基于PPPoE的带宽汇聚NAT设计与实现

汤鹏杰,李奇润,唐凤仙
(河池学院 计算机与信息科学系,广西 宜州 546300)

摘 要:PPPoE技术使得已有的以太网能方便地实现宽带接入和计费功能,但Windows操作系统内置的PPPoE及ISP只支持用户单帐号登录,用户即使登录多个宽带帐号,再开启系统的宽带共享功能,也只能共享一个帐号的带宽。为解决这一问题,使用户能够在多帐号的状态下实现高速数据传输,在PPPoE的基础上,设计并实现了一个可以汇聚多个帐号带宽的NAT系统。实验结果表明,运行该系统时,用户登录帐号越多,链路带宽越大,实现了汇聚多帐号带宽的功能,为用户和ISP实现高速接入提供了另外一种途径。

关键词:PPPoE;NAT;TCP/IP;带宽汇聚
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2011)10-0095-04

Design and Implementation of Bandwidth Aggregation NAT System Based on PPPoE

TANG Peng-jie, LI Qi-run, TANG Feng-xian
(Department of Computer and Information Science, Hechi University, Yizhou 546300, China)

Abstract: The existing PPPoE technology supports that the users access to an Ethernet and billing easily. But the Windows built-in PPPoE protocol only supports single account, ISP also support only one on this basis account login. Although the consumer logs on multiple broadband accounts, and opens the system for broadband sharing, shares only one account's bandwidth. On the basis of PPPoE, designed and implemented a NAT system that can be brought bandwidth of multiple accounts together. Experiment results show that the system can achieve the intended target.

Key words: PPPoE; NAT; TCP / IP; bandwidth aggregation

0 引 言

无论是PPPoE(PPP Over Ethernet,基于以太网的PPP)还是NAT(Network Address Translation,网络地址转换)技术,都是当今使用最广泛的局域网宽带接入技术[1],在Windows操作系统中,这两种技术都已内置,但缺点就是不能同时登陆多个宽带帐号来汇聚网络带宽,这在需要一些高速下载或上传而又没有能力升级现有带宽的网络系统中是满足不了用户需求的。因此,设计并且实现了一种能汇集多个帐号带宽的NAT系统,从而使用户既可以达到高速下载或上传的目的,又不用浪费更多的时间和精力去升级现有网络硬件系统以达到扩大网络带宽的目的,同时也为ISP(网络服务提供商)在为用户提供网络服务时开辟了一个新的思路。

收稿日期:2011-03-25;**修回日期**:2011-06-27
基金项目:广西自然科学基金项目(2011jjB70037);河池学院引进人才科研启动项目(2010QS-N001)
作者简介:汤鹏杰(1983-),男,河南郸城人,硕士,讲师,主要研究方向为计算机网络。

1 PPPoE及NAT

1.1 PPPoE技术

PPPoE是在以太网上传播PPP帧的技术,该技术的实现需要经过两大阶段,即PPPoE发现阶段和PPPoE会话阶段[2,3]。在发现阶段期间,主机在以太网中寻找并选择一个AC(Access Concentrator,访问集中器),一旦找到,就确定了AC的物理地址和会话的ID。

发现阶段的数据交互流程如图1所示。

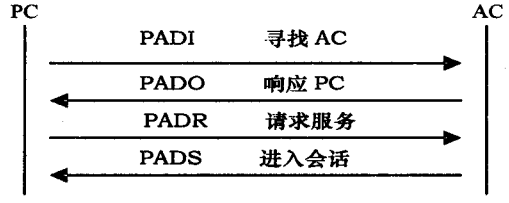

图1 PPPoE发现阶段交互过程

发现阶段成功之后,主机便确定了提供接入服务的AC,并且获取了AC的MAC及与AC进行会话的ID,接着就进入会话阶段,在此阶段期间,主机将与AC进行PPP协商及验证,协商包括两个部分的内容:LCP、NCP;至于验证,文中采用的是通用的PAP认证

