文章编号：1005－8893（2005）03－0045－04

一种 IP 控制网关的设计与实现*

姜熙炯[1]，封红旗[2]
（1. 苏州大学 计算机科学与技术学院，江苏 苏州 215006；2. 江苏工业学院 网络中心，江苏 常州 213016）

摘要：介绍了一种基于透明网桥原理，采用 Linux 内核开发的 IP 控制网关。重点分析了系统的结构、实现的关键技术及实现过程，并结合多出口的园区网拓扑结构特点，探讨了其在网络上的具体应用。

关键词：Linux；内核模块；网关
中图分类号：TP 393.5　　文献标识码：A

随着网络技术的发展，对于园区网用户而言，采用多条出口与不同的网络服务提供商进行接入，一方面可以提高网络的稳定性，避免由于单一链路的故障引起网络中断，另一方面，可以制定合理的分流策略，对进出的业务流量进行分流，均衡网络带宽，减少网络拥塞，提高网络的访问速度。在这种状况下，为便于管理这种多出口网络。用户在 Linux 下开发了许多基于 Netfilter 的透明网关系统，由于这种网关一般需要配置路由，会使网关系统效率降低[1]。由于以太网桥工作在数据链路层，是与协议无关的，它只能识别以太网帧（Frame）。对于工作在网络下的应用而言，感觉不到网桥的存在，可以认为它是物理上透明的。利用这个原理，实现了一个针对不同协议，不同端口进行控制的透明网关。

1　系统设计原理

在基于 TCP/IP 协议的网络中，当数据由应用层自上而下传递时，首先在网络层形成 IP 数据包，然后再向下到达数据链路层，由数据链路层将 IP 数据包分割为数据帧，加上以太网包头后向下发送到物理介质上。以太网包头中包含着本地主机和目标主机的 48 位 MAC 地址，链路层的数据帧就是依靠 MAC 地址来寻址的，正常情况下，网卡只响应以下两种数据帧：①数据帧的目标 MAC 地址与网卡自身的 MAC 地址一致；②数据帧的目标 MAC 地址为广播地址。

当接收到上面两种类型的数据帧时，网卡会通过 CPU 产生一个硬件中断，然后再由操作系统负责处理该中断，对帧中所包含的数据做进一步处理。但在 IP 控制网关中，要网卡接收所有的数据帧，所以需要将网卡设置为混杂（Promiscuous）模式，对接收到的每一个帧都产生一个硬件中断，以提醒操作系统处理经过该网卡的每一个数据包，这样网卡就可以捕获网络上所有的数据了[2]。因此，本系统采用两块设置成混杂模式的网卡构建一个网桥，并设计了相应的网桥内核模块，用以管理这两个以太网卡，网桥内核以 Linux 内核模块模式运行，负责处理流经两个网卡的数据，将一块网卡收到的包一律转发到另一块网卡上（见图 1），采用先进的流量采集技术，对网络流量进行管理，配合数据库系统的工作，实现流量采集与系统控制。

2　系统结构设计

2.1　系统体系结构设计

在本系统设计中，采用一台运行了 Linux 2.4.18 内核的服务器，在该服务器上配置了 3 个以太网卡，其中两个网卡设置成混杂模式，用于转发数据，构建透明网桥。另一个网卡配置 IP 地址，

* 收稿日期：2005－04－27
作者简介：姜熙炯（1972－），男，江苏丹阳人，硕士研究生。

管理员可以登录系统管理和配置网关参数，系统按功能分为 4 个子模块。

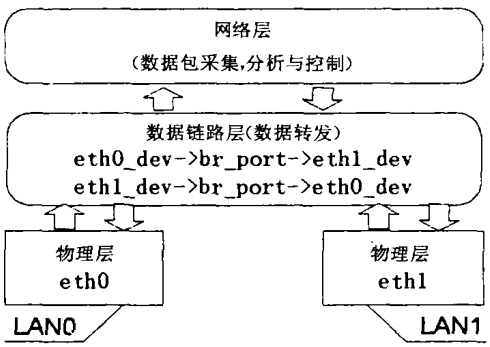

图 1 网桥模块工作原理

Fig. 1 Working principle of bridge module

网桥模块：利用网桥原理构建，处于数据链路层，负责截获所有的数据包，并通过检查每一个数据包的源 IP 地址及目的 IP 地址，并按系统指令参数对数据包进行相应的管理与控制，是整个 IP 控制网关的核心模块。

