第 33 卷 第 2 期 计 算 机 工 程 2007 年 1 月
Vol.33 No.2 Computer Engineering January 2007
·开发研究与设计技术· 文章编号：1000—3428(2007)02—0264—03 文献标识码：A 中图分类号：TP302.1

嵌入式 PCI 网卡驱动程序的设计与优化

宋有泉，高小鹏，龙 翔
(北京航空航天大学计算机学院，北京 100083)

摘 要：介绍了嵌入式 VPN 网关 ESG-1 的基本情况，讨论了 RTEMS 的 PCI 网卡驱动程序的设计要点：采用服务线程进行网络中断处理；采用生产者－消费者模型对缓冲区进行管理；采用事件驱动机制实现了网卡驱动对多个相同网卡的支持。进一步讨论了驱动程序的内存拷贝优化问题和零拷贝技术。通过测试数据分析得出优化内存拷贝和使用零拷贝技术都能提高网卡驱动程序的性能。

关键词：RTEMS；设备驱动；优化；零拷贝；内存拷贝

Design and Optimization of PCI Ethernet Adapter Driver Program in Embedded System

SONG Youquan, GAO Xiaopeng, LONG Xiang
(School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100083)

【Abstract】 This paper introduces the embedded VPN gateway ESG-1 and discusses the design outlines of device driver program for the PCI Ethernet adapter based on RTEMS. This paper adopts the service threads to dispose the network interrupt, uses the producer and consumer model to implement the buffer management and supports the multitude Ethernet adapters managed by one device driver program through the event-driven technique. The problems of optimizing the memory copy for device driver program and zero copy technique are discussed. After analyzing the test data, it concludes that the performance of Ethernet adapter driver is improved by both optimizing the memory copy and adopting the zero copy technique.

【Key words】 RTEMS; Device driver; Optimization; Zero copy; Memory copy

嵌入式系统是以应用为中心，以计算机技术为基础，能够满足应用对功能、性能、体积、成本、功耗等方面要求的专用系统[1]。嵌入式操作系统是嵌入式应用的基础平台，其性能很大程度上决定了整个嵌入式系统的性能，而嵌入式操作系统的性能很大程度上取决于操作系统内核的结构和驱动程序设计。因此，分析影响设备驱动性能的因素，研究高效的设备驱动设计方法，对于充分挖掘嵌入式系统的性能是很有意义的。

本文以北京航空航天大学嵌入式系统实验室开发的“10Mbps 系列”嵌入式 VPN 网关 ESG-1 为例，研究了嵌入式设备中 PCI 网卡驱动程序的设计与优化过程。

1 背景

1.1 ESG-1 介绍

目前国内的 VPN 设备基本都是采用“工控机+密码单元+通用操作系统+应用软件”的结构。在这种结构中，通用操作系统的稳定性、实时性不高，而且工控机的体系结构导致 VPN 成本比较高。目前，在国内的邮政、银行、医疗系统中，存在许多小型的分支机构使用拨号、DDN 专线等方式与上级机构进行互联。这种低速接入方式带宽一般都小于 10Mbps，成本是用户决定使用 VPN 的关键因素。ESG-1 专门为这种需要而设计，它采用“嵌入式处理器+密码单元+嵌入式操作系统+应用软件”的结构，既能满足用户需要，又能降低成本。图 1 是 VPN 的硬件框图。

ESG-1 硬件板的核心处理单元采用了嵌入式处理器 MPC8245；加解密单元采用加解密算法芯片；核心处理器通过 PCI 总线和两个网卡相连。采用的 PCI 网卡集成了 RTL8139D 单芯片，具有 100M/10Mbps 自适应的能力，遵循 IEEE802.3 和 IEEE802.3u 以太网标准，兼容 PCI2.1/2.2 总线标准。ESG-1 采用的嵌入式操作系统是 RTMES (Real-time Executive for Multiprocessor System)。

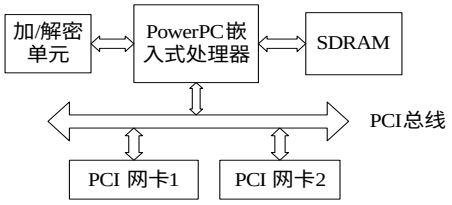

图 1 ESG-1 硬件框图

1.2 RTEMS 介绍

RTMES最初是美国军方为导弹控制领域开发的嵌入式强实时操作系统，后来由美国OAR公司提供后续技术支持。RTEMS是免费的、源码公开的嵌入式操作系统，其部分性能指标甚至超过了最著名的商业实时操作系统VxWorks[2]。RTEMS内核支持对称或者非对称的多处理器系统，采用事件驱动，基于优先级的抢占式任务调度算法，动态分配内存，支持进程间通信和同步等[3,4]。事件（Event）是RTEMS提供的一种高效的同步机制，但不能用于数据传送。