协议。

整个协商/验证的流程如图 2 所示[4,5]。

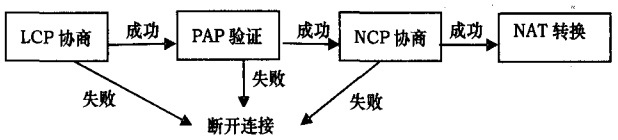

图 2 PPP 协商验证模型

如果要协商某个选项就发送 Configure-Request 报文,对方则根据自己的配置情况来进行应答:如果对收到的每个选项及全部值都接受就返回 Configure-ACK;如果每个收到的选项都是可认知的,但某些值不可接受,就返回 Configure-Nak,并在返回的报文中将值改为自己能接受的;如果选项中存在不可辨认的或不可协商的,就返回 Configure-Reject;如果收到带有未知代码的 LCP 包时,就返回 Protocol-Reject;Echo-Request 与 Echo-Reply 一般情况下用来测试链路的对端是否仍然存在,收到 Echo-Request 后得以 Echo-Reply 进行应答;Terminate-Request 为关闭一个连接,收到后必返回 Terminate-ACK,并释放 PPP 连接占用的资源。

LCP 主要协商三个选项:最大接受单元、认证协议、魔术字,魔术字仅是用来检测链路是否存在环路,最大接受单元(简称 MRU)最重要,它规定了 PPP 所能承载的最大数据长度,一般为 1492 字节。对于因特网接入,NCP 实质就是从 AC 处获取公网 IP 及 DNS 服务器的 IP。在验证阶段,主机发送用户名及密码给 AC,AC 进行核对后判断用户是否可以接入因特网。

1.2 NAT 系统结构

NAT 系统结构模型如图 3 所示[5]。

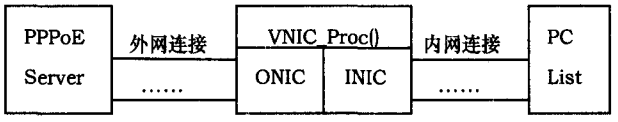

图 3 NAT 结构模型

在该结构模型中,有两种网络接口卡,第一种为 ONIC(外网网卡),每个帐号都需要一个 ONIC,一旦登录成功后,ONIC 就相当于与 AC 有了虚拟的连接,ONIC 里面包含了同 AC 进行会话时所必需的参数,以及为了便于 NAT 转换而设计的另外变量。流向 ONIC 的 IP 数据包有可能被转发至内网。第二种为 INIC(内网网卡),INIC 实质上就是内网的网关,它为内网的 PC 提供 NAT 转换,将流向外网的 UDP/TCP 数据包进行端口及 IP 转换、转发。VNIC_Proc() 过程处理所有接收到的以太帧,它会根据帧的目的地址、帧的类型来调用其他的子模块进行处理,使得 INIC 向内网表现出网关特性,ONIC 表现出 PPPoE 连接特性。VNIC_Proc() 驱动着整个 PPPoE 及 NAT,所有的网络 I/O 都在这里进行。

2 多帐号带宽汇聚 NAT 设计

2.1 网络 I/O 接口设计

"网络 I/O"模块提供了原始以太帧的收发接口,VNIC_Proc() 通过接收接口来获取所有的帧,其它的子模块则使用发送接口将帧传送到网络上去。此模块中定义了两个结构体:

```
typedef struct _IntfInfo
{
int Number;// 可用接口的数量
char * IntfName[MAX_INTERFACE];// 各接口的名字
char * IntfDesc[MAX_INTERFACE];// 各接口的描述信息
}IntfInfo;
typedef struct _Intf
{
int Index;// WinPcap 绑定的下层网卡序号
int RecvBufLen;// 接收缓存区长度 B
int SendBufLen;// 发送缓存区长度 B
int RecvTimeOut;// 接收超时 ms
int RecvPktLen;// 接收数据包的最大长度 B
}Intf;
```

在初始化网络 I/O 前,必须先知道 WinPcap 已绑定了哪些下层的网卡。下面函数可以获取网卡信息:

```
int IntfInfo_Init(IntfInfo * pIntfInfo);
```

