文章编号:1671-6833(2006)03-0081-04

Linux平台下电信级计费网关研究

朱思峰[1,2], 李惠敏[3]

(1.周口师范学院数学与信息科学系,河南 周口 466000; 2.西北工业大学计算机学院,陕西 西安 710065; 3.郑州科技职业学院计算机系,河南 郑州 450064)

摘 要:设计了基于Linux内核的电信级计费网关模型,实现了一种在x86硬件上使用Linux内核进行包过滤的计费系统原型.论文的主要研究工作:使用Netfilter框架,完成了网络层的数据包重组、以及应用层的报文头分析;实现了在内核中对HTTP协议的数据流进行解析及过滤,并提取出计费所需的重要信息;使用x86电信服务器对系统原型进行了实例测试,通过性能分析,证明了该系统的可用性.

关键词: Linux; Netfilter框架; Linux内核; 计费网关

中图分类号: TP 391 **文献标识码**: A

0 引言

在我国电信增值业务迅猛发展的大背景下,无论是移动运营商还是服务提供商,都需要一个灵活高效的性能和价格均能被运营商和服务提供商所接受的电信级计费网关产品,使用灵活可变的资费规则,来核算移动终端上网所产生的费用.目前移动运营商的管理软件,一方面价格过高并且需要购买软件开发商提供的专用硬件(如Cisco公司的Cisco Mobile Exchange[1]、华为公司的infox-WISG[2]),另外在二次开发上的灵活性较差,难以满足移动业务资费规则不断变化的需求.

为了向中小规模的移动运营商和服务提供商提供一个高性价比、高灵活性的计费网关产品,笔者主要研究了运行在X86服务器上、Linux操作系统平台的电信级网关计费信息的获取技术.

1 Netfilter框架

1.1 Netfilter框架简介

Linux内核中的网络包过滤机制经历了Ipfw(1.1.x)、Ipfwadm(2.0.x)、Ipchains(2.2.x)、Netfilter(2.3.15-至今).Netfilter是一种内核中用于扩展各种网络服务的结构化底层框架.Netfilter的设计思想是:生成一个模块结构使之能够比较容易地扩展,新的特性加入到内核中并不需要重新启动内核.这样,可以通过简单的构造一个内核模块来实现网络新特性的扩展,给底层的网络特性扩展带来极大的便利,使新的网络特性能在Linux内核中更容易被实现[3].

Netfilter作为最新版本的Linux网络包过滤机制,比以前有了很大的改进.Netfilter提供了一个抽象、通用化的框架,该框架定义的一个子功能的实现就是对各个层次的网络协议包进行过滤.Netfilter框架包含以下三部分:

(1) 为每种网络协议(IPv4、IPv6等)定义一套钩子(hook)函数(IPv4定义了5个钩子函数),这些钩子函数在数据报流过协议栈的几个关键点被调用.在这几个点中,协议栈将把数据报及钩子函数标号作为参数调用Netfilter框架.

(2) 内核的任何模块可以对每种协议的一个或多个钩子进行注册,实现挂接,这样当某个数据包被传递给Netfilter框架时,内核能检测是否有任何模块对该协议和钩子函数进行了注册.若注册了,则调用该模块的注册时使用的回调函数,这样这些模块就有机会检查该数据包、丢弃该数据包及指示Netfilter将该数据包传入用户空间的队列.

(3) 需要被缓冲接收的数据包可以传递给用户空间进行异步处理.一个用户进程能检查数据包,修改数据包,甚至可以重新将修改过的数据包通过注入内核里的某个钩子函数,重新进入网络

收稿日期:2006-04-14;**修订日期**:2006-05-18

基金项目:陕西省自然科学基金资助项目(200511061127)

作者简介:朱思峰(1976-),男,河南周口人,周口师范学院讲师,西北工业大学在读博士研究生,主要从事网络安全和多媒体网络传输技术方面的研究.

的协议栈.

所有的包过滤和NAT等技术都基于该框架.内核网络代码中不再有到处都是的、混乱的修改数据包的代码.当前Netfilter构架在IPv4、IPv6等网络协议栈中都已经被实现.针对现有移动网络的实际应用,以及计费网关需要处理的网络包内容,本文对IPv4在Netfilter中的应用进行介绍.

