

宽带接入交换设备中的 VLAN 实现技术^{*}

The VLAN Implementation Technology in the Broadband Access Switch

孙利民 张 灏 王 军

(中科院软件所多媒体网络研究中心 北京100080)

Abstract The broadband LAN is an access scheme that is adaptable to the condition of china. The VLAN in access switch can utilize effectively the bandwidth and enhance the data safety. After briefly introducing the functions of VLAN, this paper then gives the hardware platform of access switch to implement VLAN, explains the implementation technology of VLAN in detail. And finally we show how to implement multicast snoop based on VLAN

Keywords Broadband access switch, VLAN implementation, Multicast technology

1 引言

宽带接入目前主要有 HFC、ADSL 和局域网等方式,以太网接入因其使用简单、扩展性好,适合密集型住宅小区/办公大楼而备受国内关注。在我们的局域网接入实现中,接入交换设备的每个端口为一个用户独占,每个用户独享10/100Mbps 带宽,在接入交换设备中需要解决其中的带宽合理利用和数据安全问题。合理使用带宽可以从两方面考虑:一方面降低广播风暴;另一方面消息不是广播到每个用户,而是用户主动请求方式。这需要采用 VLAN 技术和组播(multicast)技术。

VLAN 在交换设备中建立虚拟的工作组,逻辑上相当于 OSI 模型第二层的广播域。在交换设备中,通常广播风暴发送到每一个交换端口,形成整个交换网络的广播风暴。VLAN 把传统的广播域按需要分割成各个独立的子广播域,将交换机某个端口或用户赋予某一个特定的 VLAN 组,在一个 VLAN 中的广播风暴不会送到 VLAN 之外,相邻的端口不会收到其他 VLAN 产生的广播风暴。同时 VLAN 可提供防火墙的机制,防止交换网络的过量广播风暴,限制用户的访问域、控制组的大小及位置等,增加了网络的安全性。

下面首先简要介绍 VLAN 实现的接入交换设备硬件平台,然后重点说明基于端口的以及基于协议和端口混合的 VLAN 的实现,最后给出如何在 VLAN 基础上实现组播协议。

2 VLAN 实现的硬件平台

接入交换设备采用如图1所示基于交叉开关的分布式存储转发结构,核心芯片是 Galileo 公司的产品。交叉开关 GT48300带有4个 Glink 接口和一个 PCI 接口,Glink 接口的双工速率为2.4Gbps,CPU 通过 PCI 接口收发包和管理交换设备。GT48300实现所有端口的无阻塞交换。交换芯片 GT48310A 采用存储转发结构,连接8个10/100M 以太网接口,带有数据 SDRAM 和控制 SDRAM,数据 SDRAM 存放接收的包,控制 SDRAM 存放 VLAN 表、地址表和控制结构等。GT48310A 根据 IEEE802.1Q 标准提供对 VLAN 的支持,包括支持多个转发数据库,完成标志的提取和插入等。CPU 和每个 GT48310A 都有设备编号,最多可有32个设备,每个10/100M 端口都有端口编号。GT48310A 之间以及每个 GT48310A 和 CPU 之间发送控制信息或数据包通过 GalNet 消息,GalNet 消息中带有目的/源的设备编号和端口号。

在所有的 GT48310A 中都维护完全相同的地址表,其简化结构如表3所示。V 是有效位,地址 Addr [47:0]对应的 VLAN 号为 Vid[11:0],对应的设备编号和端口编号分别为 Dev # 和 Port #,A 是地址老化(aged)使能标志,St 是静态地址标志。

每个 GT48310A 有一个 VLAN 表,每个 VLAN 包含4k 个表项,各设备的 VLAN 表基本一致。简化 VLAN 表结构如表1所示,VLAN Entry 是 VLAN 号,

^{*} 本文研究得到国家973课题的资助。孙利民 博士后,副研究员,主要研究领域为高性能网络和多媒体技术。张 灏 硕士生,主要研究领域为计算机网络。王 军 博士生,主要研究领域为多媒体技术和软件工程。

VLAN 包含的设备用 Dev[31: 0]表示,若 VLAN 包含第 I 号设备,则第 I 位为“1”,否则为“0”,本 VLAN 表所在设备的设备号对应位为“0”,本设备的 8 个端口