在调用 IntfInfo_Init() 后,如果返值为 1 则所有的网卡信息都保存在 pIntfInto 所指向的内存里。接着分配并填充一个 Intf 结构体,其中,Index 域指明了所选择的接口序号,其范围为[0, IntfInfo. Number -1],RecvBufLen 域指明了接收缓存区的长度,当有帧到达网卡接口时会先把它保存起来,这样可以有效地减缓由于流量过大而导致帧的丢失,RecvTimeOut 域指明了接收帧的间隔,选择 1ms,RecvPktLen 域指明要接收数据包的最大长度,在以太网下 MTU 一般都为 1500B,再加上以太帧首部 14B[6,7],故设置为 1514。最后,调用接口初始化函数 Intf_Init(IntfInfo * pIntfInfo,Intf * pIntf),表示选择接口的 Index。如果返值为 1,就表明网络 I/O 初始化成功,然后,就能通过以下两个 I/O 函数收发以太帧了:

```
int Intf_RecvPkt(char * RecvBuf);
int Intf_SendPkt(char * SendBuf,int SendLen);
```

2.2 TCP/IP 协议栈设计

由于要用到一些特殊的功能,因此,开发了一个简易的独立于 Windows 系统协议栈的 TCP/IP 协议栈。在网络接口层中,ARP(地址解析协议)可以将 IP 地址映射成 MAC 地址,内网网关 INIC 应该具备 ARP 协

议,这样可以方便其它 PC 的配置,整个 ARP 模块实现了对 ARP 请求的应答,并保存相应的 ARP 映射关系[8,9]。

在 ARP 模块中,映射关系被定义成如下结构体:

```
typedef struct _ARPCache
{
unsigned int IP;
unsigned char MAC[6];
}ARPCache;
```

向外导出了两个实现 ARP 操作的接口,一个用于 ARP 的应答,一个用于查询 ARP 表:

```
void ARP_Reply(INIC * pINIC,char * pBuf, int iLen);
unsigned char * ARP_Find(unsigned int ARP_IP);
```

ARP 模块不需要进行初始化,它只是实现 ARP 的应答请求,以确保内网的 PC 能够获取网关的 MAC,并向 NAT 模块提供 IP 到 MAC 映射的查询功能。

在网络层中,IP 数据包在网络传输的过程中,如果链路层的 MTU 小于 IP 包的大小,则此数据包就会被分成小于或等于 MTU 的若干个小数据包再进行传送,将 IP 进行分片的功能在 NAT 模块中集成,如果需要分片的话,数据包就会被分成 OOFFSet 或 IOFFSet 的大小再进行传送。

若 ONIC 接收到被分片的 IP 包,则就需重组,因为不重组的话,就无法再重新计算 TCP 或 UDP 的校验和,IP 重组模块可以实现重组功能[10,11]。

IP 数据包分片模块结构体定义:

```
typedef struct _IPFrgPkt
{
struct _IPFrgPkt * pNextPkt;// 下一个 IP 分片包
struct _IPFrgPkt * pNext;// 当前 IP 分片包的下一个内存块
int iLen;// 当前内存块存放的数据长度
char Data[IPFRGPKTLEN];
}IPFrgPkt, * PIPFrgPkt;
```

IP 分片重组管理模块标识结构体定义:

```
typedef struct _IPFrgModle
{
HANDLE hTimer;// 时钟线程的句柄
int bStatu;// IP 分片重组模块的状态
IPFrg * pIPFrgList;// IP 分片链表
MemBlk MemList;// 内存链表
CRITICAL_SECTION Lock;// 同步对 IP 分片链表的操作
}IPFrgModle;
```

PPPoE 协议的连接、维护功能都在 PPPoE 模块中实现,它向外导出以下函数接口:

```
int PPPoE_Conn(ONIC * pONIC,char * * pPBuf);    // 使用 ONIC 来进行连接
int PPPoE_DisConn(ONIC * pONIC);
// 断开 ONIC 与 AC 的连接
void PPPoE_AddDisConnected(ONIC * pONIC);
// 向 PPPoE 模块增加一个已断开的 ONIC
ONIC * PPPoE_GetDisConnected();
// 获取已断开连接的所有 ONIC
ONIC * PPPoE_GetPublicIP_Proc(char * pBuf,int iLen);  // 获取公网 IP 的处理过程
int PPPoE_PPP_Terminate_ACK_Send(char * pBuf,int iLen);
int PPPoE_PADT_Send(ONIC * pONIC);
```

