

间隙整形和设计实现

郑郁正¹ 曾家智²

(成都信息工程学院通信工程系 成都 610225)¹ (电子科技大学计算机学院 成都 610054)²

摘要 本文提出的间隙整形完全不同于现有的通信量整形。它能够有效地防止拥塞,构建无拥塞网络,并且能够利用高优先级任务已分配但暂时不用的带宽资源。间隙整形的另一个优点是容易在网卡上实现,完全不占用主机资源,效率高。文中还给出了一种可以满足间隙整形的以太网卡设计方案,研制后的测试数据完全达到设计要求。

关键词 间隙整形,拥塞,服务质量,以太网卡

Inter-frame Gap Shaping and its Design Implementation

ZHENG Yu-Zheng¹ ZENG Jia-Zhi²

(Department of Communication Engineering, Chengdu University of Information Technology, Chengdu 610225)¹

(School of Computer Science and Engineering, UESTC, Chengdu 610054)²

Abstract In this paper, the inter-frame gap shaping is first proposed, which is totally different comparing the current shaping and can be used to construct a network without congestion and utilize the bandwidth resource which has already assigned to but not yet used by high priority task. Another advantage of inter-frame gap shaping is its easy implementation on network card without occupying the host resource and higher efficiency. A network card design scheme with inter-frame gap shaping is developed in this paper and test result shows the inter-frame gap shaping network card meets the design requirements.

Keywords Inter-frame gap shaping, Congestion, QoS, Ethernet card

互联网的核心理念是端到端的透明性,即“终端智能化而网络简单化”。互联网因这个理念而迅速飞跃发展,但36年后的今天,也是因为这个理念缺乏安全性、可管理性、可扩展性和经济可持续发展性而面临严峻的挑战^[3]。互联网无法满足保证服务质量等新需求,而日益严重的安全问题已经成为燃眉之急。现在,互联网的问题已经到了非彻底解决不可的时候了。因此,联合国于2004年11月成立了联合国互联网治理工作组(WGIG);2005年8月美国自然科学基金会(NSF)正式公布了重构互联网的计划,即网络创新的全球化环境GENI(Global Environment for Network Innovation);我国在“863”和“973”指南中,也增添了新型网络体系结构的研究课题。

针对新型网络体系结构,我们提出了帧间间隙整形(下称“间隙整形”)方案。间隙整形完全不同于现有的通信量整形。间隙整形能够有效地防止网络拥塞,在其它技术的配合下,能够保证服务质量QoS,并且兼有共享资源的优点。至于间隙的确定和计算参见文[1],交换机间隙的确定将另文讨论,此不赘述。

1 间隙整形和无拥塞网络

间隙整形,就是上游节点发出一个帧后,等待一个时间间隙 Δ ,以便下游节点处理完这个帧后再发送下一个包。

这样,下游节点就不会被“淹没”。如果网络的所有链路都采用间隙整形,则网络不会发生拥塞。这样的网络称为无拥塞网络。

我们采用讨论拥塞时常用的模型和术语,把网络所有主

机单位时间发出的包数作为网络的输入 x ,而把网络所有主机单位时间收到的包数作为网络的输出 y ,此时的网络只包含路由器和交换机等网络设备和介质。那么我们就可得到如图1所示曲线和模型。图1(a)中的无拥塞网络对应一条折线。折点前 y 和 x 是线性关系,表示网络的所有链路都还有空闲的包隙时槽,即网络所有的链路信号的流量和还小于容量;折点后是水平线,表示网络的包隙时槽都已在传送包,链路传输的各种信号的流量和已经等于链路的容量。

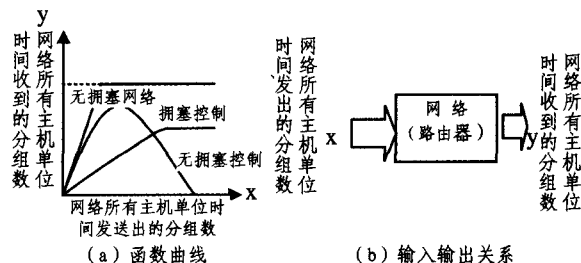


图1 描述拥塞的网络模型

在我们的网络架构下,同时采用间隙整形、虚电路结构、资源预留和准入控制时,只要能够建立虚电路,就工作在流量线性段。否则,就无法建立虚电路,也就不能通信。工作在线性段的无拥塞网络具有流量可加性,即整数个输入同时产生的输出流量等于每个输入单独产生的输出流量之和。