是否包含在该 VLAN 中用 Port[7 0]表示,V 是有效位。

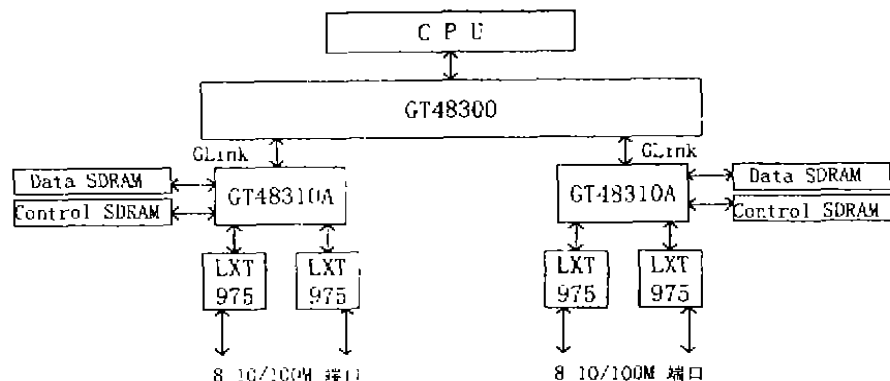


图1 接入交换设备的硬件结构

GT48310A 可以工作在 Tag 或 Basic 模式下,这两种模式的基本差别是接收包是否带有 VLAN 信息。Tag 模式下,希望接收包带有 VLAN ID,如果没有,GT48310A 将自动给它加上 VLAN ID;Basic 模式下,也可用 VLAN ID,但必须通过管理 CPU 赋值。不管 VLAN ID 是如何建立的,CPU 必需将 VLAN 信息 (Dev 和 Port 等)写入 VLAN 表。

3 VLAN 的实现技术

为了更好地理解 VLAN 的实现,下面详细讨论 VLAN 机制的基本规则及 Basic 和 Tag 模式的区别。

在 Basic 模式中,地址表的表项仅依赖于源 MAC 地址;在 Tag 模式中,地址表的表项依赖于源 MAC 地址和 VLAN ID,对于同一个地址及其不同的 VLAN ID 可能有多个表项。地址学习在 Basic 模式中是 GT48310A 在地址表中搜索接收包的源地址(SA),如果在地址表中没有发现 SA,GT48310A 在该包 CRC

校验正确后创建地址表表项(VLAN ID=0);如果发现了SA,但地址表和包中的设备号和端口号不等,则更新地址表;否则,不需要任何动作。在Tag模式中有两种情况:(1)如果接收包有非零的VLAN ID标签,以太网类型合法,表项的VLAN ID为接收包的VLAN ID;(2)如果VLAN ID=0,或接收包没有标签,表项的VLAN ID为端口控制寄存器的(Default-VLAN-ID)值。两种模式下,CPU都可以随时更改地址表。接收包交换过程为:在两种模式下,单播包根据地址表中的相应表项进行转发(Tag模式中需要检查VLAN ID),多播包和未知目的地址包根据VLAN表的相应表项转发。Basic模式中,不能给地址表赋予不同的VLAN ID,需要CPU负责为地址表表项更改VLAN ID,而Tag模式中,VLAN ID由端口控制寄存器获得缺省的VLAN ID。

3.1 Basic 模式中的基于端口的 VLAN

表 1 设备 1 VLAN 表

[illegible]

表 2 设备 2 VLAN 表

VLAN Entry	Dev[31..0]	Port[7..0]	v
2	000000000000000000000000000010	01010101	1
3	000000000000000000000000000010	10101010	1

表 3 地址表

Addr[47:0]	Vid[11:0]	Dev#	Port#	St	A	v
0x123456789abc	2	1	3	1	1	1

使 VLAN 基于端口分类,Port-Control-Extend(Protocol-bits)位能被每一个端口定义,因此对同一协议最多定义4个 VLAN。在 Tag 模式中,不同点是如果包的标记 VLAN ID 不为0,则 VLAN ID 来自标记而不考虑协议过滤机制。

表7 VLAN ID

VLAN ID Bits	Source
11	Always 1
[10:9]	PortControlExtend.Protocol bits[1:0]
8	Ethertype field bit 15
7	Ethertype field bit 13
6	Ethertype field bit 11
5	Ethertype field bit 6
4	Ethertype field bit 5
3	Ethertype field bit 4
2	Ethertype field bit 2
1	Ethertype field bit 1
0	Ethertype field bit 0