1.2 IPv4数据包在Netfilter框架中的流程

IPv4数据包在Netfilter框架中的流程如图1所示,图中小圈所在的位置即为Netfilter中的几个钩子函数.数据包从左边进入,首先就到了钩子函数NF_IP_PRE_ROUTING,然后经过路由选择,有三种可能:一种是要从另一个网络接口出去的;第二种是到本机的;第三种是没有路由而被丢弃的.如果是到本机的,那么进入钩子函数NFIP_LOCAL_IN;如果是要从另一个网络接口出去的,那么数据包就进入钩子函数NF_IP_POST_ROUTING,然后数据包再进入钩子函数NF_IP_POST_ROUTING,然后发送出去.对本机产生的数据包,则会经过钩子函数NF_IP_LOCAL_OUT,然后再发送出本机.

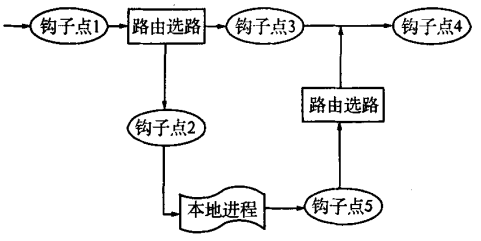

图1 IPv4网络数据包经过Netfilter框架

Fig.1 Ipv4 data packet passing the Netfilter

由于内核模块可以在任何一个钩子上注册和侦听,因此模块的注册必须要定义其注册的函数的优先级别.当任何一个钩子被核心网络代码调用时,各个模块在其上注册的函数就会按优先级的高低被调用.这样就可以对包进行各种各样的处理了.模块可以要求Netfiler作以下的处理:

(1) NF_ACCEPT: 继续传送.

(2) NF_DROP: 丢弃该包,不再继续传送.

(3) NF_STOLEN: 接管该包,不继续传送.

(4) NF_QUEUE: 将包放到队列中,一般是转发到用户空间.

(5) NF_REPEAT: 再次调用钩子.

那些被转发到用户空间的包,它们排队的处理是由内核模块ip_queque.ko来执行的,然后用户进程就可以对数据报进行任何处理.处理结束以后,用户进程可以调用nf_reinject()将该数据报重新注入内核,或者设置一个对数据报的目标动作.Netfilter的这个技术可使用户进程进行复杂的数据报操作,从而减轻了内核空间复杂度.

2 在Linux内核中实现对应用层协议的分析

计费网关的设计目标之一就是实现在Linux内核中实现对应用层协议的分析[4].具体地,系统将实现对HTTP通信的计费.HTTP作为WAP应用系统中重要的应用协议之一,在WAP网关到服务提供商之间的网络中占用着最多的带宽.而Linux内核中的网络部分并不提供对应用层协议的直接支持,应用Netfilter框架,可以在网络层接管通信数据,再对其中的应用层协议进行分析[5].

在计费网关原型中,为了在内核中完成对通过计费网关的应用层协议数据包的统计和计费,本文使用Linux内核中的Netfilter框架,对经过内核网络堆栈时对这些数据包进行截取与分析.与此同时,由于Linux内核并不提供对HTTP协议的支持,因此还必须在Netfilter的相应hook点中实现对应用层的分析.

注册一个Netfilter hook点需要调用nf_register_hook()函数,以及用到一个nf_hook_ops数据结构.nf_register_hook()以一个nf_hook_ops数据结构的地址作为参数并且返回一个整型的值.nf_hookfn函数的第一个参数用于指定hook函数类型,即hook点的位置.第二个参数是一个指针,该指针指向的指针指向一个sk_buff数据结构,网络堆栈用sk_buff数据结构来描述数据包.这上数据结构在skbuff.h中定义.sk_buff数据结构中最有用的部分是三个联合体,它们分别描述传输层包头、网络层包头,以及链路层包头,其名字依次为h、nh以及mac.通过访问这些联合体,就可以获得IP头和TCP头结构的指针,以及TCP协议中承载的HTTP数据.

根据层次化网络的定义,要在内核中获得HTTP协议的通讯内容,必须解决网络层、传输层和应用层中的协议接收和分析,相应地,在设计中有3个难点:

(1)对IP报文分片的重组:乱序到达的IP报文在到达内核网络堆栈后,必须缓冲接收并以原来的次序重新组合起来.

