件人收到一封邮件后可以知道其他人是否收到同一邮件,如果其他人也收到同样的邮件且数量巨大,那一般就是垃圾邮件。

使用某种 Checksum 算法,可以保证任意两封不同的邮件产生不同的 Checksum,这样,该 Checksum 就可以代表一封邮件,或者可以理解该 Checksum 是该邮件的指纹。

DCC 客户端将接收到的所有邮件的 Checksum 提交给 DCC 服务器,DCC 服务器按不同的 Checksum 累计所有 DCC 客户端提交的收件人数量(注意:一次提交可能不止一个收件人,因此累计也不是加 1),并将当前该 Checksum 的累计数作为响应告知 DCC 客户端。可以想象,一封正常邮件的收件人数量是很少的,而同一封垃圾邮件的收件人可能是成千上万甚至更多,如果同一 Checksum(实际就是同一封邮件)的 DCC 客户端提交数(实际就是该邮件发到具有 DCC 客户端的邮件服务器的数量)很大,超过了 DCC 客户端的最大值设置,DCC 客户端就可以认为这是一封垃圾邮件,根据其自己的策略,可以拒绝该邮件。

由于简单的 Checksum 对垃圾邮件来说效果不好,因此 DCC 的 Checksum 使用模糊的 Checksum,且随着垃圾邮件的发展而更新。从 DCC 开始使用的 2000 年底到现在,模糊 Checksum 已经修改了若干次。

DCC 客户端向 DCC 服务器查阅一封邮件需要一对大约 100 字节的 UDP 包,这个数据量甚至低于 DNS 请求,但当大量的邮件请求蜂拥而来时,也会造成延时,因此 DCC 使用分布式的服务器阵列,DCC 客户端应该选择最近、速度最快的 DCC 服务器。而各 DCC 服务器之间会交换或同步数据,这些数据只是 Checksum,且只交换或同步看来是垃圾邮件的 Checksum(比如累计数很大),因此不会引起拥塞。

对于发送给很多人的邮件,DCC 采用自动接收上报的 Checksum 自动计数的方式,而对于巨大邮件,DCC 采用收件人投诉提交的方式。注意:正常的邮件应该不会被人投诉提交,因此不影响正常邮件的发送,即不会被诬陷为巨大邮件。

DCC 建议每天邮件数量超过 10 万封的邮件服务器应该自己运行 DCC 服务器,并将本地 DCC 服务器与全球的 DCC 服务器进行数据同步。

DCC 服务器有个特殊的累计值 MANY,表示数量已经巨大,一般来说可以肯定是垃圾。每台 DCC 服务器的门槛是不一样的,比如一台只有 100 个用户的服务器收到一封同时发给 80 个用户的邮件,很可能是垃圾邮件;但一台有 10 万个用户的服务器收到一封

① http://www.ccc-servers.net/dcc/ http://www.rhyolite.com/anti-spam/dcc/

给 200 个用户的邮件,也许不是垃圾邮件。

随着时间的推移,DCC 服务器可能保留了大量的 Checksum,可能无限增长,因此对于时间很长、没有再有新提交的 Checksum,提供了"忘却"机制。

DCC 主要的功能是防止大规模群发的垃圾邮件,因此对于正常的大规模群发(例如邮件列表)有可能产生误判。

DCC 本地(客户端)使用 udp 1023 端口,远程(服务器端)使用 udp 6277 端口,因此要保证从客户端到服务器的防火墙等所有 ACL 开放这样访问的双向通道。

6.4.2 Razor

Vipul's Razor[①] 是一个分布式、协作的垃圾邮件检测和过滤网络。

和 DCC 不同的是,Razor 不是由客户端上报所有邮件的指纹,而是要求人工上报确认是垃圾的邮件。因此 Razor 提供了提交工具,如 Windows 的工具 SpamNet、Razor2 等,其中 SpamNet 是一个 Outlook 的插件。

由于是人工提交,因此 Razor 使用 Truth Evaluation System (TeS)来跟踪提交人的可信度。

需要访问远程的 TCP 端口 2703(Razor2)和 7(Echo)。Razor2 使用 TCP 的 ping 指令来查看哪个服务器离它最近。

通过用户的提交,Razor 建立一个分布式的、持续更新的、目前流传的垃圾邮件的目录,以帮助邮件客户端过滤已知垃圾邮件。通过统计和指纹随机化的检测方法可保证有效地识别变异的垃圾邮件内容。该体系基于用户对每个被交邮件的共识以及提交被撤回的情况。用户提交的内容通过可信度体系来验证。

Razor 的协议基于结构化信息串交换,类似 URI,可以用 URI 解码库解析。支持管道,这样 Razor 代理可以和服务器建立连接而消除 TCP 的 3-way 握手和 4-way breakdown 带来的隐患。该协议可以无缝导入新的指纹机制。

现在使用 Nilsimsa 模糊指纹算法,基于一段文字中发生的 n-gram 统计模型。Nilsimsa 忽略文本中与统计无关的小的变化。Nilsimsa 签名可以比较准确确定源文本的相似度(0~100%)。

短时指纹是基于协作计算随机数的有效时间的很短的指纹。短时签名基于一个随时变化的随机数从垃圾邮件中选择一段文本。这使哈希机制成为一个变化的目标,垃圾邮件发送者不可能发掘使用,因为在随机数变化后,他们不知道邮件的那个部分被哈希。

