

· 网络与通信 ·

电子新闻系统的服务机制

顾韵华 余晓 龚俭

(南京气象学院计算机系) (东南大学计算机系 南京 210096)

摘要 介绍电子新闻系统提供的新闻接收、存储及发布等基本新闻服务功能, 主要对基于 NNTP 的电子新闻服务系统 INN 的系统结构及其工作原理进行详细分析, 讨论了电子新闻系统中新闻服务的实现机制, 给出电子新闻服务系统建立和维护要点。

关键词 电子新闻 NNTP USENET 存储-转发

The Service Mechanism of Electronic News System

Gu Yunhua

(Department of Computer Science, Nanjing Institute of Meteorology)

Xu Xiao Gong Jian

(Department of Computer Science, Southeast University Nanjing 210096)

Abstract This paper gives out in detail the system architecture and operating principle of INN, an electronic news server based on NNTP, and discusses the implementation of the news service mechanism of an electronic news system. It also describes the main functions that are provided by an electronic news system, such as news receiving, saving and distributing. Still, it shows the main points of building and maintaining an electronic news server system.

Key words Electronic news; NNTP; USENET; Store and forward

随着计算机网络的日益普及, 人们越来越多地利用计算机网络公布消息、进行各种讨论等, 这种应用称为网络新闻系统。网络新闻系统中较为重要的一种是电子新闻系统, 其它的手段还包括电子论坛(listserv)、电子公告牌(bbs)和邮件列表(mailing-list)。电子新闻系统中采用的协议是NNTP(网络新闻传输协议), NNTP使用可靠的基于流传输(如TCP)客户机/服务器模式, 对新闻条款进行发布、查询、检索、邮寄、索引及失效等控制。*

目前, 在Internet上主要以USENET新闻系统, 按主题分类方式提供电子新闻服务, 用户可访问的新闻组达数千种。USENET是由许多个USENET网点构成的新闻发布系统, 运作机制为存储-转发。建立USENET网点的关键在于建立其新闻服务器, 一个完整的电子新闻服务器应具备完善的新闻接收、存储、管理、转发功能, 并提供用户阅读新闻与参加讨论的方法。本文以中国教育科研计算机网(CERNET)华东(北)地区网的电子新闻系统为背景讨

论新闻服务器的系统结构及运行机制。

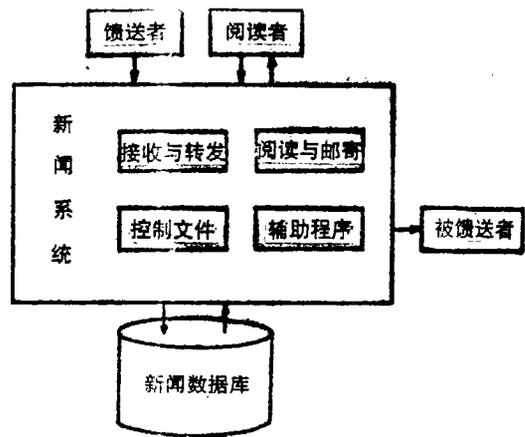


图1 系统模型

* 本文的研究内容受 CERNET 应用课题资助
顾韵华 讲师, 主要从事数据库、计算机网络的教研工作

收稿日期: 1996-04-01

1 系统模型

CERNET 华东(北)地区网电子新闻系统的建立采用了 Internet 中常用的 INN(InterNetNews)系统,使用它可建成一个 USENET 网点的服务器。INN 是一个基于 NNTP 协议的新闻服务系统,是在 UNIX 环境下主要以 C、Shell 程序设计语言开发的,可运行于 Internet 的常见平台 SunOS、Ultrix、Solaris 等。其提供的新闻服务可归结为:(1)在一台计算机上建立本地新闻数据库;(2)提供对新闻数据库的维护管理手段;(3)提供与其它 USENET 网点连接以互相交换新闻条款的功能;(4)支持用户的新闻阅读;(5)支持用户邮寄自己的新闻条款。