(2)高效接收TCP数据报:计费网关系统本

身并不关心 TCP 协议，但由于 HTTP 协议是承载于 TCP 协议之上的，因此系统也需要具备分析 TCP 协议流的功能．电信计费不能允许过高的延时，为了使 TCP 数据流能在内核的 Netfilter 模块中得到快速处理，则必须设计一个高效接收 TCP 数据报的算法．

(3)从 HTTP 通讯中获取付费内容的信息：HTTP 报文中可能携带着需要用户付费获取的信息，因此在内核中还需要分析 HTTP 报文，取得付费内容的信息．

3　IP 包的重组

Netfilter 的 5 个钩子函数都位于 IP 层，因此内核里数据流的截取点位于 IP 层．IP 协议的功能是路由选路，因此 IP 数据包的到达次序并不能确定．IP 数据包的捕获和分析的实现并不难，但为了分析应用层协议，系统必须对乱序到达的 IP 数据包进行重新组合．

参考 Peter M．Ewert 等人提出的使用的高效 IP 包重组算法[6]和 Linux 内核源代码，在设计计费网关时，负责 IP 包重组的主要数据结构是 ipfrag．每一个分片用 ipfrag 结构表示，未组装完的所有 IP 包用 iqp 结构表示，如图 2 所示．

算法的处理流程如图 3 所示，分 3 个步骤：

(1) 从 ip _ local _ deliver 把所有 IP 数据报的分片传递给 ip _ defrag()．在 ip _ defrag 中，这些分片在先被放入一个分片缓存结构中，直到所有的分片到达系统，进而取得完整的数据报提交给传输控制协议．

必须考虑的异常情况是，如果分片在最大的等待时间(ipfrag-time，30 s)里没有完全到达，则此块 IP 数据报将因超时而被丢弃．

(2) 使用一个存放 ipq 结构的哈希表作为分片缓存结构．每个 ipq 结构代表一个被分片的 IP 数据报，而单独的数据报分片被存入链表，每个分片所处的位置与它们被分片前所处的位置对应．

当没有足够的缓冲空间来存放这些 IP 数据报的分片时，将调用 ip _ evictor()来消除 ipq 条目，直到到达最低临界阀值 sysctl _ ipfrqg _ low _ thresh．

(3) ip _ frag-queue 把新的分片添加到现有的分片中，当分片全部到达后，即 pq - en = = pq - > mea 时，重组 IP 数据报．