PPPoE_Conn()实现用户登陆的功能,调用 PPPoE_Conn()时,就不能再在其他地方调用它,因为 PPPoE 模块只允许一个帐号登陆结束后再登陆另一个帐号,返值为 PPPoE_SUCCESS 则表示登陆成功。

PPPoE_GetDisConnected()用于获取已断开连接的 ONIC,登陆失败或者一个 PPP 连接被断开,那么对应的 ONIC 就会被插入 PPPoE 模块中,直到调用此函数才被移除。

2.3　带宽汇聚

如图 4 所示,设有 n 个 ONIC 登陆成功,则有 n 条到 AC 的链路,此时公网 IP 地址分别为(IP_1,IP_2,…,IP_n),每条链路都具有 x Mb/s 的带宽,主机 PC 将 INIC 作为网关,并且 PC 与外网的计算机 S 通信,S 的带宽足够多。如果要汇聚流量,PC 与 S 的通信就必须满足一些要求:PC 的应用层程序应该向 S 发起 m 个连接,NAT 将这 m 个连接分配到 n 条链路上去,转换成(IP_1,IP_2,…,IP_m)与 S 的连接。当客户下载文件时,PC 可以启动 m 个连接,分别请求下载 S 上某文件的 m 个数据块,这 m 个请求是相互独立的,NAT 将每个连接分配到各个链路上。在 S 看来,请求的主机就像是(IP_1,IP_2,…,IP_m),S 分别将数据块传送给(IP_1,IP_2,…,IP_m)。但在 PC 看来,通信的目的地只有一个,与 S 的每个连接都具有 x Mb/s 传输的速度,故文件传送的总带宽为 $m * x$ Mb/s。

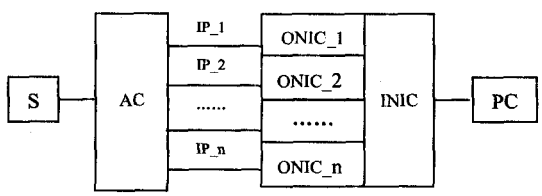

图 4　多帐号流量汇聚模型

其实,对于其他一些 P2P 服务也能达到类似的效果,它们服务相同、请求独立[12],只要进行多链接,就

能汇聚多个链路的带宽。

3 实验结果及分析

本实验使用了虚拟机 VM 6.0.2 来搭建网络环境。采用 ROS 2.9.27 作为 PPPoE 服务器;将 PPPoE 服务器的上下行流量限制为 1 Mbps/s,然后启动 NAT 进行登陆,运行虚拟机的 Windows2003,将网关的地址设计为 INIC 的 IP,并开启迅雷下载。测试结果如表 1 所示。

表 1 不同帐号个数下的下载速率

帐号个数	帐号带宽(kbps)	下载速率(kbps)
1	1000	987.6
2	1000	1924.4
3	1000	3915.2

从测试结果来看,在帐号带宽同为 1Mbps 的情况下,下载速率随帐号个数的增加而呈线性增长,达到了系统设计的目的,实现了多帐号带宽汇聚的功能。

4 结束语

文中在 PPPoE 连接获取公网 IP 后,设计并实现了一种可以汇聚多帐号带宽的 NAT,并最终测试成功。由于时间、水平的限制,还有些需要改善的地方,整个设计可以在 NDIS 内核驱动中实现,而文中是使用 WinPcap 在应用层上实现的,所以效率不及前者,此外,由于一般情况下 IP 头选项几乎没用,ICMP 协议大都对网络进行差错报告,所以文中所设计的系统暂时不支持这两样功能。

参考文献:

[1] 王艳平. Windows 网络与通信程序设计[M]. 第 2 版. 北京:人民邮电出版社,2009.

