

第33章 NNTP：互联网新闻组

作者：Daniel Baker

本章内容包括：

- 互联网新闻组
- 新闻组和层次
- 网络新闻传输协议
- 大量广告(Spamming)和新闻黑洞(Blackholing)

本章将讨论与新闻组相关的各种主题。新闻组是互联网通信的最为古老方式之一，它目前仍在使用。此外本章还讨论了如何与网络新闻传输协议(Network News Transfer Protocol, NNTP)服务器建立连接。

虽然NNTP并不是新闻组传输的原始协议，但是它使得所有早期的传输方法如UNIX到UNIX拷贝协议(UNIX——to——UNIX Copy Protocol, UUCP)被淘汰。

33.1 互联网新闻组

Usenet(Usenet News)，由两名Duke大学的研究生在1979年创建，是当前历史最长也是最流行的网络服务之一。Usenet允许在整个互联网范围内的组成员间用基于文本的方式进行交流。在这个方面，它非常类似于公共的、组电子邮件。

Usenet讨论是非实时的，文章从一个地点传输到另一个地点有时产生延误。然而，通常这个延误是微不足道的，不能阻碍互联网用户进行网上讨论的能力。

近年来，由于Usenet极为流行以及互联网快速增长，Usenet经历了实实在在的成长痛苦。许多年以前，设置一个新闻服务器仅需在互联网上随意接入一台电脑，并提供20~100M磁盘空间。然而今天，绝大多数Internet服务提供商(ISP)已拥有多个Usenet服务器，它们通常要运行一对骨干新闻发布版块(news box)负责用户的地址与互联网上其他地址间的通信，此外还需要包括若干客户服务器用来向最终用户提供Usenet文章数据。为了提供高质量的Usenet的服务，一个Usenet客户服务器必须保持不间断的、数兆带宽的互联网的连接，至少0.5GB的内存以及若干高端的微处理器。

技术局限性不是导致Usenet存在问题的仅有根源。缺乏集中式管理、强加的商业信息以及缺乏教育的用户都使Usenet较难扩大到其所需的规模。

注意 Usenetll(<http://www.usenet2.org/>), 一个新的Usenet，声称它通过加强法规及控制能够解决大多数Usenet存在的问题，使其能向正确的方向发展。

尽管还没有人能提出更好的方法，但大家都一致公认Usenet存储和传输文章的方法并不是很有效的。目前，每一个Usenet服务器都要保存一份发布在其体系内中的每一篇文章。

文章通过一种称为对等的方法从一个服务器传输到另一个服务器。新闻管理员可以与其他站点的对等层次的管理员进行谈判，决定他们与其他站点在Usenet的哪个层次上进行新闻交流。通常情况是，一个新闻服务器拥有1~200个对等服务器。一个服务器收到的每一篇文

章，都会与所有与它的对等服务器进行连接（除了把文章传输给它的那个服务器），然后用NNTP命令ihave将该文章传输给对等站点。

注意 由于这种存储方式的原因，即使一个很小的Usenet的文章，都会占用惊人的存储空间。举例说明，假如在互联网上有1500个Usenet服务器，一个合理的2500字节的文章在互联网范围内将占用3.75M存储空间(这个数字仅计算了存储这篇文章的服务器上的被占用空间，而在阅读该文章的客户机所占用的存储空间都不计算在内。一个CD音质的歌曲(大约3.5MB)发布到Usenet，将占用5.25GB的空间。一张运动的画片(大约1.2GB)发布到Usenet，将占用1.95TB的空间。

33.2 新闻组和层次

Usenet讨论组可以划分为成千上万个新闻组，传统上可以按其类别组织成分层的格式。然而，存在着一些层次最高的“根”层次。主要的层次被称为“七大”。

表33-1显示了“七大”最高层次，表32-2展示了其他流行的层次。

表33-1 “七大”最高层次

名 字	描 述
comp	计算
misc	其他各方面的
news	Usenet管理
rec	娱乐
sci	科学
soc	社会
talk	一般讨论

表33-2 其他流行的层次

名 字	描 述
alt	不受控制的“可选”新闻组
bionet	生物
clari	从路透社、美联社等实况转播的新闻信息
k12	包括12个年级的幼儿园

为了减少混淆，将不同的语言划分到不同的层次既符合习惯也较为方便。因此，不同的国家通常位于不同的层次，如表33-3所示。

表33-3 国家层次

名 字	描 述
ch	瑞士
fr	法国
de	德国

绝大多数顶级层次都受到严格控制以确保仅仅被授权的管理员才可以创建和删除新闻组。然而，最大的层次alt没有任何正式和有效的方法去这样做。无论如何，一个新闻服务器的管理员必须在接受询问的时候决定是否加强对层次的限制，然后该管理员必须对他或她的服务

