文章编号：1007-757X(2008)7-0050-02

利用 Netfilter 框架和 TC 实现 P2P 流量控制

张卿　陆晓峰

摘　要：介绍了 P2P 应用及其优缺点，对 Linux 内核的 Netfilter/Iptables 技术和带宽控制技术进行了分析，分析 P2P 协议并提出一套完善的识别方法。

关键词：P2P；Netfilter；带宽控制；Iptables

中图分类号：TP393　　**文献标识码**：A

引言

P2P 是 peer-to-peer 的缩写，称之为对等联网。IBM 对 P2P 的定义是：P2P 系统由若干互联协作的计算机构成，且至少具有如下特征之一：系统依存于边缘化(非中央式服务器)设备的主动协作，每个成员直接从其他成员而非服务器的参与中受益；系统中成员同时扮演服务器与客户端的角色；系统应用的用户能够意识到彼此的存在，构成一个虚拟或实际的群体[1]。

简单地说，P2P 直接将人们通过互联网直接交互。由于 P2P 应用不依赖于某个具体的服务器，传统的网络管理方法显得力不从心，因此对 P2P 应用的监控与管理变得尤为重要。

1　Netfilter 框架与 Iptables 技术

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包过滤子系统[2]。netfilter 为每种网络协议(IPv4、IPv6 等)定义一套钩子函数（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。一个数据包通过 netfilter 框架时，它将经过如图 1 所示流程。

数据包选择工具Iptables也称为用户空间，是具有可扩展性的规则配置工具。Netfilter框架提供了一系列的表，每个表由若干个链组成，而每条链中可以由一条或多条规则组成。那么Netfilter是表的容器，表是链的容器，而链又是规则的容器，Iptables则是对这些规则进行操作的工具。

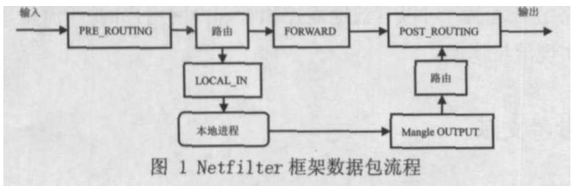

图 1 Netfilter 框架数据包流程

2　Linux 内核的带宽控制技术

Linux从Kernel 2.1.105就开始支持QoS，即Linux内核本身已经提供了相当完整和强大的带宽管理代码[3]。Linux主要使用路由工具包(IPROUTE2)的流量控制命令(Traffic Control command)来进行带宽控制。

流量控制模块在概念上可以划分为3部分：队列策略(Queue Discipline)、分类(classes)和过滤(filters)。其中每一个网络设备(如以太网卡eth0)都与一个队列策略相关，队列策略控制数据报在这个网络设备上怎样来排队，例如，sch_fifo是一个最简单的队列策略，它只有一个队列，所有分组都按其到达的顺序来排列，更加复杂的队列策略可能会使用过滤来区分不同类别的数据报，并对其分别处理。

数据报通常按下面的原则来处理：当一个队列策略的排队函数被调用时，它将在系统中寻找一个与自己相匹配的过滤，然后用这个过滤将数据报分成对应的分类。而那些没有相应过滤的数据报，就被系统划分到默认的分类中去。

3　P2P协议的识别

要实现对P2P应用的管理，必须先识别哪些是P2P数据流，哪些不是。目前很多的P2P应用软件都是开放源代码的，我们可以很方便地知道它们所使用的协议，而另外一些则可以通过抓取数据包来分析它们的协议。表1列举了部分P2P应用程序及其特征码。

P2P下载工具凭借其速度上的优势，已经完全替代了原先的Ftp下载工具，迅雷，网际快车等传统下载工具也开始加入P2P功能。但网络上各种各样的P2P下载工具基本上都是参照BitTorrent等开放源码的P2P工具，在协议上大同小异，此处不在赘述。但也有些软件比如迅雷，它的数据经过加密处理，不容易简单识别。