网桥控制模块：监听来自数据库的请求，负责激活内核模块以及内核模块与数据库模块的通讯，获取所有重要的数据包原始信息，如时间、源 IP、协议等，并写入数据库。

数据库模块：监听来自认证服务器的请求，与网关管理模块进行通信，将来自 web 客户端的请求送入内核，以及将来自内核的流量统计写入数据库，对用户的流量信息，用户权限等信息进行统一存储。

系统管理模块：系统管理模块包括：系统参数设定，管理员模块、操作员模块和用户模块等部分，负责数据库操作，以及少量的与数据库模块通信任务。

整个网关系统各功能模块逻辑关系见图 2。

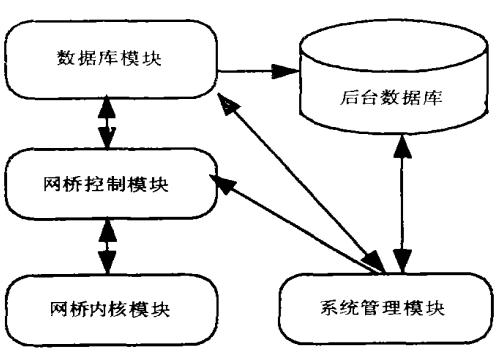

图 2 网关各模块间关系

Fig. 2 Relationships between the module structures

2.2 数据采集方法与技术

在 IP 控制网关的应用中，对数据流量的控制与管理是其它服务的基础工作。数据采集是否迅速有效是其中的关键，本系统采用存贮转发加数据统计的方式，网卡在混杂模式下接收网络上的所有数据，然后把数据包交给上层程序处理，由于网关一般架设在内外网链路的主干部位上，其数据采集是否高效是衡量系统功能的重要指标。传统上，在数据采集过程中，一般采用将数据包头全部拷贝到用户层空间，然后再进行数据分析，这样会消耗相当大的系统资源[3]。为了提高网关的数据采集效率，我们对操作系统的网络协议栈进行了修改，根据功能需要，开发了专用于数据采集和网络控制的网络层功能。系统只在底层提取源 IP 地址、目的 IP 地址、源端口、目的端口、协议、时间戳、数据包长度等信息，每个数据包只需拷贝 20 字节，远远小于拷贝整个数据包头的方案。同时采用了底层数据双缓冲技术，数据采集模块维护着两个缓冲，使系统不会因为等待缓冲区而丢失数据，同时也减少了数据丢失的可能性。

对于每个接收的数据包，同时根据一定的决策算法进行决策。然后根据结果处理数据包，在控制系统中一般只支持转发和丢弃两种动作，要么转发要么丢弃。对于多条规则就得去多次匹配，这就很大地增加了运算量。为了提高决策算法的效率，系统采用了分页式的决策算法，这种算法由底层提供决策，具体的策略由上层制定，这样系统在最坏情况下的计算量是一定的。随着规模的上升计算量不会显著上升，可以减少转发的平均时延，降低控制系统对网络性能的影响，提高网络的性能。

3 系统模块的设计与实现

3.1 透明网桥模块设计

网桥内核模块（Bridge）在整个系统中处于核心地位，主要管理两块网卡，一块网卡收到的包一律转发到另一块网卡界面上；网桥内核模块负责系统的转发控制、流量处理、带宽控制、ACL 访问管理等。对于系统制定的不同策略，可对数据采取不同的处理方式，主要处理动作有通过（PASS）、丢弃（DISCARD），在具体的转发包的过程中，通过相应的数据包处理函数来实现，其对于数据包处理流程见图 3。