器做出相应的配置。然而，新闻服务器管理员的有冲突的选择常常导致层次之间的不一致性。例如，一些服务器可能认为某些服务器存在，然而其他的不存在，因为新闻服务器管理员在层次规则上的宽容度不同。这些不一致对用户产生了极大的困扰，使 Usenet 的管理变得复杂。

提示 一个不错的想法是在发出任何控制消息之前应先对一个层次管理新闻组的政策有充分的理解。控制消息是Usenet服务器的新闻组管理脚本语言从语法上所分析的Usenet消息。控制消息可以要求Usenet的一个服务器创建一个新闻组、删除一个存在的新闻组、取消一个Usenet报文或者其他的管理请求。

确定 Usenet 层次所覆盖的主题后，用户应可以轻松地找到一个新闻组来讨论其所感兴趣的特定话题。例如，超过 350 个新闻组在其名字中包括 `unix`。其中最著名的新闻组就是 `comp.unix.solaris` 和 `com.unix.aix`。这样名字为一个组所关注的主要话题提供了一个清晰的提示。`comp` 表明一个与计算机相关的讨论组，而 `unix` 表明第二层次中的组与基于 `unix` 的操作系统相关。`solaris` 代表 Sun 公司的 Solaris。

关于一个组内所允许讨论的内容，层次的名字应给予用户关键性的提示。例如，在 `alt.binaries` 层次的每一个组中发布的都是一些二进制代码，例如计算机程序、音频文件、图片等。除非另做特别说明，一个组的内容可以放心地假设为 ASCII 码的英语文本。

提示 将二进制内容发布到非二进制的新闻组是一种极端恶劣的行为。例如，将一个音频的文件发布到 `alt.tv.seinfeld` 将会给新闻管理员带来极大的麻烦也违背了层次化的分层组织系统的初衷。然而，将同样的文件发布到 `alt.binaries.sounds.tv` 就比较合适，这也是一种将文件传遍整个 Usenet 的最有效途径。

至此，我们已经知道一些 Usenet 的基本概念以及它的组织方式，接下来看看 Usenet 是如何传输文章数据的。

33.3 网络新闻传输协议

网络新闻传输协议 (NNTP) 是新闻服务器与新闻客户间传输 Usenet 文章数据的最常见途径。NNTP 是一个 ASCII 协议通常在 TCP 端口 119 进行的通信。尽管用户可以直接与新闻服务器以 NNTP 连接 (正如本节中所描述的)，然而通常情况下用户的所有 NNTP 服务由一个新闻用户代理来处理 (包括发布、阅读等)。UNIX 用户经常使用的 NNTP 客户软件是 `tin` 和 `trn`，然而 Windows 用户通常使用 Netscape 和微软的 Outlook Express。

本节涵盖 NNTP 相关的各种话题，包括：

- 获取新闻组列表
- 获取特定的 Usenet 文章
- 发布消息到 Usenet

本节将使用户对 NNTP 的功能基础结构有一个深入的认识。

33.3.1 获取新闻组

为了从 NNTP 服务器获取新闻组的列表，用户必须首先建立一个到 NNTP 服务器的连接，进入阅读模式，然后请求“活跃”列表。

建立一个连接，输入：

```
unixbox% telnet news.cuckoo.com 119
```

用户应该看到下列输出：

```
Trying 10.0.0.1...
Connected to news.cuckoo.com.
Escape character is '^]'.
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
➡ ready (posting ok)
```

这个时候，键入：

```
MODE reader
```

将有下列输出：

```
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
➡ ready (posting ok)
```

当继续进行的状态栏出现后，可以键入命令显示所有的 Usenet组列表：

```
LIST active
```

这个时候，NNTP服务器将显示所拥有的所有 Usenet组的列表。

```
215 Newsgroups in form 'group high low flags'.
alt.test 0000000967 0000000947 y
comp.unix 0000000343 0000000343 y
comp.unix.admin 0000094331 0000094080 y
comp.unix.advocacy 0000068422 0000068413 y
comp.unix.aix 0000162768 0000162368 y
comp.unix.amiga 0000019141 0000019132 y
comp.unix.aux 0000023507 0000023502 y
comp.unix.bsd.bsdi.announce 0000000090 0000000091 m
comp.unix.bsd.bsdi.misc 0000009195 0000009192 y
comp.unix.bsd.freebsd 0000000201 0000000197 y
comp.unix.bsd.freebsd.announce 0000000955 0000000955 m
```