表1 部分P2P应用程序及其特征码

应用程序	特征码	备注
eDonkey	0xe3 0x9a/0x96	第1，2个字节
eMule	0xc5 0x91/0x92/0x93	第1，2个字节
Kad	0xe4 0x50/0x59	第1，2个字节
Gnutella	GNUTELLA 或 GND	明文检查
KaZaA	KaZaA	明文检查
SoulSeek	Xx xx 00 00 yy zz 00 00	前8个字节格式，其中xx xx 为16位负载－4，yy!=0,zz 任意或数据长度为8字节。
BitTorrent	0x13 BitTorrent protocol	第1个字节加明文检查
KAMUM	KamunPeers protocol	明文检查
百度下吧	BaiduP2P	明文检查
Ares Galaxy	PUSH SHA1:	明文检查

作者简介：张　卿，苏州大学计算机科学与技术学院，工程硕士，江苏 苏州 215006
陆晓峰，苏州大学计算机科学与技术学院，副教授，江苏 苏州 215006

4 利用Netfilter框架和TC实现P2P流量控制

4.1 在Netfilter中构造新的匹配函数

在linux2.6内核中，利用Netfilter框架的Iptables技术必须先向内核注册自己定义的处理函数：int ipt_register_match(struct ipt_match *match);注销时使用函数Void ipt_unregister_match(struct ipt_match *match)。这两个函数的核心是struct ipt_match结构，它有如下成员：

(1) struct list_head list;
初始化为{NULL,NULL}，由核心使用

(2) const char name[IPT_FUNCTION_MAXNAMELEN-1];
Match的名字，该名称必须和模块的名称相匹配。

(3) u_int8_t revision;

(4) int (*match)();
指向具体匹配函数的指针，返回非0表示匹配。

(5) int(*check entry)();
在使用本Match的规则注入表中之前调用，进行有效性检查，如果返回0，规则不会加入到iptables中。

(6) iptables void(*destroy)();
在包含本Match的规则从表中删除时调用，于checkentry配合可用于动态内存分配和释放。

(7) struct module *me;
表示当前Match是否为模块(NULL为否)。

完成注册之后，主要的工作便是在match函数中过滤网络数据包并判断是否为P2P数据包。部分关键代码如下：

```
static int match(const struct sk_buff *skb,const struct net_device *in,const struct net_device *out,const void *matchinfo,int offset, unsigned int protoff,int *hotdrop)
{
    switch (ip->protocol)
    {
case IPPROTO_TCP:
{
p2p_result=p2p_parse_tcp(skb);//检查tcp类型数据包
    return p2p_result;
}
case IPPROTO_UDP:
{
p2p_result=p2p_parse_udp(skb);//检查udp类型数据包
    return p2p_result;
}
default: return 0;
    }
}
```

4.2 为iptables添加命令行匹配

用户使用Iptables命令行实现对内核netfilter的具体操作，为了添加新的命令行匹配选项我们需要调用函数register_match来进行注册，它的入口参数为结构iptables_match，包含以下主要成员：

(1) struct iptables_match *next;
Match链，初始为NULL 。

(2) ipt_chainlabel name;
Match名，和核心模块加载类似 。

(3) const char *version;
版本信息。

(4) size_t size;
Match数据的大小，必须用IPT_ALIGN()宏指定对界。

(5) size_t userspacesize;
由于内核可能修改某些域，因此size可能与确切的用户数据不同，这时就应该把不会被改变的数据放在数据区的前面部分，而这里就应该填写被改变的数据区大小；一般来说，这个值和size相同。

(6) void (*help)();
当iptables要求显示当前match的信息时，就会调用这个函数。

(7) void (*init)();
初始化，在parse之前调用。

(8) int (*parse)();
扫描并接收本match的命令行参数，正确接收时返回非0。

(9) void (*final_check)();
当命令行参数全部处理完毕以后调用，如果不正确，应该退出。

(10) void (*print)();
当查询当前表中的规则时，显示使用了当前match的规则的额外的信息。

(11) void (*save)();
按照parse允许的格式将本match的命令行参数输出到标准输出。

Iptables将命令行输入转换为程序可读的格式，然后在调用libiptc库提供的iptc_commit()函数向内核提交该操作请求。在libiptc/libiptc.c中定义了iptc_commint()，它根据请求设置了一个struct ipt_replace结构，用来描述规则所涉及的表(filter)和HOOK点(FORWARD)等信息，并在其后附接当前这条规则的一个struct ipt_entry结构。组织好这

(下转第49页)