RTEMS 采用了微内核的设计思想，内核只实现了最基本的操作系统元素，内核和用户空间共享相同的地址空间。PCI 设备驱动程序不属于内核空间，而属于用户空间，但是可以

作者简介：宋有泉(1978－)，男，硕士生，主研方向：嵌入式操作系统；高小鹏，讲师；龙 翔，教授、博导

收稿日期：2006-01-29 **E-mail：**syq6buaa@126.com

直接使用内核空间的所有资源。

2 网卡驱动设计

在嵌入式网卡驱动程序设计中，最重要的特征是采用服务线程进行中断处理，以提高网卡对中断响应的实时性。本文采用生产者－消费者模型对缓冲区进行管理，保证网卡驱动有效工作；采用事件驱动机制，实现了用一个驱动程序去管理多个相同网卡的方案。

2.1 服务线程

RTEMS 的任务是基于线程的。线程主要由线程控制块和堆栈寄存器等构成，占用资源少，而且调度、切换、启动的时间都较短。本网卡驱动程序遵循了 RTEMS 任务基于线程的特点，采用服务线程实现网卡的接收和发送过程。

接收服务线程实现网卡接收数据的过程。中断服务程序检查到网卡接收中断后，发送事件给接收服务线程。发送服务线程实现网卡发送数据的过程。当上层软件需要发送数据时，就通过网卡的发送接口函数向发送服务线程发送事件。服务线程接收到事件后，由阻塞状态转变成就绪状态，然后由 RTEMS 的实时调度算法调度执行。

网卡驱动中使用服务线程的优点有：中断服务程序将网卡接收过程转交给接收服务线程，这样加快了中断处理过程，有利于提高嵌入式系统对中断响应的实时性。上层发送任务将网卡发送过程转交给发送服务线程，然后返回进行下一个数据包的发送过程，这样有利于提高网卡的发送效率。

2.2 缓冲区管理设计

RTL8139D 芯片有一个接收缓冲区起始地址寄存器和 4 个发送缓冲区起始地址寄存器，每个缓冲区都有一个状态寄存器。此外，接收缓冲区还有当前读取地址和写入地址寄存器。通过设置网卡的起始地址寄存器可以在内存中为网卡建立一个接收缓冲区和 4 个发送缓冲区。

PCI 网卡驱动程序通过 DMA 方式在网卡片上缓冲区和内存缓冲区之间进行数据传送。内存缓冲区管理是网卡驱动程序设计中很重要的部分，本文采用生产者－消费者模型进行缓冲区管理，缓冲区组织成循环缓冲区结构。如图 2 所示。

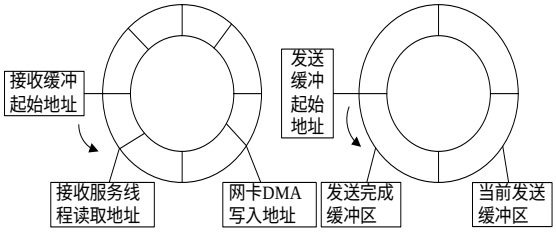

图 2 接收和发送缓冲区的生产者－消费者模型

整个接收缓冲区被设置成了一个循环缓冲区。在接收过程中，网卡 DMA 发送过程是生产者，它按照协议规则不断地向接收缓冲区写入数据帧；接收服务线程是消费者，它按照协议规则不断从接收缓冲区中读取数据帧。读取地址寄存器用来记录接收服务线程读取数据的位置；写入地址寄存器记录了网卡 DMA 写入数据的位置。当网卡检测到写入地址反超读取地址时，就会发出接收缓冲区溢出中断。中断服务程序接收到中断后，进入接收缓冲区溢出处理过程。

发送缓冲区被分成 4 个独立的发送缓冲区，每个发送缓冲区一次只发送一个数据帧，4 个发送缓冲区依次轮循进行数据帧的发送。发送服务线程是生产者，它不断从上层接收数据帧写入发送缓冲区；网卡 DMA 接收过程是消费者，它不断从发送缓冲区中读取数据帧发送到网络。发送服务线程往发送缓冲区写入一帧，当前发送缓冲区向前移动一个缓冲区；网卡每发送完一帧后，会发出发送完成中断，中断服务程序通过查询发送状态寄存器就可以知道当前发送完成的缓冲区。如果当前发送缓冲区反超发送完成缓冲区，则发送缓冲区溢出。当发送缓冲区溢出时，发送服务线程必须等待网卡 DMA 接收过程；当有可用的发送缓冲区时，发送服务线程就会被唤醒进行数据发送过程。