当结束操作后，键入：

```
QUIT
```

NNTP服务器将结束会话并有如下输出：

```
205 Bye!
Connection closed by foreign host.
unixbox%
```

注意 绝大多数的服务器拥有巨大数量的新闻组。在一个服务器的可用文件中包含3000 ~ 6000个新闻组一点也不奇怪。当用户发出LIST active命令，服务器可能会产生一个数兆字节大小的文件。所以，为了节省连接双方的带宽起见最好避免，访问该文件。

服务器指明了新闻组的列表格式是“组 高 低 标志(group high low flags)”；高和低是NNTP客户从NNTP服务器当前能看到的文件编号区域。标志则显示了对该新闻组的控制状态；其中“m”意思为中等的，而“y”的意思是严厉的。

提示 除了本章中提到的那些还有更多的NNTP命令。当连接到NNTP服务器上的时候，客户可以键入HELP来获取关于那些额外命令的帮助。绝大多数的客户都会得到一连串

的命令以及使用它们的必要参数。

33.3.2 获取消息

用户可以通过连接、选择组接着发出一个文章的标志号来从 NNTP服务器上获取 Usenet消息。在接收到文章以后，用户可以继续通过标识号ID来请求其他文章，或者发出下一条命令。

为建立连接，键入：

```
unixbox% telnet news.cuckoo.com 119
```

用户将看到下列显示：

```
Trying 10.0.0.1...
Connected to news.cuckoo.com.
Escape character is '^]'.
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
➤ ready (posting ok)
```

这时候键入：

```
MODE reader
```

下列结果将显示：

```
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
➤ ready (posting ok)
```

当继续进行的状态栏显示后，键入命令告诉新闻服务器用户想访问那一个新闻组：

```
GROUP alt.test
```

服务器将回应询问中所提到的组的一些信息：

```
211 7440 947697 955337 alt.test
```

在回应中的第二个数字是当前服务器包含的组中的文章数。第三和第四个数字是文章号码区域的开始号码和结束号码。在服务器所拥有的文章总数和文章号码区域之间可能存在着不一致，因为一些文章在收到后被服务器取消掉了。这些文章可能有一个文章号，但在服务器上已经不存在了。

这个时候，用户可以从给定区域内选择任何号码来访问文章，如下所示：

```
ARTICLE 955283
```

用户将看到如下显示：

```
220 955283 <07N%2.156$m6.184626@news.distributed.net> article
Path: news.cuckoo.com!uuneo.neosoft.com!ultraneo.neosoft.com!
➤ news.distributed.net!not-for-mail
From: root@distributed.net (Root)
Newsgroups: alt.test
Subject: Hello!
Message-ID: <07N%2.156$m6.184626@news.distributed.net>
Date: Sun, 16 May 1999 20:46:01 -0800
NNTP-Posting-Host: unixbox.distributed.net
NNTP-Posting-Date: Sun, 16 May 1999 21:42:36 PDT
Xref: news.cuckoo.com alt.test:955283
```

```
Hello, world! This is a usenet post in the alt.test newsgroup,
```

```
a newsgroup designated for testing of all sorts!
```

```
I'm glad this works!
```

```
-root
```

这个时候，用户可以继续进行NNTP会话，或键入QUIT命令终止会话。

正如上所示，为了接收文章数据，用户必须指定一个组及其文章号码。

显示中的Path(路径)标题区域显示了该消息到达用户的 NNTP客户服务器所经过的路径。路径从右到左指明了消息所经过的主机的名称或其别名，各名称之间由惊叹号隔开。在上一示例中，发布到news.distributed.net的文章，到达示例中的NNTP客户服务器news.cuckoo.com经历了中间的两步跨越。

NNTP-Posting标题区域指明那一个客户连接到NNTP客户服务器。这个示例说明该文章是由主机unixbox.distributed.net在1999年5月16日星期天21:42:36PDT时刻发布的。