2 服务质量 QoS

由于流量可加性,我们可以分别讨论某个主机的某一个输入数据流在网络中传送的性质,完全无须考虑网络成千上

万的其它传输对它的影响,这使网络的复杂性大大降低。

在讨论服务质量(QoS)时,我们可以分别讨论可靠数据流、实时数据流和孤立数据。在传送文本数据时,可以通过滑动窗口协议来解决包的出错丢失和保序问题,所以接收方能够得到和发送方完全相同的文本。在传送实时音视频数据时,传送出错丢失问题,在当前技术条件下,一般是可以容忍的。抖动可以通过实时传输协议 RTP 解决。延迟对于单向传送不是问题;对于多媒体交互,例如电话,必须解决。延迟包括两部分:介质传播时间和节点处理时间。在当前技术条件下,一般是可以容忍接收接受的。至于孤立数据由于流量和数据流相比非常小,不必每次传送时预留容量资源,只须简单地在总分配容量中留有百分之几的余额即可。

我们采用四级优先级。最高为孤立数据,依次为三种数据流:紧急文本、实时音视频和一般文本^[2]。紧急文本指紧急并且重要的数据。当间隙到时,只传送存在的最高优先级包。由于一般文本优先级最低,我们可以不分配资源。由于间隙整形的特点,可以利用高优先级已分配但暂时不用的带宽资源。因此,不仅能保证服务质量 QoS,而且兼有共享资源的优点。

3 当前以太网卡的帧间隔

按照 IEEE 802.3 标准,网卡发送的每两个帧之间必须有一个帧间间隔 IPG(Inter-Packet Gap),最小的帧间隔为 12 字节(96 位)的传输时间。部分网卡的帧间间隔可以通过软件编程调整,进一步调小,最短为 84 位传输时间。这样,对于 10Mbps 的速率来说,间隙时间就在 8.4 μ s 至 9.6 μ s 的范围内;而对于 100Mbps 的速率来说,间隙时间就在 840ns 至

960ns 的范围内。现有以太网卡的帧间隔用于接收方判断帧的结束。在半双工系统中,帧间隔还提供让其它网卡竞争发送帧的机会。但是本文提出的间隙整形却是为了防止网络拥塞和保证 QoS,而调整这个帧间(增大到 ms 级)。

综上所述,间隙整形可以利用当前的以太网接口芯片进行设计实现。本文以当前用得较多的 Realtek 公司的 RTL8139(D)L 芯片为例进行讨论。

4 设计方案

由于选用符合 802.3 网络协议集的 RTL8139 芯片设计间隙整形网卡,必须从软件和硬件两个方面修改 RTL8139 的网卡设计方案。

首先是关于 RTL8139 发送缓冲区的问题。由于 RTL8139 内部缓冲区中的帧是按先进先出方式发送的,如果缓冲区中有较多低优先级别的帧排在前面等待发送,后进入缓冲区中,或者还在主机中的高优先级别帧就无法按优先级顺序及时发送。想要在 RTL8139 上实现按优先级别发送,只能借助软件实现。定义 RTL8139 正在发送的帧为 F_s ,而 RTL8139 缓冲区中下一个等待要发送的帧为 F_w 。在任何时候,在 RTL8139 的发送缓冲区中,至多只存入一个待发帧,即 F_w ,其它待发帧都保存在主机中。修改网卡驱动程序,当 RTL8139 发送完 F_s ,经过时间间隙 Δt ,将 F_w 变为 F_s ,此时驱动程序从主机中选择一个最高优先级别的帧包送入 RTL8139,成为 F_w 。这样,只要在 F_s 发送期间能及时将待发帧写入 RTL8139 缓冲区中成为 F_w ,那么 F_w 总是级别最高的帧,满足了间隙整形的理论要求。