本文是在对红外人群光流研究的基础上发现低视角下存在的问题，然后提出一种理想化的模型对低视角光流进行了验证和改进。实际情况中对于人群的光流的分析是非常复杂和有挑战性的，由于对光流来说还存在很多的不确定性，诸如在正常的监控情况下，人员的走动方向和走动速度都是不受控制的，而且由于红外相机目前分辨率本身不是很高(文中用的相机分辨率仅为160*120)，对光流的计算也有很大的影响；光流算法本身理论上的精度问题，加上相机本身缺乏标定参数，重叠和遮挡造成的问题，拍出的图像仅包括二维的信息，因此对人多情况下的光流进行低视角的校正还需要进一步深入的研究，必须结合人的分割和三维信息的辅助才能更好的对光流进行改进，进而更精确的完成人群的运动分析。

参考文献

[1] Berthol Horn, Bdan Schunck. Detenming Optical flow. Artificial Intelligence[Z]. 1981. 185-203.

[2] Barron M, Fleet D J, Beauchemin S S. Performance of optical flow techniques [J]. International Journal of Computer Vision, 1994, 12 (1): 43-77.

[3] Michael J. Black. The robust estimation of multiple motions[J]. Computer Vision and Image Understanding,1996, (1):75-104.

[4] 马颂德,张正友.计算机视觉——计算理论与算法基础[M].北京:科学出版社,1998.

[5] 陈邦忠,王秀坛,涂建平等.红外图像序列的运动目标检测[J].探测与控制学报,2002,24 (3):11-13.

[6] 李熙莹,倪国强.红外图像的光流计算[J].红外与激光工程,2002,31(3):189-193.

[7] 胡以静,李政访,胡跃明.基于光流的运动分析理论级应用[J]. 计算机测量与控制,2007,15(2):219-221.

[8] 阮秋琦.数字图像处理[M].北京:电子工业出版社,2001.

[9] STILLER C, KONRADJ. Estimating motion in image sequences[J]. IEEE Signal Processing Magazine, 1999:70-91.

(收稿日期：2007-11-30)

(上接第51页)

些数据后，iptc_commit()调用setsockopt()系统调用来启动核心处理该请求。

4.3 使用TC命令实现对P2P数据的流量控制

在实现了对P2P数据包的识别之后，我们就可以通过Linux的带宽控制工具TC来实现对P2P应用的流量控制，将命令写成脚本文件在系统启动时自动加载。其中关键的命令如下：

```
tc qdisc add dev eth1 root handle 20: htb default 20 //定义新建下载最顶层根用户规则

tc class add dev eth1 parent 20: classid 20:1 htb rate 94000kbps ceil 940000kbps //定义10:1总下载带宽94M

tc class add dev eth1 parent 20:1 classid 20:10 htb rate 40kbps ceil 100kbps prio 0//建立一个通道，最小保证带宽40K，最大占用带宽100K。

tc qdisc add dev eth1 parent 20:10 handle 201: pfifo

tc filter add dev eth1 parent 20: protocol ip prio 100 handle 2010 fw classid 20:10 //建立队列策略和过滤

iptables -F -t mangle

/sbin/iptables -t mangle -A POSTROUTING -m ip2p --edk -kazaa -bit -j MARK --set-mark 2010 将edk,kazaa和bit应用的数据打上标记并丢到相应的通道里。
```

5 结论

通过实验该控制系统达到了预期的目标，一方面可以及时准确地发现常用P2P应用的数据包如BitTorrent、eMule，BitComet等；另一方面可以将上述P2P应用程序的下载速度很好地控制在用户设定的允许值之内。利用Linux的带宽控制技术可以为内网用户设定不同的优先级，将该系统安置在路由器之上，可以有效地监控和管理内网用户的P2P流量，实现流量均衡。

参考文献

[1] 张联峰,刘乃安,钱秀槐,等.综述:对等网(P2P)技术[J].计算机工程与应用,2003,39(12):142—145.

[2] 余青霓,周刚.Linux防火墙[M].北京:人民邮电出版社,2000.

[3] 向培素,田珂.带宽控制技术分析及其实现[J].西南民族大学学报,2003,29(4): 406-410.