Razor 支持多个预处理器。预处理器在计算哈希前修改垃圾邮件的文本,包括对

① http://razor.sourceforge.net

Base64 或 QP 编码邮件的解码、把 HTML 转换为纯文本等。垃圾邮件发送者会使用多种技术以不同编码变化隐藏垃圾。预处理器通过对 MUA 实际看到的内容做哈希来击败垃圾邮件发送者的变化。

Razor 支持多引擎。引擎是封装了一个特定类型的过滤服务的逻辑单元。目前 Razor 有 4 个引擎：基本引擎、基于邮件体的 SHA1 指纹的引擎、Nilsimsa 指纹引擎和短时哈希引擎。在需要的时候还可以无缝地插入新引擎。

指纹使用 64 个数字的编码，而不是 16 个(十六进制数)，相当于是用 64 进制编码，减少了约 1/3 的网络传输。

Razor 具有一个透明的后端部件：TeS。TeS 是一个可信度系统和启发式识别模板的组合，针对每个指纹，给出提交人的可信度(0～100)。用户可以在其 Razor 配置中设置可接受的可信度。服务器公布建议可信度。TeS 已经排除了合法巨大邮件。

Razor 接受新垃圾邮件的整个邮件体文本。这使 Razor 可以每 n 小时计算新的短时指纹作为新指纹机制和预处理器带来的数据库的种子。注意：在检查过程中，Razor 不接受合法邮件的内容，在检查邮件时只有指纹被发送。

Razor 还允许用户撤回他们不认为是垃圾邮件的提交。撤回请求送到 TeS 中，改变指纹的可信度，如果必要，会从数据库中删除。

Razor 要求提交人注册，并给提交人建立可信资料，根据提交人的可信度，对垃圾邮件的提交和撤回具有不同的权重。提交要求用户使用 CRAM-SHA1 认证机制认证。

Razor 引进了内容分类的概念。一个内容类是相同内容的不同变化的邮件集合。处理提交时，相关服务器首先匹配存在的内容类。另外，Razor 将每个 MIME 附件按一个单独的内容类看到，这样垃圾邮件的 MIME 附件会被单独跟踪，这对病毒邮件特别有用。

6.4.3 Pyzor

Pyzor[①] 和 Razor 类似，是一个利用摘要算法产生的邮件指纹检测和屏蔽垃圾邮件的协作化网络系统。

Pyzor 一开始是 Razor 的 Python 实现，但后来使用了不同的协议从而分道扬镳。

Pyzor 允许全世界的 Pyzor 兼容用户提交垃圾邮件并比较接收到的邮件，以自动确定其是否垃圾邮件。Pyzor 对邮件体做数字指纹，然后查询中央服务器，看是否有别人提交了相同的指纹并认为是垃圾邮件，如果有，就应该对该邮件做特殊处理，例如转发到垃圾邮件箱或在邮件头上打上垃圾标记。

Pyzor 建议用户建立垃圾探针邮箱，专用于接收垃圾邮件向 Pyzor 服务器提交。与

① http://pyzor.sourceforge.net

DCC、Razor 等分布协作分析系统类似,提交的垃圾邮件越多、越快,分析系统越准确。

Pyzor 需要使用 udp 和 tcp 的 24441 端口。使用 udp 和服务器通信,但服务器响应使用 tcp 连接。

Pyzor 使用步骤如下。

(1) 对进入的、要处理的邮件进行黑名单处理。用黑名单屏蔽所有确定的垃圾邮件。一般的技术是维护一个商业垃圾邮件发送者的列表,并在这一步屏蔽掉。例如,对于来自 mediatrec.com 的所有邮件,可能想都屏蔽掉,只需将其列入黑名单即可。

(2) 对进入的邮件进行白名单处理。白名单允许接收所有确认不是垃圾的邮件,这可能包括所有用户认识的并想接收其邮件的地址,或者订阅的邮件列表。

(3) 基于内容过滤。

基于内容的过滤检查邮件的邮件体,并使用非确定的、基于启发式的内容处理方法发现垃圾邮件。

一般要求在基于内容的过滤前一定要使用白名单,以保证已知合法来源的邮件不会被误判为垃圾邮件。任何使用启发式或非确定方案的过滤系统都无法保证不发生误判,只有黑名单和白名单才不会误判(注意:白名单可能被垃圾发送者利用而发生漏判)。因此如果白名单不够完全,就有可能因为误判导致合法的邮件无法收到。

Pyzor 就是一种基于计算邮件某些部分的摘要(指纹)来匹配其他人已经接收并提交的邮件指纹来判断垃圾邮件的一种基于内容的过滤方法。其他一些反垃圾邮件产品,例如 OpenSource 的 SpamAssassin 就使用了 Pyzor 作为其一个部件,即判断是否是垃圾邮件的一个依据。

分布协作的内容指纹分析搜集垃圾邮件的统计数,提取和总结其特征值,并以此作为依据对邮件进行判断。这种技术在对付由蠕虫或病毒爆发造成的垃圾邮件时有非常好的效果,这类邮件由于数量庞大,容易被搜集和分析,同时各封邮件的内容基本不会有变化,因此特征值也保持不变,容易被检验出来。但如果垃圾邮件流传的规模比较有限,或者内容会根据不同的收件人动态变化,这种技术也就无能为力了。另外,该技术要求邮件服务器的管理员和用户能够自觉提交垃圾邮件样本以进行分析,而国内的用户往往没有这方面的习惯和意识,因此该技术在国内的作用也是有限的。