下面例子在两个 VLAN 中过滤 IPv4 包,(1) 端口 0、1、2、3 为一个 VLAN,端口 4、5、6、7 为另一个 VLAN;(2) IPv4 EtherType 为 0x800;(3) 设置端口 0、1、2、3 的 Port-Control-Extend 为 Protocol-En=1, Protocol-Bits=01,设置端口 4、5、6、7 的 Port-Control-Extend 为 Protocol-En=1, Protocol-Bits=10;(4) 在 VLAN 表中加入下面的表项。端口(0、1、2、3)接收到任何 EtherType 为 0x800 的包,它仅在该 VLAN (A40)中广播。

我们设置交换设备工作在 Tag 模式下,各个 10/100M 端口的缺省 VLAN 互不相同,CPU 和连接路由交换机的端口在所有端口缺省的 VLAN 中,这样每个端口的广播包不会发送到其它端口,只发送给 CPU 和连接路由交换机的端口,可有效降低广播风暴。同时,提供灵活的 VLAN 界面,可配置每个 VLAN 包含的设备和端口,给定地址对应的 VLAN 号。

表8 VLAN 表

VLAN Entry	Dev[31..0]	Port[7..0]	v
A40	00000000000000000000000000000000	00001111	1
C40	00000000000000000000000000000000	11110000	1

4 基于 VLAN 的组播

GT48310A 基于二层的交换,能够提取 IGMP 包并转发给 CPU,我们在 VLAN 上实现了 IGMP

snoop,其状态转换如图2所示,IGMP 守护进程维护一个组播表,为每个多播组(组播地址)建立一个表项,记录它对应的 VLAN ID、包含的各设备的端口号、每个端口生存周期等。

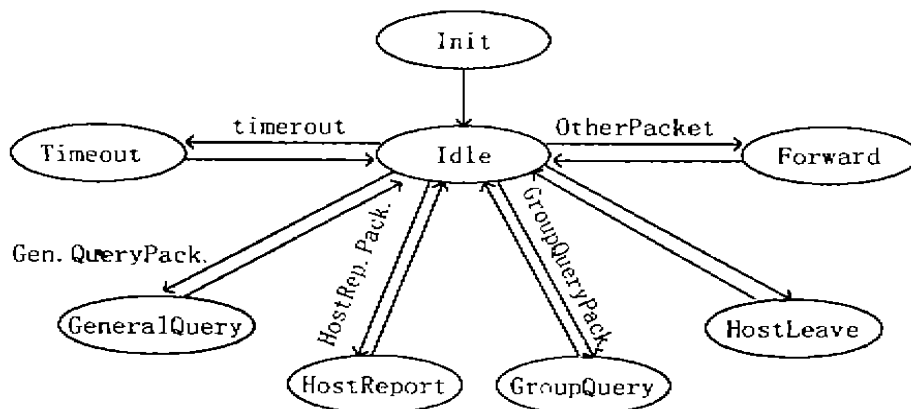


图2 IGMP 守护进程状态转换图

在初始化(Init)状态,写所有 GT48310A 的 VLAN 表,创建主机多播组“224.0.0.1”的 VLAN ID=1,设置端口生存周期无限大;然后,进入空闲(Idle)状态,等待接收 IGMP 包,根据 IGMP 包的类型进入不同的状态并进行不同的处理,处理完之后又回到空闲状态。

当接收到路由交换设备发送的组成员查询包时,首先检查路由器多播组“224.0.0.2”是否存在,如果不存在,创建路由器多播组“224.0.0.2”的 VLAN;如果存在,检查该端口是否包含在“224.0.0.2”的 VLAN 中,不包含就在该 VLAN 中加入此端口,然后,设置/

(下转第61页)

结束语 通过在 ns2 上的模拟,我们研究了出现网络链路总流量自相似的条件,说明 TCP 的阻塞控制机制可以成为网络自相似现象的决定性因素,而不必要高可变性。而由于模拟配置很简单,可以很方便地改变其中流速、延迟、排队策略等因素,考察这些因素对自相似现象的影响,这也是我们今后工作的方向。

参考文献

- 1 Veres A, Boda M. The Chaotic Nature of TCP Congestion Control. IEEE Infocom'2000, Tel Aviv, Israel, Mar 2000
- 2 Leland W E, et al. On the Self-Similar Nature of Ethernet Traffic. ACM SIGCOMM'93, San Francisco, CA, USA, Sep. 1993
- 3 Paxson V, Floyd S. Wide Area Traffic: The Failure of Poisson Modeling. IEEE/ACM Transactions on Networking, 1995, 3: 226~244
- 4 Willinger W, et al. Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. IEEE/ACM Transactions on Networking, 1997, 5(1): 71~86
- 5 Crovella M E, Bestavros A. Self-similarity in world wide