用next命令也可以来请求服务器下一序号的文章，并显示它的标题及正文。

33.3.3 发布消息

用户可以通过与NNTP服务器建立连接、启动读者模式，接着发出一个POST命令来往Usenet上发布消息。一旦这些准备结束，文章就可以像通常情况一样地输入。仅为完成文章发布需要填写From、Newsgroups和Subject标题区域。然而，许多其他标题都是标准化的，仅由新闻客户端来使用。

```
unixbox% telnet news.cuckoo.com 119
```

```
Trying 10.0.0.1...
```

```
Connected to news.cuckoo.com.
```

```
Escape character is '^['.
```

```
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
```

```
A ready (posting ok)
```

```
MODE reader
```

```
200 news.cuckoo.com InterNetNews NNTP server INN 1.7.2 08-Dec-1997
```

```
➡ ready (posting ok)
```

现在，用户需要让NNTP服务器知道用户想进入文章发布模式：

```
POST
```

NNTP服务器将如下显示回应用户的请求：

```
340 Ok
```

这个时候，可以开始用户的Usenet文章发布了。在输入文章数据之前，用户必须先输入标题，至少包括From、Newsgroups和Subject几项。

```
From: Jennifer <jennifer@cuckoo.com>
```

```
Newsgroups: alt.test
```

```
Subject: Another test!
```

```
This is a test of posting messages to Usenet by connecting directly
```

```
➡ to a NNTP server!
```

```
-Jennifer
```

用户可以在一空白行上输入一个句点来结束用户所发布的文章：

```
.
```

如果用户能够顺利完成以上各步，NNTP服务器将确认用户发布成功：

```
240 Article posted
```

这时候用户可以退出会话。

```
QUIT
```

```
205 Bye!
```

```
Connection closed by foreign host.
```

```
unixbox%
```

如上所示，发布一个Usenet消息仅有少量的信息是必须的。

值得注意的是，发布中的最后那个句点是结束符。它必须单独占据一行并且位于行的开头，这样NNTP服务器才能意识到这个句点标志着文件的结束。

下一小节涉及到了目前在Usenet中的一些关键性问题。因为是沿袭Usenet的传统设计，对于整个基础结构没有官方的管理员；而由特定的层次管理员负责管理事宜，因此问题总有各种各样的不同解决方法。

33.4 大量广告(Spamming)和新闻黑洞(Blackholing)

大量广告(“Spamming”)被发布到Usenet上是一个严峻的问题。不守规矩的公司贪图免费传媒的好处将大量的广告发布到并不适合它们的组。这些不适宜的广告占据了大量的新闻服务器资源，使终端用户必须烦恼地过滤大量广告才能获得自己想要的内容。

一些新闻管理员试图通过新闻黑洞技术来解决这个问题，换句话说，就是忽略那些接收大量Usenet广告的、创建Usenet广告的、中继Usenet广告的站点。顾名思义，黑洞技术就是指借助标记为“黑洞”的站点减少网络拥塞。任何加强了黑洞策略的站点将忽略被标记为黑洞的站点。这就给那些有责任阻止大量广告的站点增加了压力，如果一个ISP的新闻服务器没有能成功地阻止一个用户滥用其权利，它将被标记为“黑洞”。

另一项解决大量广告问题的技术是取消不适宜的文章。一些被称为“cancelbots”的程序负责确认发布的商业性信息，对其进行监测，若存在有问题的文章，则发出控制命令请求Usenet服务器将其删除。

注意 MAPS RBL(<http://maps.vix.com/rbl>)是组织此类黑洞技术的最大的组织。

33.5 小结

在提供公众性的讨论论坛之外，多年来，Usenet也传播了大量的知识及信息，使所有的互联网用户群都可以方便地获取。例如Deja(<http://www.deja.com/>)网站保存了Usenet发布的所有信息，允许用户查询多年来的Usenet数据。Deja文档除了包括Usenet所涵盖的广泛话题外，同时也包括与用户能想象到的几乎所有话题相关的文章。近期这些功能的实现进一步加强了人们对于Usenet应用前景的信念。

Usenet是拥有巨大潜力、受人欢迎的服务。然而，Usenet若要取得成功、易于管理、便于

扩充，还需要解决许多重要的问题。但毫无疑问，这些问题终会以某种方式得到解决，并导致Usenet走向繁荣、发挥出它的最大潜力。

下一章涵盖了运行 Web 站点服务所涉及到的各种管理上的问题。特别是，用户将学习到有关 Apache Web 服务器、Netscape 企业级 Web 服务器、微软的互联网信息（Internet Information）服务器的基础知识。