(收稿日期：2008-01-20)

SHENG Pan-long, ZHAO Yu-ming(Inst. of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: Optical Flow is a most important method for moving object detection, and most usually used in parameter evaluation and object tracking. Infrared images always have large noises and low contrast, so few people use optical flow for moving object detection in infrared images, especially for pedestrian detection. This paper uses optical flow to detect people's moving speed and direction by using infrared images. Because in low angle situation, the results of optical flow always have near-far effect and the results have large errors. So, this paper proposes a method to reduce this kind of errors caused by near-far effect, and we have also done some tests on the acquired infrared images. The results showed this method had a good performance and the optical flow results were more accurate after compensation.

Keywords: Infrared image; Optical flow; Pedestrian detection; Low view; Differential method

ZHANG Q ing, LU Xiao-feng (Computer Science and Technology School, Suzhou University, Suzhou, Jiangsu 215006, China)

Abstract: This paper firstly introduces the application of P2P, and then, analyzes the Netfilter framework, the technology of Iptables and the technology of Traffic Control. With analyzing the characteristics of the P2P protocol, this paper puts forward a method to identify the data of P2P.

Keywords: P2P; Netfilter; Traffic control; Iptables

LIU Qiang XUE Hui-feng(Northwestern Polytechnical University, Xi' an, 710072, China)

Abstract: The air defense operational command and control system is complex. This paper analyzes the influence of the intelligent command and control system on the air defense effect from the battle field overall situation. The battle elements are simplified to be offensive and defensive sides. The system is composed of three operational Agents which are command and control system, the assaulting planes and the ground-to-air missiles. Through the simulation experiments by NetLogo4.0 software on the air defense system, this paper analyzes the influence of the command and control system intellectualization on the modern air defense effect air when the other operational conditions are certain

Keywords: CAS theory; Intelligent command and control system; Air defense; Effect assessment

SU Zhong-bin, ZHENG Ping, SUN Hong-min, ZHANG Ji-cheng (Beijing Institute of Technology, Beijing, China)

Abstract: In order to solve the problems of weak organ classification and worse code flexibility, the paper uses visualization and assembly technologies to design a crop growth system. It presents some characteristics of the system such as flexibility, easy to manage, and reuse of the codes. The system provides function-expanded interface for the simulation system, and lays the foundation for simulation soybean mass growth state. Therefore, it has great significance in visualization research of virtual crops.

Keywords: Assembly technology; Object-oriented; Soybean structure; Growth system

HUA Shan, HE Li-ming, TIAN Zuo-hua(Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: The control of airship fuel battery is requested to be real-time and highly accurate, and to keep high stability in extreme and changeful conditions. The CompactRIO system developed by National Instruments (NI) is introduced in the paper as the control system of fuel battery. The principle of fuel battery control and the design of control system based on CompactRIO are introduced. The system accomplished expected effect in the elementary debug.

Keywords: Fuel battery; CompactRIO; LabVIEW

MA Yan[1,3], WANG Wen-dong[1,3], LI Zhu-lin [2](1.Department of Computer Science, Yan'an University, Yan'an 716000, China; 2. 402 Division of the Second Artillery Engineering College, Xi'an 710025, China; 3.Software R&D Center, Yan'an University, Yan'an 716000, China)

Abstract: The meaning of the four-tier structure and the key technology are explained, and some suggestions on deploying the four-tier structure project based on ASP.NET are presented with combination of the application of the practical office automation system of Yan'an University.

Keywords: Four-tier structure; ASP.NET; ORM; Operation logic tier; Data access tier; Database

Learner's Garden

CUI Cong, ZHANG Zhong-neng (Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: This paper introduces an innovative matrix input method for sales orders in the traditional ERP system. This matrix-based method is extremely applicable for processing orders of multi-attribute products, such as clothing, footwear and zippers. Based on the general features of commercial sales orders, it begins with building a fundamental order model and then describing in detail the design and implementation of the matrix-order input method. It investigates the possible applicability and advantages of the method in the modern manufacturing industry. Finally, it concludes with insightful analysis of the performance and scalability of the matrix-based order input system.

Keywords: ERP; Sales order; Matrix; Multi attributes

Address: 1954 Huashan R d ., Shanghai, P. R. China
Zip Code: 200030
Tel: 86-21-62933230 Fax: 86-21-62933230
Email: smcaa@online.sh.cn URL: http//www.Smcaa.online.sh.cn
IP: 202.96.210.198 Publisher: Shanghai Microcomputer Application Association
Code Number: M 6329 Distributor: International Book Trading Corporation (P.0.Box 399,Beijing)