[2] 赵雪峰. 基于 PPPoE/PPP 协议的带宽接入客户端拨号软件的实现[D]. 北京:中国地质大学,2005.

[3] RFC2516. A Method for Transmitting PPP Over Ethernet (PPPoE)[S]. [s.l.]:Network Working Group,1999.

[4] 邹 航,杨元晔,苟光磊. NAT 网络地址转换技术分析[J]. 重庆工学院学报,2007,21(7):89-91.

[5] 肖辽亮. NAT-PT 簇负载平衡的设计与实现[J]. 计算机技术与发展,2006,16(3):80-82.

[6] 王 南,孙保锁,王月平. P2PSIP 系统中 NAT 穿越方案的研究与设计[J]. 计算机技术与发展,2009,19(10):66-69.

[7] RFC3022. Traditional IP Network Address Translator[S]. [s.l.]:Jasmine Networks,2001.

[8] 郭士秋. IP 协议体系[M]. 北京:电子工业出版社,2002.

[9] Stevens W R. TCP/IP 详解 卷 1:协议[M]. 范建华,张涛,译. 北京:机械工业出版社,2007:33-34.

[10] WinPcap Team. WinPcap Documentation[EB/OL]. 2002. http://www.winpcap.org/docs/docs_411/html/main.html.

[11] 罗军舟,黎波涛,杨 明,等. TCP/IP 协议及网络编程技术[M]. 北京:清华大学出版社,2004:22-26.

[12] Comer D E. 用 TCP/IP 进行网际互联(第一卷:原理、协议与结构)[M]. 林 瑶,蒋 慧,杜蔚轩,等译. 第 4 版. 北京:电子工业出版社, 2001.

(上接第 94 页)

高,达到了预期效果;对长词的切分能力更强。

同时,算法也还存在不足之处。中文是一种较为复杂的语言,其结构复杂多样,用法灵活多变,应用也无处不在,这就决定了其词库非常庞大,而且要求高效;另外,随着社会、经济的飞速发展,大量的新词语往往随机涌现,这些新词语必然在词库中尚未收录,这给词库的更新及维护带来了新的挑战。因此,词库的建立是个巨大而艰难的工程,对词库的维护、更新及新词语的识别还有待进一步研究。

参考文献:

[1] 孙茂松,邹嘉彦. 汉语自动分词研究评述[J]. 当代语言学,2001(1):22-32.

[2] 李淑英. 中文分词技术[J]. 商丘科技职业学院学报,2007(36):95-95.

[3] 张春霞,郝天永. 汉语自动分词的研究现状与困难[J]. 系统仿真学报,2005,17(1):138-147.

[4] 金在全,赵 照,杜秀全. 一种改进的增字最大匹配算法[J]. 科学技术与工程,2007,18(7):4161-4164.

[5] Li Haizhou, Yuan Baosheng. Chinese Word Segmentation [C]//Language, Information and Computation (PACLIC 12). [s.l.]:[s.n.],1998:212-217.

[6] Xue Nianwen. Chinese Word Segmentation as Character Tagging[C]//Computational Linguistics and Chinese Language Processing. [s.l.]:[s.n.],2003:29-48.

[7] 徐 飞,孙劲光. 中文分词切分技术研究[J]. 计算机工程与科学,2008,30(5):126-128.

[8] 文庭孝,邱均平,侯经川. 汉语自动分词研究展望[J]. 现代图书情报技术,2004(7):6-10.

[9] 冯书晓,徐 新,杨春梅. 国内中文分词技术研究新进展[J]. 情报检索,2002(11):29-30.

[10] 曹卫峰. 中文分词关键技术研究[D]. 南京:南京理工大学,2009.

[11] 宋国柱,陈俊杰. 基于双字词的动态最大匹配分词算法的研究[J]. 太原科技大学学报,2009,30(3):199-202.

[12] 龙树全,赵正文,唐 华. 中文分词算法概述[J]. 电脑知识与技术,2009,10(5):192-193.

[13] 尹 锋. 汉语自动分词研究的现状与新思维[J]. 现代图书情报技术,1998(4):22-26.