6.5 辅助工作

1. 集中的分领域内容分析

传统的独立的内容过滤方法对于分辨垃圾邮件效果很差,而分领域的内容分析则可以弥补这一点。分领域的内容分析技术根据邮件的内容中所涉及的领域将邮件分类,并

对不同类别的邮件分别分析其特征。预先将邮件分类，以在更小的范围内总结和提取邮件的特征，可以更加有效地获取垃圾邮件的内容特征，从而提高内容分析的准确性。

2. 打分机制

在前面的章节里，已经提到了目前大量的反垃圾邮件技术和方法，但除了黑名单和白名单外，所有方法，包括可追查性检查在未完全实施前，都可能发生误判，而对于用户来说，少量的漏判是可以忍受的，但即使是极少量的误判用户可能都不愿接受。因此，为了提高反垃圾邮件判断的正确性，在实际构造反垃圾邮件系统时，往往不会采取单一技术的一票否决制，而是综合采用多种反垃圾技术，每种反垃圾邮件技术都会判断一封邮件是否是垃圾邮件，并据此给出一个分值，所有的判断完成后，将全部分值加权累加，最后根据计算得出的总分值判断一封邮件是垃圾邮件的可能性。

但仍然要强调的是：即使使用打分机制，仍然存在误判的可能性，只不过可以将分值极高的邮件认为是垃圾邮件，分值极低的邮件认为是正常邮件，而中间分值的邮件建议采用标记的方法，而不是直接拒绝，由收件人去判断它是否垃圾邮件。

3. 降噪

智能降噪技术利用过去累积的关于垃圾邮件的知识来修正当前对垃圾邮件的判断结果。例如，一个合法邮件发送者平时都是发送正常邮件，但有一天反垃圾邮件对它的判断结果里突然出现了大量垃圾邮件的情况，这有可能是因为某种原因的错误判断，也就是可能是噪声，因此，可以对其进行降噪处理，即认为他发送的邮件是垃圾邮件的概率相对比较低。相反的，如果一个地址平常大量发送垃圾邮件，但是突然有一天垃圾邮件非常少，那么也会认为它可能是采用了新的逃避反垃圾邮件检查的方法，需要相应增加它的垃圾邮件概率。

4. 互动

系统的所有引擎和子引擎应该是互动的，即可追查检查、智能 RBL、智能集中分析和智能内容分析得出的垃圾邮件和非垃圾邮件都会作为贝叶斯自学习的样本，同时贝叶斯过滤器认为是垃圾邮件的邮件会作为智能集中分析等子引擎的提交邮件，充实智能集中分析系统的特征库。

5. 垃圾探针邮箱

垃圾探针邮箱又被成为蜜罐系统(Honey Pot)。这是一种收集垃圾邮件样本的有效方法，简单说，就是主动构建一个陷阱邮箱，并诱使垃圾邮件制造者向该邮箱发送垃圾邮件，从而获取分析用的垃圾邮件样本。

首先,垃圾邮件发送者要得到邮件地址,可能是从邮件列表得到,从网页得到,或者交换和购买。

邮件地址收集软件可以自动扫描新闻组、邮件列表和网页,从中找到任何看起来像邮件地址的字符串,存入数据库或文件,作为发送垃圾邮件的目标地址列表,这个列表可能非常大。当用户向新闻组提交一篇文章,或把邮件地址放到任何网页上时,就可能被邮件地址收集软件捕捉到。

根据垃圾邮件发送者搜集、共享邮件清单的方法,可以建立多种垃圾探针邮箱,这些探针邮箱有的可以100%保证是垃圾邮件,有些可能会引来一些正常邮件。

- 隐蔽邮件账号:在邮件系统上建立完全不对外公开的隐蔽邮件账号,开放不存在用户的退信。这样,邮箱地址搜索方式(例如随机发邮件等候退信或无退信)会得到这些隐蔽邮件账号。由于隐蔽邮件账号完全不对外,因此凡是向这些账号发来的邮件肯定100%是垃圾邮件。
- 非显示邮件账号:在网页上公布探针邮箱账号,但该账号的文字颜色与网页背景颜色相同,正常的人是无法看到该账号的,但用于搜索垃圾邮件发送目标的软件可以发现,因此该探针邮箱地址收到的邮件必然100%是垃圾邮件。
- 网站注册:使用专门的邮件账号,不用于任何其他目的,仅用于在各网站上注册。对于不负责任的网站,会将注册用户的邮件地址泄漏出去,成为垃圾邮件发送者的目标。只要过滤掉网站来的邮件,剩下的就肯定是垃圾邮件。
- 订阅邮件列表:使用专门的邮件账号,同样不用于其他目的,仅用于订阅邮件列表。对于不负责任的邮件列表,会将订阅人的邮件地址泄漏出去,成为垃圾邮件发送者的目标。收件系统只要将从邮件列表发来的邮件过滤掉,剩下的就肯定是垃圾邮件。另外,有些垃圾邮件发送者可能直接利用邮件列表发送垃圾邮件,这时正常的邮件列表邮件和利用邮件列表的垃圾邮件就更难以区分了,因此邮件列表应该要求必须使用列表的注册用户作为发件人才予以转发。
- 其他公开方式:例如在论坛上公开探针邮箱等。但这种方法可能引来合法的邮件,无法100%保证收到的是垃圾邮件。

在上述公开方式中,有可能引来正常邮件,因此针对垃圾邮件收集程序不可能理解邮件地址的情况,可以将用户名定义为:人可以理解,看到这个邮件地址,正常人就不会向该地址发送邮件的用户名,例如NoPeply。