采用生产者－消费者模型进行缓冲区管理，解决了网卡 DMA 和网卡服务线程之间的同步问题。

2.3 支持多网卡设计

支持多网卡，是指用一个网卡驱动程序来管理多个网卡。本网卡驱动程序为了减少线程调度和切换的开销，所有的网卡共享一个接收服务线程和一个发送服务线程。

RTEMS提供了多种线程间通信机制，但是适合于多网卡驱动程序的只有两种机制：消息机制和事件机制。消息机制是在通信线程之外建立一个消息队列，通过传递不同的消息来达到线程间通信的目的。消息机制因为发送方和接收方都必须对外部消息队列进行操作，所以效率比较低。事件机制是直接操作线程控制块的事件列表来传递信息，所以这是一种高效的同步方式[4]。RTEMS中每个线程最多可以有 32 个事件，不同事件是可以叠加的，接收方可以对事件值进行解析，从而获得多个叠加在一起的事件。通过分析可知，消息机制的优点在于线程间传递数据；事件机制的优点在于线程间传递同步信息；执行任务或中断服务例程和服务线程之间的通信主要是同步信息，而不是数据信息。因此，本文驱动程序采用事件机制实现对多网卡的支持。

每个网卡具有发送和接收两种服务，因此可以为每个网卡设置不同值的事件。当执行任务需要网卡服务时，就可以通过发送该网卡定义的事件给服务线程。服务线程接收到事件后，通过对事件值进行解析，就可以判断出请求服务的网卡，然后提供该网卡的服务。

3 性能优化

本文首先为 RTEMS 操作系统设计并实现了 RTL8139D 网卡驱动；然后对内存拷贝部分进行优化，提高了网卡性能；最后在网卡驱动程序中采用了零拷贝技术，进一步提高了网卡的性能。

3.1 优化 memcpy

在ESG-1 中，MPC8245 和内存之间数据总线是 32 bits。在RTL8139D网卡驱动中，每次数据发送或接收都存在缓冲区和mbuf之间的内存拷贝过程，RTEMS中提供了memcpy函数进行内存数据拷贝[4]。memcpy在数据拷贝时，如果源地址和目标地址都是 4B的整数倍时，按字方式进行内存拷贝，那么一次可以拷贝 4B；否则就按字节方式进行内存拷贝，一次只能拷贝一个字节，这样内存拷贝的速度将大幅下降。

为了提高网卡性能，驱动程序能够保证接收缓冲和发送缓冲的起始地址是 4B 的整数倍。但是在 RTEMS 的 TCP/IP 协议栈中，分配的 mbuf 的数据块起始地址是 2 的整数倍，不一定是 4 的整数倍。因此，驱动中使用 memcpy 函数进行内存拷贝并不能发挥出网卡的最大性能。为了提高网卡驱动性能，本文对驱动的内存拷贝过程进行了优化，其主要思想是：如果目的地址和源地址都是 4B 的整数倍时，按 4B 对齐进行拷贝；否则按 2B 对齐进行拷贝。这样设计的好处是：当地址不是 4B 整数倍时，也还是能够一次拷贝 2B，效率比 memcpy 高出一倍。

3.2 零拷贝

零拷贝技术可以减少数据拷贝次数和共享总线操作的次数，消除通信数据在存储器之间不必要的中间拷贝过程，有效地提高了通信效率[5]。

RTL8139D 的 PCI 网卡工作在主设备模式，通过 DMA 可以直接访问内存。该网卡芯片只支持一个接收缓冲区，在 DMA 过程中，是根据前一个数据帧的结束位置决定下一个数据帧的存放起始位置，而不是由驱动决定下一个数据帧存放起始位置，因此接收过程很难使用零拷贝技术。网卡发送缓冲分为 4 个独立的缓冲区，缓冲区的位置可以由驱动程序动态指定，所以网卡发送过程可以使用零拷贝技术。其主要思想是，将发送缓冲区起始地址寄存器直接指向发送数据块，然后启动网卡 DMA 过程，这样去掉了将数据从 mbuf 拷贝到接收缓冲的过程。当网卡发送数据完成后，通过中断将发送数据内存块回收。

要实现零拷贝，必须将发送的数据放置到一块连续的内存区，但是 RTEMS 的 TCP/IP 协议栈把数据部分分开放置，因此还必须对 RTEMS 的 TCP/IP 协议栈进行一些修改，使得每个数据帧存放在的一块连续的内存区且起始地址是 4 的整数倍。