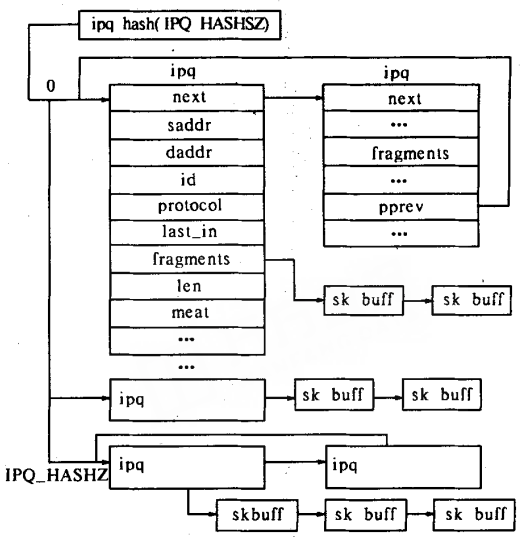

图 2　ipq 结构

Fig.2　The structure of ipq

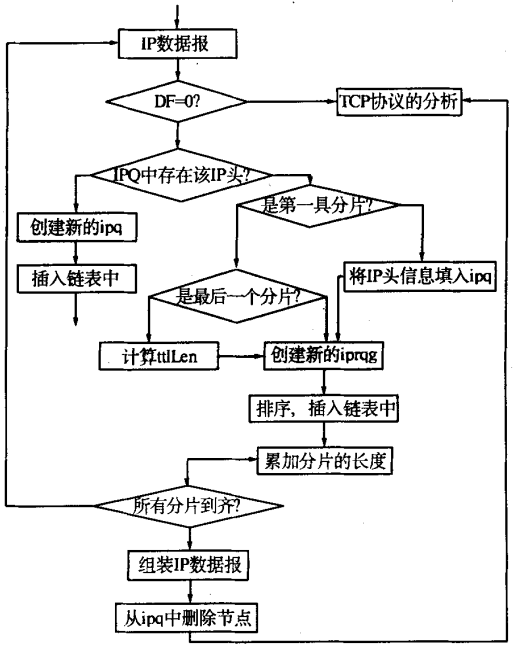

图 3　IP 包重组算法

Fig.3　The arithmetic of IP packet recombine

4　在 Linux 内核中处理 Http 协议

本计费网关所关心计费内容是由 WWW 服务器提供的，而 WWW 服务器使用的主要协议是 HTTP 协议．HTTP 协议是基于请求/响应模式的．一个客户机与服务器建立连接后，发送一个请求给服务器，请求方式的格式为，统一资源标识符、协议版本号，后边是 MIME 信息包括请求修饰符、客户机信息和可能的内容．服务器接到请求，给予

相应的响应信息,格式为一个状态行包括信息的协议版本号、一个成功或错误的代码,后边是MIME信息包括服务器信息、实体信息和可能的内容.

基于客户/服务器模式的信息交换过程,分四个过程,建立连接、发送请求信息、发送响应信息、关闭连接.在WWW服务中,"客户"与"服务器"是一个相对的概念,只存在于一个特定的连接期间,即在某个连接中的客户在另一个连接中可能作为服务器.WWW服务器运行时,一直在TCP80端口监听,等待连接的出现.在计费网关监控的网络中,HTTP协议下客户/服务器模式中信息交换的实现的4个过程所完成的工作是:

(1) 建立连接.连接的建立是通过申请套接字(Socket)实现的.客户打开一个套接字并把它约束在一个端口上,如果成功,就相当于建立了一个虚拟文件.以后就可以在该虚拟文件上写数据并通过网络向外传送.

(2) 发送请求.打开一个连接,客户机把请求消息送到服务器停留端口上,完成提出请求动作.

(3) 发送响应.服务器在处理完客户的请求之后,要向客户机发送响应消息.

(4) 关闭连接.客户和服务器双方都可以通过关闭套接字来结束TCP/IP对话.

根据HTTP通讯的特点,计费网关在接收TCP数据流之后重组HTTP报文,可以获得通讯双方的源地址/端口与目的地址/端口和HTTP报文中的MIME文件信息.

根据HTTP协议的无连接、遵循请求/响应模型、使用MIME内容等特点,计费网关就可以在HTTP会话中提取付费内容的文件信息[7].当HTTP服务器响应请求时,HTTP使用类MIME报文格式来封装数据.本文规定一个HTTP响应只能包含一个数据块,并使用"multipart/mixed"的标准MIME类型.

从服务器返回的HTTP响应的都将带有http消息的头部,计费网关通过对表单信息进行字符串匹配,就可以很容易地得到文件名、文件类型以及文件的大小.文件的内容被编码为二进制数据,其中字符"ThisRadomFile"为边界(Boundary).具体算法是:①获取HTTP响应报文头部中的二进制表单信息;②取边界(Boundary)信息;③取表单域名部分,转换成普通字符串;④提取Content-Length及Content-Type,转换成普通字符串;⑤提取Name及Filename,转换成普通字符串;⑥判断Data部分是否为空:是则执行8),否则执行7);⑦记录付费内容的信息;⑧操作结束.

5 结论

针对当前电信级计费网关使用专有设备和专有软件系统架设成本高、规则添加不够灵活以及提高工作效率的成本高的境况,本文设计并实现了一个基于Linux内核的电信级计费网关原型系统,该原型系统主要对WAP应用系统中常见的HTTP协议通讯进行计费处理.

为了测试计费网关原型的运行性能,使用Intel AdvancedTCA(先进电信计算架构)硬件平台.计费网关原型在此硬件平台上进行测试,充分满足了电信级的测试强度,从而保证了系统测试实验数据的有用性.要模拟电信级的通信量,测试实例使用Apache项目组提供的开源HTTP协议的性能测试工具ab(Apache Benchmark).为了对Linux的虚拟内存、中断请求、上下文切换、CPU使用率进行检查,测试使用了Unix/Linux下的性能分析工具vmstat.