作为具体的方法,为了方便,可以使用一个垃圾邮件探针邮箱,给这个探针邮箱设置多个别名,这样所有的垃圾邮件都用这个邮箱保存,而对外仿佛是很多个邮箱,容易迷惑垃圾邮件发送者。同时,在不同的发布地点使用不同的探针邮箱别名(或干脆用不同的探针邮箱地址)还可以分析出垃圾邮件发送者如何收集邮件地址,这些地址又是如何被传播的。

6.6 反-反垃圾技术方法

前面的介绍了一些传统的反垃圾邮件技术。这些技术虽然能起到一定的效果，但都存在着这样或那样的不足，在单独用于防止垃圾邮件时效果都不理想。垃圾邮件制造者可以通过有针对性地改进垃圾邮件制造和发送技术等方法，来逃避这些方法检查。垃圾邮件制造者可能采用的反-反垃圾邮件方法包括以下几种。

1. 利用动态 IP 地址来发送垃圾邮件

垃圾邮件制造者通过利用动态分配的 IP 地址来对抗黑名单、动态黑名单等基于 IP 地址封锁的反垃圾邮件技术。尤其在宽带网、智能社区高速发展的今天，很多宽带网接入的用户使用的是动态分配的 IP 地址，每次用户联网时使用的都是不同的 IP 地址，因此在这种情况下，基于 IP 地址封锁的反垃圾邮件技术一方面会导致大量的无辜用户被误封锁，另一方面也无法对垃圾邮件制造者实现有效的封锁。

2. 反内容分析

- 关键词变形：传统的内容分析技术依赖于对关键词的提取。针对这一点，垃圾邮件制造者可以有针对性地对关键词进行变换，从而达到对抗检查的目的。例如，某个内容过滤系统可能将“赚钱”作为关键词，如果垃圾邮件制造者在“赚钱”两字中插入标点符号，将其变成“赚……钱”之类的形式，使内容过滤系统无法正确对关键字进行提取，就可以逃避这类内容分析技术的检查。
- 内容图片化：面对简单的关键字变形方法，反垃圾邮件系统可以通过增强内容分析器的分析能力来化解。但是如果垃圾邮件制造者更进一步，将关键内容用图片来表示，内容分析技术就无能为力了。因为无论何种内容分析技术，都是基于文本内容来进行分析的。而要将图片转化为文本进行分析，现有的技术一来还不够成熟，转化正确率还不够高；二来这种转化本身的计算量也非常大，会大大增加垃圾邮件处理的成本，因此在现阶段对于图片内容的分析和过滤还是个难题。
- 内容附件化：和内容图片化类似，这种技术将邮件内容放在某种反垃圾邮件系统无法直接处理的文档中，作为附件和垃圾邮件一起发送。例如，Word 文档，由于微软一直没有公开其具体格式，所以其他人都无法正确解析文档里面的内容。对于这种附件，内容过滤系统也是无能为力的。
- 信头、正文随机化：使用随机内容产生器产生随机的信头、内容和附件名。

- 图片加噪：对附件图片做随机的、不影响图片阅读效果的噪声处理，可以躲避所有内容分析。

3. 分散发布源

垃圾发送者合作相互转发，使垃圾邮件不会重复从某一个 IP 地址出现，以逃避反垃圾邮件检查。

4. 结合蠕虫的功能

将垃圾邮件发送分散到世界各地可被蠕虫侵染的计算机，可防止智能分析中对发送邮件 IP 地址的分析。

除了这些有针对性的反-反垃圾邮件方法外，垃圾邮件制造者们还有一个更有效的绝招，那就是隐藏真实发件人地址。对真实发件地址进行隐藏之后，收件人无法找到垃圾邮件制造者的真实身份，这样，即使垃圾邮件被发现或过滤，垃圾邮件制造者本身也不会有任何危险和损失。再加上制造垃圾邮件几乎没有任何成本，只要反垃圾邮件系统达不到百分之百的过滤效果，垃圾邮件制造者们依然可以肆无忌惮地发送垃圾邮件，最多不过是多发点而已。也就是说，前述的各种反垃圾邮件技术，虽然能起到一定的反垃圾邮件效果，但对垃圾邮件制造者的威慑力为零，因而也就无法真正解决垃圾邮件问题。

6.7 彻底解决反垃圾邮件问题：源头认证

前面已经讨论了，要想真正彻底的解决垃圾邮件问题，只有真正对垃圾邮件制造者们形成有效的震慑才行。那么，如何才能达到这一点呢？显然，这首先需要依靠法律，运用法律武器对那些垃圾邮件制造者进行严厉制裁，通过法律制造出一条难以逾越的高压线来震慑垃圾邮件制造者；另外，必须找到一种方法来解决邮件伪装的问题，特别是要确保邮件发件人是真实可靠、未经伪装的，从而可以作为法律惩罚的依据。这两者结合在一起，就可以形成对垃圾邮件制造者的有效震慑。

所谓源头认证，就是通过某种方法来防止邮件伪装，特别是确保发件人不是伪装的，从而彻底断绝垃圾邮件发送者躲避法律制裁的路。但由于电子邮件系统本身原本就存在着缺陷，因此，虽然早就有人想到利用源头认证技术来反垃圾邮件，但是长期以来，一直未能真正出现一种有效的源头认证技术，直到近几年可追查性等新技术出现以后，源头认证技术才真正实用化。