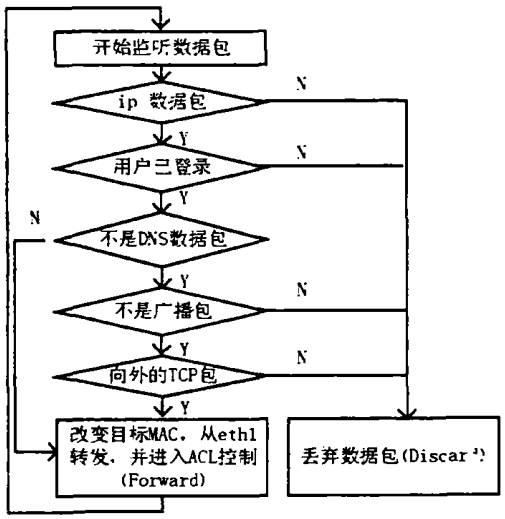

图 3 数据包处理流程

Fig. 3 Processing of network data

网络接口核心层通过 dev_queue_xmit（）函数向上层提供统一的发送接口[4]，无论是 IP，还是 ARP 协议，通过这个函数可把要发送的数据进行传递。dev_queue_xmit（）做的工作最后会落实到 dev→hard_start_xmit（），而 dev→hard_start_xmit（）会调用实际的驱动程序来完成发送的任务。其中数据流入和流出可以共享同一段代码，唯一的区别就是它们的源/目的地址需要交换。网桥的管理由网关守护进程负责，网桥模块使用 insmod 或 modprobe 插入系统内核时，可以设定系统内存大小、网桥锁定时间、用户重定向的 IP 地址等参数。

3.2 网关守护进程模块设计

网关守护进程模块在本系统中处于呈上启下的作用，一方面监听来自数据库服务器的同步请求，另一方面，通过 ioctl 将用户请求传入内核的网桥模块。当网关守护进程启动之后，首先读入所有可管理的 IP 地址段，并将其传送给网桥模块，然后连接数据库守护进程，读入所有已登录 IP，并添加到 bridge 内核模块中，这样就实现了当网关守护进程意外退出之后的错误恢复。经过以上 2 步，就完成了网关守护进程的启动，这时，网关守护进程之后进入监听状态，接收数据库服务器命令，完成在线 IP 管理功能，并接收数据库服务器传送来的新的在线 IP，添加到内核中等功能。

3.3 数据库守护进程模块设计

数据库模块一方面监听来自认证服务器的请求，处理用户的登录和退出请求。同时负责与网桥守护进程定时进行同步，将用户的网络流量记录入数据库，系统在转发冲报文的同时将流量数据写入数据库系统。为了让管理员能实时地监控网络，系统采取直接写入数据库的方式。同时为了减少对报文的处理时间，在系统缓存中建立一个流量表，将单位时间内同一连接的流量数据累加，然后定时将流量写入数据库系统，采用这种方式，减少了流量表中的记录数量，提高了系统流量统计的效率。数据库模块还可以接受来自认证服务器的请求，读入 IP 地址、用户名和登录模式，并将此信息送给网关守护进程，实现用户登录与退出。为获得在线用户流量信息，数据库模块定时向网关守护进程发送 IP 刷新命令，并根据用户策略统计数据，然后记录入数据库。

3.4 系统管理模块设计

系统管理模块提供两种方式，一种方式是调试方式，提供给系统管理员对网桥系统进行底层调试工作，以方便准确快速发现系统出错原因，并在系统安装过程中用以监控网桥运行状态。另一种方式提供给用户管理员，采用 B/S 结构，以方便管理员对系统进行各项性能的日常管理，主要包括的管理模块见图 4。