1.1 新闻数据库

新闻数据库是本 USENET 网点收到的新闻条款的集合。对新闻数据库最常用的操作是存储和检索,结合 USENET 的新闻组命名规则,将新闻数据库设计为一个包含若干棵新闻条款树的新闻条款林。

1.2 新闻接收与转发系统

这是新闻服务器的最主要部分,它负责:(1)侦听 NNTP 口,处理进来的连接;(2)为新闻条款生成文件存储之;(3)处理控制消息;(4)控制转发新闻条款拷贝及条款失效处理。

1.3 新闻阅读系统

负责处理用户阅读新闻及邮寄条款的请求。

1.4 运行控制工具

这是提供给新闻管理员对新闻系统的管理工具。主要可进行新闻组的增、删,以及对新闻数据库的活动索引文件的管理等。

2 系统工作原理

新闻接收转发系统和新闻阅读系统物理上以两个进程 innd 和 nnrpd 来实现, innd 是主服务器进程,完成新闻接收与转发工作; nnrpd 是处理新闻阅读和邮寄的服务器进程。INN 工作,实质是 innd 进程处于活动状态。一般 innd 作为后台进程运行,其主要功能及设计要点如下。

2.1 连接的处理

根据 NNTP 协议, INN 工作时使用 TCP 的端口 119(以下称为 NNTP 口)。进程 innd 运行时,侦听 NNTP 口,处理所有进来的连接。NNTP 连接分两类:馈送连接和客户连接。馈送连接是来自本网点的新闻馈送者的连接,客户连接则是来自本

USENET 网点用户的连接。

innd 对两类连接进行不同的处理。当一个有效的馈送连接进来,表明有馈送条款到达, innd 对发送来的新闻条款检查合法性,进行存储处理,并将它们送入排队以按控制要求转发拷贝到被本网点馈送的其它 USENET 网点去;当进来的是客户连接时, innd 进程将该连接交给新闻阅读服务进程 nnrpd 去处理,而 innd 进程致力于侦听及处理馈送连接,这样可提高系统整体效率。

2.2 到达新闻条款的处理

新闻条款到达新闻服务器主机有两个途径:一是来自本网点的新闻馈送者,一是来自本网点的用户。USENET 允许一则条款被发布到多个新闻组,称之为交叉邮寄条款。条款的结构主要参照 RFC1036 的规定。innd 主要依据新闻条款的格式以及新老程度判别新闻条款的合法性。新闻服务器对进来的合法的每个新闻条款都将其存入新闻数据库。具体地,对每个非交叉邮寄条款以一个文件存储;对交叉邮寄条款,无必要在每次收到后都以文件存储之,仅在第一次接收时创建一个文件存储它。其后则是建立一个到第一次所建文件的连接。

INN 支持 USENET 的仲裁新闻组,通过控制文件 moderates 确定仲裁新闻组的仲裁者,保证新闻条款的质量。

2.3 控制信息的处理

控制消息是 USENET 的新闻管理员用来管理 USENET 的动作的特殊新闻条款,有创建新闻组、删除新闻组及中止条款等。innd 负责按 USENET 新闻系统要求处理所有控制消息。

2.4 控制新闻条款失效处理

新闻数据库中老的条款要及时删除,才能保持新闻的 spool 区有足够空间存放新条款。innd 允许新闻管理员根据新闻组的重要程度和 USENET 主机的硬盘容量在控制文件 expire.ctl 中设定不同新闻组的条款有不同的失效时间。条款的失效处理是定期进行的,交叉邮寄的条款在最后一个连接删除后文件才被删除。

2.5 转发新闻的控制

在 NNTP 协议出现之前,消息发布可以邮件列表方式或 UUCP 方式进行。NNTP 保留了邮件列表方式,同时推荐了 NNTP 方式。NNTP 方式是通过 USENET 新闻系统,主机间相互作用决定新闻交换的一种新闻发布方式,它比其它两种方式优越,主要在于它能大大降低网络通信量。INN 支持上述三种

方法, 在其控制文件 newsfeeds 中可设定到不同的被馈送网点采用不同的发布方法。该控制文件由 innd 启动时读入内存, 负责解释, 调用相应的发布程序转发条款。这种转发机制使各 USENET 网点形成新闻网。