6.8 质询-回应技术

质询-回应(Challenge-Response)技术是一种传统的身份识别技术。简单地说,这种技术就是识别方向被识别方问一些只有这两方才可能知道正确答案的问题,通过被识别方回答的正误来判断被识别方的身份。例如,在著名小说《林海雪原》中,我侦察员杨子荣伪装成土匪深入敌穴,敌匪首座山雕为了验证杨子荣是不是真正的土匪,和杨子荣有一段著名的精彩对白:

座山雕瞪着像猴子一样的一对圆溜溜小眼睛,撅着山羊胡子,直盯着杨子荣。八大金刚凶恶的眼睛和座山雕一样紧逼着杨子荣,每人手里握着一把闪亮的匕首,寒光逼人。座山雕三分钟一句话也没问,他是在施下马威,这是他在考查所有的人惯用的手法,对杨子荣的来历,当然他是不会潦草放过的。老匪的这一着也着实厉害。这三分钟里,杨子荣像受刑一样难忍,可是他心里老是这样鼓励着自己,"不要怕,别慌,镇静,这是匪徒的手法,忍不住就要露馅,革命斗争没有太容易的事,大胆,大胆……相信自己没有一点破绽。不能先说话,那样……"

"天王盖地虎。"

座山雕突然发出一声粗沉的黑话,两只眼睛向杨子荣逼得更紧,八大金刚也是一样,连已经用黑话考察过他的大麻子,也瞪起凶恶的眼睛。

这是匪徒中最机密的黑话,在匪徒的供词中不知多少次的核对过它。杨子荣一听这个老匪开口了,心里顿时轻松了一大半,可是马上又转为紧张,因为还不敢百分之百地保证匪徒俘虏的供词完全可靠,这一句要是答错了,马上自己就会被毁灭,甚至连解释的余地也没有。杨子荣在座山雕和八大金刚凶恶的虎视下,努力控制着内心的紧张,他从容地按匪徒们回答这句黑话的规矩,把右衣襟一翻答道:

"宝塔镇河妖。"

杨子荣的黑话刚出口,内心一阵激烈的跳动,是对?还是错?

"脸红什么?"座山雕紧逼一句,这既是一句黑话,但在这个节骨眼问这样一句,确有着很大的神经战的作用。

"精神焕发。"杨子荣因为这个老匪问的这一句,虽然在匪徒黑话谱以内,可是此刻问他,使杨子荣觉得也不知是黑话,还是明话?因而内心愈加紧张,可是他的外表却硬是装着满不在乎的神气。

"怎么又黄啦?"座山雕的眼威比前更凶。

"防冷涂的蜡。"杨子荣微笑而从容地摸了一下嘴巴。

……

这就是一段典型的质询-回应,当然其结果我们都知道了,这次质询-回应的结果是失败的,因为杨子荣通过了座山雕的考验。但那是因为杨子荣已经预先通过其他方法知道了回应的正确内容。必须声明,质询-回应方法本身还是一种有效的身份认证方法。

不过具体到邮件系统中,传统的质询-回应方法是没有办法使用的,因为那要求反垃圾邮件系统预先和全部可能的发件人定下一个(或多个)秘密问题及其答案,这显然是不现实的。但对于这个问题,可以换个角度来考虑。虽然电子邮件系统本身存在很多漏洞,邮件地址很容易伪装,但是其投递过程还是很可靠的,即只要收件人邮箱存在并且没有被塞满,还是能保证邮件被投递到收件人邮箱中的。因此,可以对传统的质询-回应方法稍作修改,不是直接问投递人秘密问题,而是将邮件中的发件人地址作为收件人,回发一封质询邮件,问他:有封邮件声称是你发的,是不是真的?如果回答是,那么,就可以认为这封邮件的发件人地址是真实的,反之,那发件人地址肯定是伪装过了。

这种方法实现起来比较简单,也确实能解决邮件地址伪装的问题,因此一经提出就被不少系统尝试采用。但很快大家就发现这种方法存在重大的缺陷:

- 很多系统会自动产生邮件,比如邮件列表等。而这些自动产生邮件的系统,绝大多数都不具备回应质询的能力。这就会使反垃圾邮件系统认为这些系统自动产生的邮件是伪装过的邮件,被错误地丢弃。这就引起了用户的不满。而要让这些系统增加回应质询的能力,需要做大量额外的工作,这并不现实。
- 更严重的问题发生在两个都采用了质询—回应技术的邮件系统之间。假设有这样的两个系统,分别为 A 和 B。现在 A 上的用户 x 发了一封邮件给 B 上的用户 y。B 收到 x 发给 y 的邮件后,就会回送 A 一封给 x 的质询信。而 A 收到这封信后,又会产生一封给 B 的质询信,B 收到质询信后,又会给 A 产生一封质询信……这样就会在 A 和 B 之间一直不断地循环产生质询信,而 x 给 y 的邮件始终无法到达 y 手中。

面对这些棘手的问题,质询—回应系统难以给出理想的解决方法。另外,这种质询系统也增加了邮件发送者的负担,给邮件发送带来了不便,会引起用户的抵制。因此,这种方法并不是一种实用的反垃圾邮件方法。

6.9 DKIM

DKIM(DomainKeys Identified Mail,邮件域名密钥验证)技术是雅虎公司提出的一种源头认证技术。通过这种技术,反垃圾邮件系统能够做到以下两点:验证邮件发件人