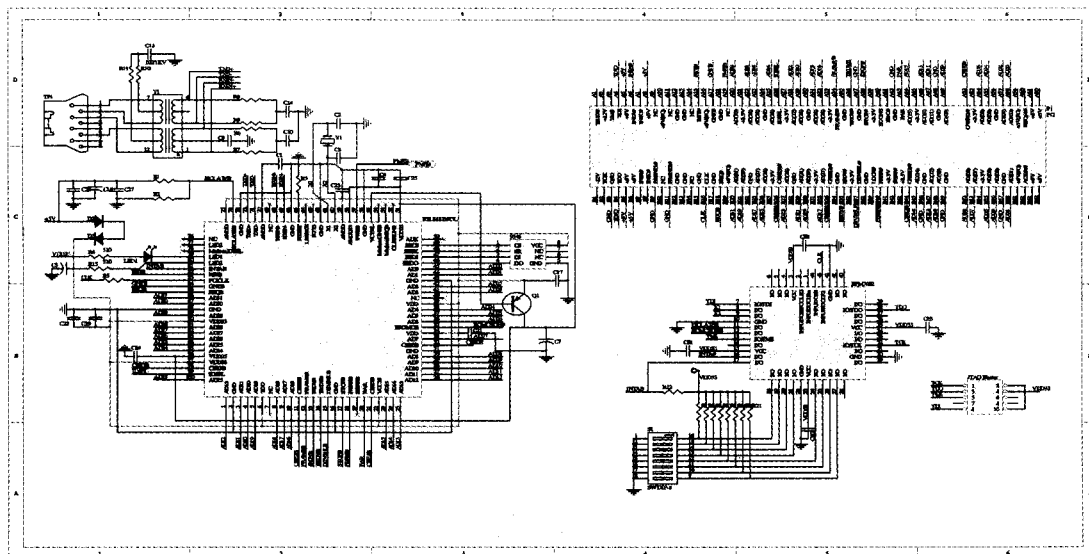


图 2 新网卡原理图

第二个问题是在 F_s 发完后,要实现时间间隙 Δt 延时。由于 RTL8139 符合 802.3 协议,如果按 RTL8139 的原有设计方案,在它发完 F_s 后只延时 96 位时间,就将 F_w 自动变为 F_s ,同时开始将 F_s 发送出去。RTL8139 是通过编程将 F_w 将自动变为 F_s ,为了延时更多时间,实现间隙整形,在编程初始化 RTL8139 时,取消 F_w 变为 F_s 的自动方式,改为程序控制方式。RTL8139 在发完 F_s 的同时,产生一中断请求信号 INTAB 给 PCI 总线,这样主机才能感知写入下一个 F_w ,如果设置一延时电路,将 INTAB 信号延时 Δt 后再送 PCI。当主

机感知到中断,检查 RTL8139 中有 F_w 时,马上将 F_w 变为 F_s ,即可实现间隙整形发送了。

在实际应用中,RTL8139 中 F_w 和 F_s 不是任何时候都存在,比如开始两者都没有,又比如当将 F_w 转为 F_s 时而主机中没有要发送的帧时,就不可以产生 F_w 。这样,在设计驱动程序时,就有必要设置两个标志 MFS 和 MFW。MFS=1 时表示 F_s 存在,MFS=0 时表示 F_s 不存在;MFW=1 时表示 F_w 存在,MFW=0 时表示 F_w 不存在。

(下转第 176 页)

过程也消耗了不少时间。本文的 RCA 算法多出的训练时间是由多次迭代封闭测试以及调整过程所消耗的。虽然调整后训练时间增加了,但能大幅度地提高分类器地分类性能。

结束语 本文提出了根据 ARC-BC 封闭测试的分类结果来调整分类器规则置信度的算法 RCA (Rule Confidence Adjustment),实验结果证明了本文提出的 RCA 算法是有效的,能大幅度地提高分类效果。接下来的工作我们将考虑如何进一步改善关联规则分类器的生成质量,提高挖掘频繁项集的效率。

参考文献

1 Liu B, Hsu W, Ma Y. Integrating Classification and Association

Rule Ming. SIGKDD, New York, 1998
 2 Aaiane O R, ntonie M. Classifying text documents by associating terms with text categories. Thirteenth Australasian Database Conference(ADC'02), Melbourne, Australia, 2002
 3 Li W, Han J, Pei J. CMAR: Accurate and efficient classification based on multiple classification rules. ICDM, San Jose, California, 2001
 4 Rakesh Agrawal Ramakrishnan Srikant. Fast Algorithms for Mining Association Rules. In: Proceedings of the 20th VLDB Conference Santiago, Chile, 1994
 5 HAN Jiawei, PEI Jian, YIN Yiwen. Mining Frequent Patterns Without Candidate Generation. ACM-SIGMOD, Dallas, 2000
 6 Sebastiani F. Machine Learning in Automated Text Categorization. ACM Computing Surveys, 2002, 34(1): 1~47
 7 http://www.nlp.org.cn/docs/download.php? doc_id=1023

(上接第 35 页)

网卡驱动程序在两种情况下对网卡进行发送操作:其一是当 INTAB 发送中断,其二是主机产生了新的发送包。

发生第一种情况时,令 MFS=0;如果 MFW=1,则将 F_w 转为 F_s ,并令 MFW=0, MFS=1;如果主机中有待发帧,则将最高优先的帧写入 RTL8139,并令 MFW=1。

发生第二种情况时,如果 MFW=0,则将最高优先的帧

写入 RTL8139,并令 MFW=1;如果 MFS=0,则执行第一种情况的操作。其它条件时都不做进一步操作。

参照 Realtek 公司提供的 RTL8139 设计 Demo 方案,利用 Altera 公司的 EPLD 芯片 EPM7032 对 INTAB 中断信号实行数字延时,图 2 是满足间隙整形要求的网卡原理图,图 3 是中断数字延时硬件原理图。