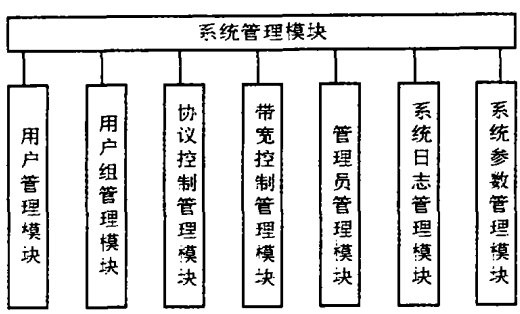

图 4 系统管理模块结构

Fig. 4 Structure of system management module

4 系统应用

目前，在园区网的建设中，随着 ISP 接入提供商的增多，对用户的管理及网络接入都有了较大的选择面。而对于网络管理而言，又需要采用统一的认证管理，采用这种 IP 透明的控制网关，有利于对于多出口网络进行统一管理，而且本网关可以认为是一个网桥设备，它对高层协议是透明的，能转发任何网络层协议的数据流，用其互连起来的网络是 1 个单个的逻辑网。因此，在多出口的网络系统

中，可以将其接入到网络任意设定的线路上，几乎不需要对原网络拓扑做任何的改动。IP 控制网关可放置在中心交换机的外部、边界路器的内部，由边界路由器负责策略路由的设定，既可以设定基于源地址的策略路由，使本网络中指定网络访问特定的 ISP，或设定基于目的地址的策略路由，对访问特定地址的采用特定的路由线路。同理，如果用户目前没有边界路由器，可采用核心的 3 层交换机（如 CISCO Catalyst 6509），由其制定策略路由，将 IP 控制网关串接在网络上，同样可以起到网络管理作用（见图 5）。

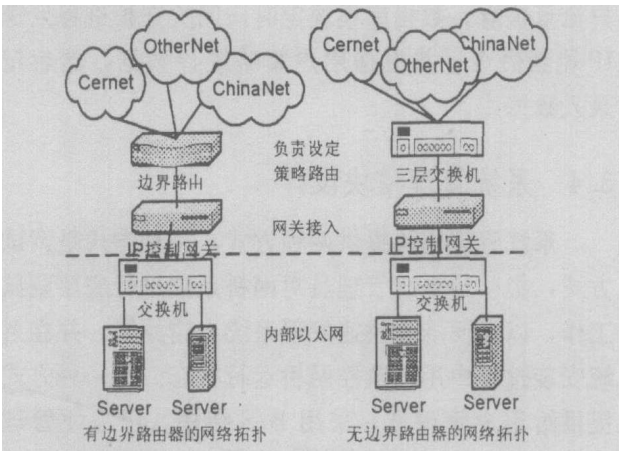

图 5 网关系统连接拓扑图

Fig. 5 Network topology with gateway

5 结束语

本文所论述的 IP 控制网关系统，由于其是从系统底层开发，独立运行，使系统具有较好的性能和强大的功能，同时又具备了较好的安全性，在实践中得到了初步应用。由于其具有良好的捕捉分组和控制分组转发的能力，可以进行流量监视、流量分析等，在此基础上可开发具有包过过滤功能的防火墙系统。由于目前网络流量的不断增大，对于网络服务质量（QoS）的需求也越来越高，下一步的工作就是在优化透明网关的的基础上，开展服务质量方面的研究。

参考文献：

[1] 姚晓宇，赵晨. Linux 内核防火墙 Netfilter 实现与应用研究[J]. 计算机工程，2003，29（5）：112－113.

[2] 郑卫斌，丁会宁. Linux 的网络转发性能研究[J]. 西安交通大学学报，2004，38（2）：124－127.

[3] Richard Steven W. TCP/IP 详解[M]. 卷 1：协议. 范建华，胥光辉，张涛，等，译. 北京：机械工业出版社，2000.

[4] 毛德操，胡希明. Linux 内核源代码情景分析[M]. 杭州：浙江大学出版社，2001.

Designation and Realization of One IP－Controlling Gateway

JIANG Xi－jiong[1]，FENG Hong－qi[2]

（1. Computer Science and Technology School，Suzhou University，Suzhou 215006，China；2. Network Center，Jiangsu Polytechnic University，Changzhou 213001，China）

Abstract：The article introduced one IP_controlling gateway basing on the Linux kernel，analysed the structure of this gateway and the key technology of realization，and put forward the way to realize this idea. And finally linking with the structural characteristics of the multiway campus network，this article also discussed the application of this gateway to the campus network.

Key words：Linux；kernel module；gateway