是否确属于他所声称的邮件域；保证邮件内容本身的完整性(即没有被篡改)。一旦邮件域通过了验证，就可以用它与邮件头中的 From 字段几行比较，从而发现地址是否被伪造。如果地址是伪造的，那么这封邮件就是垃圾邮件，可以直接丢弃而无需惊动用户。如果地址是真实的，系统就会记住该邮件域，并在反垃圾邮件系统内部产生一条对应的策略，供以后的判断使用。

对于那些经常向消费者寄送交易信息的公司，如银行、公共服务公司、电子商务公司等，这种验证的好处就更大了，因为它可以帮助这些公司保护它们的用户免受邮件欺诈攻击。在这类攻击中，欺骗者将欺诈邮件的发件地址伪装成这些公司，并在内容中模仿这些公司的格式和语气，诱使用户相信其中的内容，并套取用户的账号和密码等账户信息。对这些公司而言，保护用户免受此类攻击就意味着用户的满意和忠诚，用户服务费用的降低，以及品牌的保护。

对于用户而言，无论使用雅虎邮件这种大型邮件提供商还是小型 ISP 提供的邮件服务，发送者身份的保护技术意味着他们可以信任电子邮件了，可以重新利用邮件这种当代最有效的通信手段来完成各种通信任务。

6.9.1 版权声明

Domainkeys 其实是由无数业界伙伴正在合作进行的一项共同开放电子邮件认证标准的项目成果，业界合作在其中起到了至关重要的作用。很多业界领袖在其中扮演了重要的角色，包括 Alt-N 科技，AOL，Brandenburg Internetworking，思科，EarthLink，IBM，Microsoft，PGP Corporation，Sendmail，StrongMail Systems，Tumbleweed，VeriSign 和雅虎等。正是这些公司的参与才使这项基于签名的电子邮件认证技术的产生成为可能。雅虎会继续保持与这些机构以及 IETF 等组织在 DKIM 草案标准化工作上的合作，以使该技术能够成为业界认可的验证邮件发件人身份的最佳技术。DKIM 技术正在 IETF 标准化过程中，最终将成为一项 IETFInternet 标准。

作为历史性参考，雅虎已经将 DomainKeys 框架作为一项 Internet 草案提交，题目是《 draft-delany-domainkeys-base-03. txt 》。雅虎的 DomainKeys 知识产权可以在基于以下协议的基础上授权：

- Yahoo! DomainKeys Patent License Agreement
- GNU General Public License version 2. 0 (and no other version).

雅虎的知识产权包括下列的专利和专利申请：

- U. S. Patent Number 6,986,049,2006 年 1 月 10 日
- U. S. Patent Application Serial Number 10/805,181,2004 年 3 月 19 日
- PCT Application PCT/US2004/007883,2004 年 3 月 15 日

• PCT Application PCT/US2005/008656,2005 年 3 月 15 日

雅虎公司已经将上述授权声明与 RFC 2026 一起提交 IETF 讨论。

6.9.2 参考实现

除了 Internet 草案外,雅虎公司还开发了一个 DKIM 的参考实现,其可以作为一个插件集成到现有的 qmail 等 MTA 系统中。该实现可以在网站 http://domainkeys.sourceforge.net/下载。Sendmail 公司也为其邮件系统开发了一个 DomainKeys 的实现。实际上,Sendmail 已经提供了一个开放源码的 domainkeys 实现,并且正在积极地征求该实现的使用意见和志愿开发人员。

6.9.3 工作原理

对于邮件发送方而言,在发送邮件前首先必须对邮件进行签名。这需要两步来完成:

(1) 初始化。邮件域的管理者(通常是公司或邮件服务提供商内负责邮件系统运行的部门)为所有要签名的信息产生一对(或者多对)公钥和私钥。公钥通过 DNS 对外公布,而私钥则被储存起来,供 DomainKeys 邮件系统使用。这一步骤在图 6-1 中被标为"A"。

(2) 签名。当邮件域内的合法用户对外发出一封邮件时,DomainKey 邮件系统会自动使用储存的私钥来为该邮件产生一个数字签名,并将该签名作为一个字段添加到邮件的邮件头中。然后,再将该邮件发往收件人所在的邮件服务器。这一步骤在图 6-1 中被标为"B"。

对于邮件接收方而言,需要 3 个步骤来完成对邮件签名的验证:

(1) 准备工作。DomainKeys 接收服务器首先从邮件头中取出 DomainKey 数字签名和发件人信息,然后根据发件人信息通过 DNS 系统取回对应的公钥。这一步骤在图 6-1 中被标为"C"。

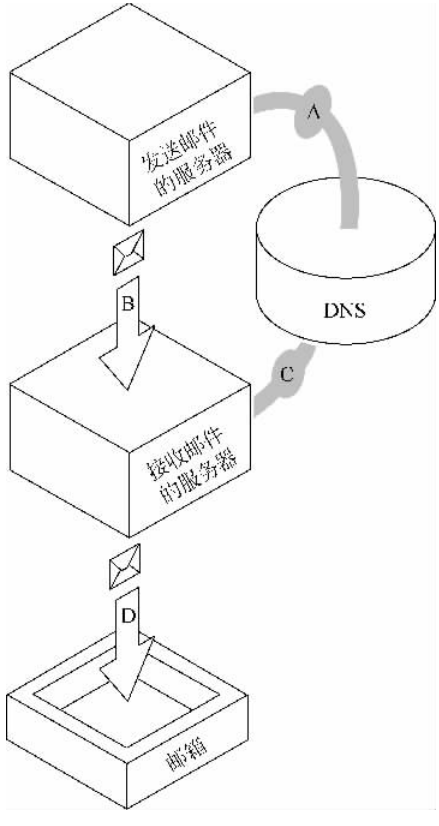

图 6-1 DKIM 工作模型

(2) 验证。接收邮件服务器使用获取的公钥来对数字签名进行验证。如果验证通过,就证明该邮件确实是从其所声称的邮件域发送来的,并且其邮件头和邮件体在投递过程中都没有被改动。