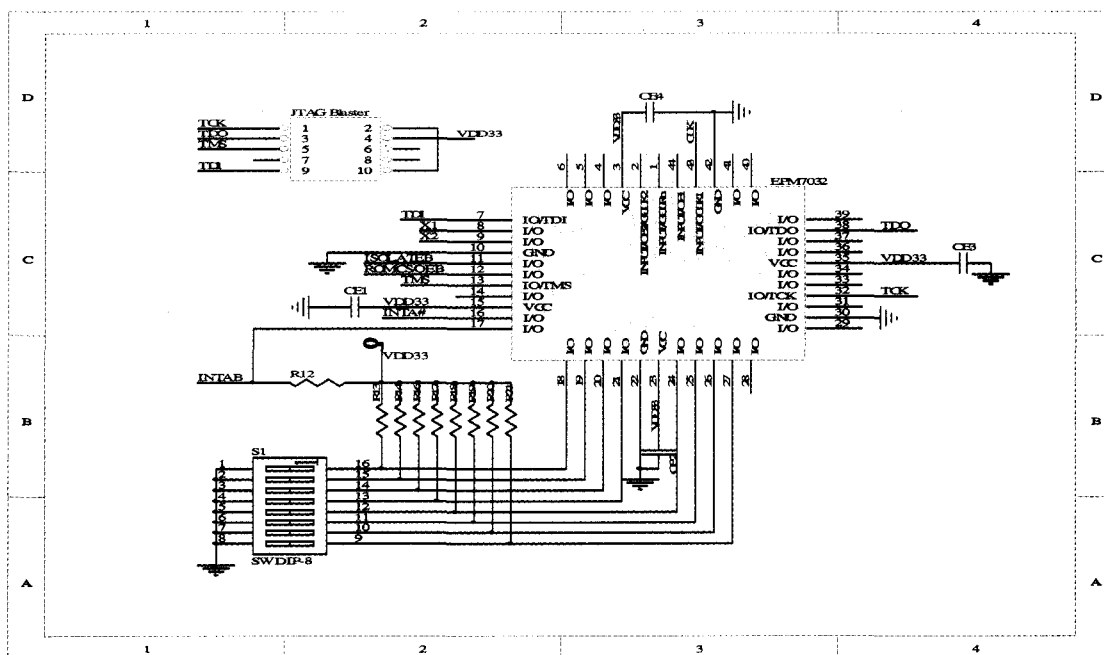


图 3 中断信号处理原理图

实际网络中,下游节点处理包的服务时间 t_s ,与下游节点机器有关,不是一个确定不变的常数值,间隙整形需要下游节点,要标明 t_s 的实际值。由于 t_s 随节点而变,显然网卡对 RTL8139 的中断信号 INTAB 延时也应该是可变的。为了调试使用方便,在网卡上设置了一个配置多选择开关 S1,共有 8 个独立输入,分别为 S(1)、S(2)、...、S(8)。当 S(8)有效时,不延时 INTAB 信号;当 S(7)有效时,INTAB 信号不能到达 PCI 总线上,即永久延时;当 S(7)和 S(8)都无效时,由 S(1)、S(2)、...、S(6)组合可选择 64 种预先设定的延时值之一,实现对 INTAB 延时。

5 实现与调试结果

为了证明间隙调整设计的有效性,我们直接利用 RTL8139 网卡原有驱动程序,使用计算机网络 ping 命令进行测试。

首先断开 S(8),实测传输延时时间与普通 RTL8139 网卡一致。

第二步,接通 S(8),断开 S(7),ping 命令超时,符合设计的功能。

第三步,断开 S(7)和 S(8),S(1)接通,其它的都断开,预设延时值为 $0x8000=32768$,测得 ping 延时为 1.0ms,而普通网卡小于 1.0ms,说明延时有效。

第四步,断开 S(7)和 S(8),其它的也全断开,即预设值为 $0xFC00=64512$,约为第三步预设值的两倍,测得 ping 延时为 1.9ms,延时也约为第三步的两倍。

上述实验结果表明,设计完全达到要求。实际使用时,如果要求的时间更长些,通过修改程序中的预设可选值即可。

参考文献

1 曾华荣. 现代网络通信技术[M]. 成都:西南交通大学出版社, 2004. 450~466
 2 曾家智,徐洁,吴跃,等. 服务元网络体系结构和微通信元系统构架. 电子学报, 2004, 32(5): 745~749
 3 曾家智,赵继东,蓝勇钢. 微通信元系统架构和 QoS. 计算机应用, 2004, 24(6): 10~12