测试由两组数据对比产生:一组是没有部署计费程序的路由器(纯转发路由器)的系统状况;另一组是部署计费程序后的路由器(计费网关)的系统状况.在并发连接数为250(略大于电信级需求)时,纯转发路由器和计费网关原型的系统状况,它们的CPU使用率、中断请求数、上下文切换数以及网络带宽使用情况如表1所示.

表1 性能达到峰值时纯转发路由器和计费网关原型的系统状况

Tab.1 The system status of pure router and accounting router in high-point of performance

类型	系统占用CPU时间/%	中断请求数	上下文切换数	网络带宽占用 *KByte/s*
纯转发路由	7	8 805	3 594	82 137
计费网关原型	98	4 860	2 973	47 282

由表1可以看出,在第一组测试中,通过纯转发路由器的网络带宽没有达到理想的千兆网关的极限,纯转发路由的CPU不处于满荷的工作状态,网络性能的瓶颈不在于纯转发路由,而是在HTTP服务器.而对于第二组测试,计费网关原型

(下转第88页)

Windows XP SP2 Firewall Technology and Its Application

Zhang Yu - feng, Zhai Guang - qun

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: This paper provides the operation method of Windows XP SP2 firewall and application embodiment. It studies Windows XP SP2 firewall (ICF) technology, through inserting, installing, setting up personal firewall and ICF respectively under different networkings. The result proves: The firewall of Windows XP SP2 almost has all the merits of other kinds of personal firewalls, and it takes up less systematic resources, is less likely to be attacked, and has better network security.

Key words: Windows XP SP2 firewall; Network security; The attack guards against

(上接第84页)

的CPU系统时间已经到达98%,这表明CPU已经满荷工作,此时通过计费网关原型的网络带宽和它所处理的响应数应能反映计费网关原型的实际处理性能.测试结果表明:以每秒处理3 300个HTTP请求/响应对,带宽达到47MByte/s的处理能力,基于x86硬件和Linux的计费网关是完全可以满足一个大中型服务提供商的业务需求的.

参考文献:

[1] 张 宏.移动业务管理平台CMX[EB/OL].http://www.cisco.com/CN/network_telecom/2005_06_18.shtml, 2005.6.18.

[2] 李 涛. 无线综合业务网关[EB/OL]. http://www.huawei.com/ProductView_06_02_03.shtml,2006.2.3.

[3] 张 海,李彭军,李 宸. 基于Netfilter框架的计费网关[J]. 计算机应用,2002, 58(12):23~26.

[4] 林子惠. Linux平台下电信级计费网关的研究与实现[D]. 西安:西北工业大学, 2005. 35~48.

[5] 郑 芸. 基于Linux平台下的Email监控系统[J]. 西安交通大学学报, 2002,67(3)45~48.

[6] 杨润华. 高性能IP宽带计费网关的设计与实现[J]. 计算机应用研究,2003, 54(5):23~26.

[7] 尹远洪,张建中. 计费网关中的规则处理模块的设计和实现[J]. 福建电脑, 2004,26(6):89~92.

Research on the Carrier Grade Accounting Router Using Linux

ZHU Si - feng[1,2], LI Hui - min[3]

(1. Department of Mathematical and Information Science, Zhoukou Normal University, Zhoukou 466000, China; 2. Department of Computer Science, Northwestern Polytechnical University, Xi'an 710065, China; 3. Department of Computer, Zhengzhou Science and Technology College, Zhengzhou 450064, China)

Abstract: This paper researches into the actuality the carrier grade accounting router using linux, designs a model of accounting router based on linux kernel, and implements a prototype of accounting router based on linux kernel running in the x86 hardware. The main research contents are as follows: (1) Using the Netfilter framework, dada pack recombine in network level and the application protocol analysis were implemented. (2) Kernel module use Netlink socket to communicate with the user application while the Netfilter to do the application protocol analysis and filtering. (3) Finally, it tested the prototype of the accounting router system with carrier grade x86 server, and proved the availability with the performance data.

Key words: Linux; Netfilter; Linux kernel; accounting router