(3) 投递。接收邮件服务器根据签名验证的结果和本地策略来对邮件进行投递。如果签名通过了验证,并且邮件本身也通过了其他反垃圾邮件的检查,那么就可以将邮件投递到用户邮箱中。如果签名验证失败,或者邮件中没有包含签名,那么就可以将邮件丢弃,或者隔离。这一步骤在图 6-1 中被标为"D"。

一般而言,DomainKeys 签名应该在邮件接收服务器端被验证。但也可以在 MUA 中对签名进行验证,并根据结果进行相应的处理。

6.9.4 常见问题

1. DKIM 对于反垃圾有何帮助?

DKIM 对于反垃圾邮件的帮助体现在下面几点。

首先,它允许邮件接收方可以丢弃或隔离那些来源于已知对邮件进行签名的邮件域的未签名邮件,这样就可以有效抑制垃圾邮件或欺诈邮件。其次,它使邮件接收方具有了验证发件域的能力,这就使得建立一个共同信用数据库成为可能,各个邮件组织可以共享该数据库,对发件域的信用进行评价,并根据该数据库来指定对应的反垃圾邮件策略。例如,某个 ISP 可以将它们统计的从 www. example. com 发来的垃圾邮件和正常邮件比例与那些尚未收到过 www. example. com 的邮件的 ISP 们共享,后者可以利用该信息来对收到的来自 www. exmaple. com 的邮件进行反垃圾邮件处理。最后,通过消除发件人地址伪装,可以使电子邮件具备可追查来源的能力。垃圾邮件制造者们不希望被追踪,所以他们不得不只攻击那些不具备 DomainKeys 检查能力的邮件服务器。

2. DKIM 对于阻止欺诈攻击有何帮助?

那些饱受欺诈攻击困扰的公司可以将其发出的所有邮件用 DomainKeys 签名,并且对外公开宣布这一点。这样,其他邮件服务商就可以对邮件进行检查,并且自动丢弃那些声称来自于这些公司的邮件域的未签名邮件。例如,假如 www. example. com 公司对它发出的所有邮件使用 DomainKeys 进行签名,雅虎公司就可以在它的反垃圾邮件系统中增加一条策略,自动丢弃所有未签名或签名错误的声称来自 www. example. com 的邮件。这样就保护了 www. exmaple. com 的顾客和潜在消费者免受欺诈攻击。

3. 如果垃圾邮件制造者也用 DomainKeys 对垃圾邮件签名怎么办?

正希望如此! 如果他们这么做了,那么 Internet 组织就可以利用前述的信用数据库

来对他们的邮件进行隔离和丢弃。通过消除邮件来源的不确定性,可以大大简化反垃圾邮件的工作量,对所有的反垃圾邮件技术都有很大帮助。

4. DomainKeys 能验证什么?

DomainKeys 检查邮件头的 From 和 Sender 中的邮件域信息,以此来保护用户并且给予用户最好的使用体验。微软 Outlook 等桌面邮件客户端可以在它们的用户界面上显示邮件头信息。如果用户能够对这些邮件域建立信任,那么它们自然也会对用于保证这种信任的各种系统产生信任。

5. 为什么要对整个邮件签名?

DomainKeys 对整个邮件进行签名,这使接收邮件服务器能够验证邮件的完整性,即邮件在传输过程中是否被篡改或改变过。通过对邮件头和邮件内容进行签名,DomainKeys 使那些重复使用来自可信任域的邮件中的某一部分来骗过签名检查的手段不再可行。

6. DomainKeys 对邮件加密吗?

DomainKeys 对邮件并不进行加密,它只是在邮件头中增加一项数字签名,用于对邮件进行验证。

7. DomainKeys 使用哪种公钥/私钥对技术?

目前,DomainKeys 使用 RSA 算法来产生公钥/私钥对。实际产生的密钥长度由邮件域的管理者来决定。

8. 谁来产生 DomainKeys 需要的公钥/私钥对?

DomainKeys 邮件系统所使用的公钥/私钥对应该由邮件域管理者,或者代表域管理者的代理或服务提供机构来产生。

9. DomainKeys 需要第三方公证机构(CA)来对公钥进行签名吗?

DomainKeys 不需要第三方公证机构的签名。在通常的公钥/私钥保密系统中,需要第三方公证机构对公钥进行签名或认证,以使那些公钥的获取者能确保所得到的公钥确实是来自其声称的产生者。而在 DomainKeys 系统中,利用 DNS 系统作为公钥的发布源,由于在 DNS 系统中只有域管理员才能在对应域中发布信息,DomainKeys 的用户可以确保通过 DNS 获取的公钥确实是来自其对应的邮件域,所以就不再需要第三方公证机构的签名来保证这一点了。当然,通过增加第三方公证机构的签名,可以为 DomainKeys

系统提供更高的安全保证，因此并不排除在将来的 DomainKeys 系统中应用第三方公证签名的可能性。

10. 如何撤销 DomainKeys?