- web traffic - evidence and possible causes. In: Proc. of ACM Sigmetrics'96, 1996: 160~169
- 6 Taqqu M S, Teverovsky V, Willinger W. Estimators for long-range dependence: an empirical study. Fractals, 1995, 3(4): 785~798
- 7 Abry P, Veitch D. Wavelet Analysis of Long-Range-Dependent Traffic. IEEE Transactions on Information Theory, 1998, 44(1)
- 8 Morris R. TCP Behavior with Many Flows. IEEE International Conference on Network Protocols, Oct. 1997
- 9 Claffy K C, Braun H W, Polyzos G C. A parameterizable methodology for internet traffic flow profiling. IEEE Journal on Selected Areas in Communications, 1995, 13: 1481~1494
- 10 Floyd S, Jacobson V. Random Early Detection Gateways for Congestion Avoidance. IEEE/ACM Transactions on Networking, Aug. 1993
- 11 Feldmann A, Gilbert A G, Willinger W. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. Computer Communication Review, 1998, 28(4)
- 12 Feldmann A, et al. Dynamics of IP traffic: A study of the role of variability and the impact of control. In: Proc of the ACM/SIGCOMM'99, Cambridge, MA

(上接第68页)

修改路由端口生存周期为查询间隔。最后,将该包转发给 ID=1 VLAN 对应的所有端口(主机)。

当接收到主机组成员关系报告包时,IGMP 守护进程首先判断该多播组是否存在,如果不存在,为该多播组(地址)创建新的 VLAN,并将新的 VLAN 写到所有 GT48310A 的 VLAN 表;如果存在,检查该端口是否在 VLAN 中,如果不在,加入 VLAN 中,修改所有 GT48310A 的 VLAN 表;然后,设置/修改该端口的生存周期为组成员资格间隔,并将该包转发给“224.0.0.2”VLAN 对应的所有端口(路由器组)。

当接收到主机离开某个多播组包时,IGMP 守护进程转发该包给“224.0.0.2”VLAN 对应的所有端口(路由器组)。

当接收到路由交换设备发送的特定组成员查询包时,IGMP 守护进程将该包转发给特定组对应 VLAN 的所有端口(主机),修改该特定组 VLAN 的所有端口生存周期为离开查询时间间隔。

当接收到其它 IGMP 类型包时,IGMP 守护进程转发该包给除进入端口外的其它所有端口。

IGMP 守护进程设置一个定时器用于更改组播表项的每个端口生存周期。当接收到一个 IGMP 包或定时器到期时,修改表项中每个端口的生存周期。

这样,可以为每个电视频道(或 VOD 点播视频

流)建立一个多播组,每个用户选择电视频道(或 VOD 点播视频流)实际就是动态加入/离开一个多播组。我们通过实验室的两个卫星接收器,成功进行了计算机终端选择卫星电视频道的测试。

结束语 VLAN 的标准正在逐渐地统一和完善,它的种类比较多,各厂家的实现也互不相同,但大多数厂家支持 IEEE802.1Q 标准。我们在 Galileo 公司提供硬件支持的基础上,在宽带接入交换设备中实现了基于端口的 VLAN、基于协议和端口混合的 VLAN 以及基于 IP 组播的 VLAN。在我们正在研发的宽带接入路由交换设备中,可以支持更高层次的 VLAN、IP 组播协议的实现更为简单。

参考文献

- 1 Draft Standard P802.1Q/D11 IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks July 30, 1998
- 2 Fenner W. RFC2236 Internet Group Management Protocol, Version 2. Nov. 1997
- 3 Parkhurst W R. 网际工作组译: Cisco 组播路由与交换技术. 机械工业出版社, 1999
- 4 Rubin A. Linux Device Drivers. O'Reilly & Associates, Inc. 1998
- 5 Galileo Inc. GT-48310/GT-5.311/GT-48312 Switched Ethernet Control for 100BASE-X Revision 1.2 April 03, 2000
- 6 Galileo Inc. GT-48300 GalNet-II Crossbar Switch Revision 1.1 June 16, 1999