4 实验数据与分析

VPN 硬件平台包括处理器 MPC8245、128MB 内存。处理器时钟频率是 266MHz，系统总线时钟频率是 133MHz。软件平台是嵌入式操作系统 RTEMS 4.6。

实验过程：以 VPN 作为网关，从一个子网络节点向另一个子网节点发送 TCP 数据包，并且改变发送数据包的大小。测试结果如图 3 所示。

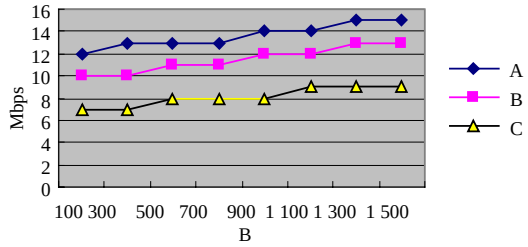

图 3 发送包大小和速度曲线

A 曲线是在驱动程序采用了零拷贝的情况下，进行 TCP 数据传输的性能曲线。B 曲线是在对网卡驱动程序进行了 memcpy 优化的情况下，进行 TCP 数据传输的性能曲线。C 曲线是在没有对网卡驱动作优化的情况下，进行 TCP 数据传输的性能曲线。

A、B 曲线存在差异的主要原因是：A 在网卡发送数据时，直接将网卡发送缓冲指向发送数据块，减少了数据的内存拷贝过程。B 则是将数据块先拷贝到分配好的缓冲区，然后再启动网卡发送。

B、C 曲线存在差异的主要原因是：在内存拷贝过程中，如果地址存在非 4B 的整数倍时，B 采用按 2B 对齐进行内存拷贝，而 C 采用的是系统提供的 memcpy 函数进行内存拷贝。在这种情况下，memcpy 实际上是按单字节进行拷贝的。

通以上过分析，可以得出影响网卡设备驱动性能的重要因素是：内存访问速度。提高内存拷贝速度其它途径有：(1)减少读写切换次数，例如连续从内存读取字节数和 SDRAM 的一个 burst 周期字节数一致。(2)使用 cache，本文是在只启动了指令 cache 的情况下进行测试的，而在数据 cache 也启动的情况下，网卡驱动性能会更高。

5 结论

本文基于嵌入式 VPN 目标板，为 RTEMS 操作系统设计开发了 RTL8139D PCI 网卡的驱动程序。在驱动设计中，采用服务线程进行网络中断处理；采用生产者－消费者模型对缓冲区进行管理；采用事件驱动机制实现了网卡驱动对多个相同网卡的支持。为了提高网卡性能，本文先对驱动程序的内存拷贝进行了优化，提高了网卡性能；然后在网卡发送过程中采用内存零拷贝技术进一步提高了网卡性能。在嵌入式 VPN 平台上进行测试，经过优化后的 PCI 网卡驱动程序运行稳定、性能较高，达到了设计要求。

参考文献

1 杜　旭，顿新平，黄　建. 一种嵌入式系统驱动架构的分析及实现[J]. 计算机工程与应用, 2004: 40(25): 116.

2 Straumann T. Open Source Real Time Operating Systems Overview[C]. Proc. of the 8th International Conference on Accelerator & Large Experimental Physics Control System, San Jose, California, 2001.

3 RTEMS C User's Guide[R]. OAR Corp., 2000.

4 BSP and Device Driver Development Guide[R]. OAR Corp., 2000.

5 可向民，龚正虎，夏建东. 零拷贝技术及其实现的研究[J]. 计算机工程与科学, 2000, 22(5): 18.

(上接第 258 页)

根据集团企业信息化的总体规划，财务信息化的下一个目标是实现财务数据与业务数据(生产、物流、人力资源等)的一体化，从而打破信息孤岛，达到局部与整体、财务与业务处理之间的高度协调一致。

集中式集团财务信息系统的成功实施，使武钢在经营活动全面核算的基础上，通过全面预算、集中资金管理、财务分析等细化分析管理和监控方式，对企业集团经营活动的成果进行深入分析，及时发现企业集团在资金运作、成本费用控制等方面的缺陷，在日益激烈的市场竞争环境中充分发挥财务监管的作用，为企业的决策提供支持。

参考文献

1 毛蕴诗，李新家，彭清华. 企业集团：扩展动因、模式与案例[M]. 广州：广东人民出版社, 2001.

2 刘俊勇. 企业集团财务信息化问题研究[J]. 中国会计电算化, 2003, (5): 10-13.

3 陈新忠. 浅谈企业集团财务控制的完善[J]. 财务与会计, 2004, (4): 16-17.

4 浦　颖. 企业集团财务管理与资金控制[J]. 农资科技, 2004, (2): 36-39.