DomainKeys 系统允许一个邮件域在 DNS 系统中同时公布多个公钥。这就让那些公司可以为它们在同一个域下运行的不同邮件服务器使用不同的密钥对，或者可以利用多个密钥对来实现密钥的撤销、更换或过期等功能。这样，邮件域管理者可以任意撤销一个公钥，并且使用新的密钥对来进行 DomainKeys 签名。

11. 为什么不直接使用 S/MIME 技术呢?

S/MIME 技术是用来进行端对端的邮件签名和加密的，因此它在设计上独立于发送和接收服务器。而 Domainkeys 正与其相反。实际上，Domainkeys 应该是 S/MIME 技术在服务器对服务器方面的补充，而非替代。另外，由于很多对保密性要求高的行业内已经使用了 S/MIME 技术，所以必须保证 DomainKeys 和 S/MIME 技术能够共同使用而不会相互影响。最后，Internet 上使用的绝大多数的邮件服务商、客户端软件和服务器软件还没有支持 S/MIME 技术。实际上，要使 S/MIME 真正成为一项广泛使用的标准，要比 DomainKeys 难度大得多。

12. 邮件列表如何兼容 DomainKeys?

那些不对邮件内容或邮件头进行改动的邮件列表无须任何改动就可以配合 DomainKeys 工作。而那些对邮件内容或邮件头进行了改动的邮件列表服务器需要自行产生密钥对，并对邮件重新进行签名。

13. 如何实现 DomainKeys 技术?

通常，DomainKeys 技术应该由公司、ISP 或者邮件服务提供商内部负责部署和营运邮件服务器的部门来负责实现。某些公司可能由邮件服务器提供商来处理他们的邮件。随着 MTA 软件的提供商在其产品中增加对 DomainKeys 的支持，DomainKeys 的实现将会变得更加轻松和方便。

14. 目前有哪些 MTA 软件已经支持 DomainKeys 技术?

Sendmail 公司已经为他们的商业和免费版本的 MTA 软件发布了相应的 Domainkeys 插件。Qmail 也已经有了一个 domainkeys 功能的补丁。www 的创造者 CERN 也已经为 MS Exchange 2003 发布了一个 Domainkeys 的 C＃代码库，此外，如 Port 25 公司的 PowerMTA、Etype. net 的 acSMTP、ActivSoftware 公司的 XMServer、

OmniTI公司的Ecelerity和StrongMail、Alt-N Technology公司的MDaemon MTA for Windows、Postfix、Communigate Pro、IronPort、Merak Mail等产品都已经提供了支持Domainkeys的版本。最后,雅虎公司也发布了一个基于开放源代码的DomainKeys参考实现,供其他的MTA软件集成使用。

15. 应该如何部署DomainKeys?

在安装了支持DomainKeys技术的MTA软件之后,可以从几种密钥发布选项中选择一种。选择之后,产生的公钥部分应该发布到邮件域的_domainkey subdomain TXT记录中,私钥保存到MTA中。之后可以手工对DNS记录策略和选择符进行测试,或者也可以利用一些自动测试系统来进行测试。

16. 对于那些不使用所属邮件域的SMTP服务器发送邮件的客户,该如何应用DomainKeys技术?

DomainKeys依靠域管理者来对邮件中的发件人域进行认证。如果用户由于某种原因(如25端口被封)而无法使用域内授权的SMTP服务器来发送邮件,可以有以下几种解决方法。

- 在邮件域服务器中开放587端口,接受任务提交请求。通过给予用户一个提交邮件的路径,可以帮助邮件域管理员对域进行授权管理。
- 为有特别需要的用户单独产生一对用户专用密钥。用户可以使用它来直接在客户端对邮件进行签名。前述的任务提交服务也应该支持用户提交自己的密钥,并用它来对用户的邮件进行签名。
- 用户可以利用其他的邮件头信息来帮助识别用户身份。例如,邮件头中的Reply-to允许收件人的邮件客户端来选择回信的地址;Sender-to字段则定义了进行SMTP投递时所用的邮件地址。用户可以考虑在邮件头的From中声明自己的邮件域,并在Sender字段中声明实际用于发送邮件的邮件地址。但是必须注意,这种邮件头可能会被反垃圾邮件系统怀疑是垃圾邮件,因为垃圾邮件可能会利用这种手段对自己进行伪装。
- 用户可以选择发送一封无签名的邮件。这不是一个好的长期解决方案,但在Internet大多数邮件都采用验证方方案之前,这还是一种可行的手段。不过,如果选择了这种方案,用户必须经常检验自己邮件是否最终抵达了目的地,以防止自己的邮件因为未签名而被当作垃圾邮件丢弃。

DomainKeys是一种保证邮件地址未经伪造的有效方法。但它依然不是一项理想的方法,原因如下。

- 这项技术需要在现有的DNS系统中增加定义项,这就涉及到DNS系统的升级和

改造，这并不是一件容易实现的事情。

- DomainKeys 技术需要签名者显式地对其签名进行声明，并需要其他采用了 DomainKeys 技术的系统增加相应的策略才能对未签名邮件进行过滤，而否则就无法对未签名邮件进行过滤。这显然不利于在实际中对 DomainKeys 技术进行应用和部署。

因此，还需要寻找其他更好的源头认证技术。从第 7 章开始，将介绍一种实用的源头认证技术——可追查性检查。