2.6 传输循环的阻止

USENET 中, 新闻传输算法的一个重要部分是阻止传输循环。传输循环是指新闻条款在由各个 USENET 网点组成的连通图反复地向每个网点发送。innd 阻止循环的方法是条款第一次到达本网点时, 在其 Path: 标题行加上自己的主机名, 表明本网点已收到该条款, innd 从不接收一个本网点主机已收到过的条款。

2.7 NNRPD 进程

nnrpd 被 innd 进程引用, innd 将一个客户连接移交给 nnrpd, nnrpd 确认其为一个有效连接后, 将发起该连接的客户主机设置为一个“阅读者”(reader), “阅读者”主机的用户可访问新闻服务器上的新闻数据库, 主要进行阅读(检索)和邮寄(插入)。nnrpd 将对用户命令的响应输出到用户的标准输出上, 其对用户邮寄条款命令的处理是将条款传送给 innd, 仅 innd 直接对新闻数据库操作。邮寄条款可被本服务器转发拷贝出去, 本地用户就参加入了新闻组的讨论。

3 电子新闻系统的建立

3.1 系统环境

CERNET 华东(北)地区网(以下简称地区网)的电子新闻服务器建在地区中心的 S1000e 计算机上, 该机内存 128MB, 硬盘 10GB, 操作系统 Solaris 2.4。选择该机作新闻服务器重在考虑新闻数据库占用大量外存及保证较快的响应速度。

3.2 新闻的访问控制

指对新闻数据库中数据的存取控制, 这是新闻服务的核心问题。首先, 新闻条款来源, 即何数据可存入新闻数据库。许多 USENET 网点可作为新闻馈送者, 只要按一定程序办理订阅手续, 即可获得订阅网点的新闻馈送。基于 CERNET 国家中心的新闻服务器已可提供上千个常见新闻组的新闻及安全因素, 目前地区网新闻服务器只从 CERNET 国家中心

USENET 网点订阅新闻, 有 1000 多个新闻组; 并建立了地区网的组, 发布中英文消息, 用户可以邮寄条款参加讨论, 也可引入仲裁机制; 其次, 条款的生存期, 其设置的恰当与否直接影响服务器主机的运行。考虑条款存在价值和硬盘容量, 我们将绝大多数新闻组的条款的生存时间定为 14-30 天(该数字可视具体计算机系统而设定), 控制消息由于其特殊作用, 设定的保存时间可更长些。最后, 访问权限问题, 可设置能够访问新闻系统的用户, 为限制邮寄条款数, 设定大部分用户只读新闻而不能邮寄条款。用户使用新闻系统的方法有 telnet 到端口、新闻阅读器和浏览器等。

3.3 新闻的本地化

新闻的新闻本地化主要考虑中文条款的处理问题。INN 接收的新闻条款可以是 7bit 或 8bit 数据, 利用这点可进行中文条款的邮寄、存储及转发。中文条款的邮寄除了需中文操作系统支持外, 与英文条款是相同的, 其存储和转发的处理也与英文条款相同。地区网新闻服务器现已建立了可发布中文条款的新闻组, 组名为 ECN, 发布地区中心的要闻。

4 结束语

本文对基于 NNTP 的电子新闻系统 INN 以“存储-转发”工作机制为用户提供新闻接收、存储、转发等基本新闻服务进行了较详细的分析, 正由于采用了非集中化的新闻服务方式, 使电子新闻这种网络应用为越来越多的用户所接收, 这对人们讨论、合作方式的转变产生了一定的影响。目前电子新闻服务系统存在的主要不足是缺乏像网络管理工具那样方便高效的访问控制手段, 新闻管理员难以确切了解新闻数据库状态。

参考文献

- 1 Brain K, Phil L. Network News Transfer Protocol, A Proposed Standard for the Stream-based Transmission of News. RFC: 977, 1986
- 2 Horton M, Adams R. Standard for Inter Change of USENET Messages RFC: 1036, 1987
- 3 Harley H, Rick S. Internet 参考大全(第1版). 北京: 学苑出版社, 1994