

双网实时切换的以太网控制器 IP 核

方湘艳^{1,2}, 陈敬东², 熊庭刚², 袁由光²

(1. 哈尔滨工程大学计算机学院, 哈尔滨 150001; 2. 中国船舶重工集团公司第 709 研究所, 武汉 430074)

摘要: 针对高可靠性网络通信中网卡的双冗余备份技术存在网络切换延时问题, 提出一种在网络控制芯片内完成网络切换的方法。通过网络链路层对接收的数据进行协议解析, 判断网络通道和接收数据的正确性进行网络切换。实现在网络通道出现故障或接收数据出错时实时切换到正常网络通道的功能。

关键词: 实时切换; 数据缓存区; 状态机

Ethernet-controllers IP Core with Real-time Switch between Net

FANG Xiang-yan^{1,2}, CHEN Jing-dong², XIONG Ting-gang², YUAN You-guang²

(1. School of Computer, Harbin Engineering University, Harbin 150001;

2. The 709th Research Institute, China Shipbuilding Industry Corporation, Wuhan 430074)

【Abstract】 The paper presents the method to complete network switch in chip of network controller in allusion to network switch delay in the redundancy backup technology of the network cards in the network communications. The network channels is switched according to the state of network channels working and data receiving circuit when datum received are decoded in network link layer. This implements the function of real-time switch when network equipment is trouble in physics layer or datum received is wrong.

【Key words】 real-time switch; data buffer; state machine

1 概述

随着网络技术的成熟, 以太网已成为各种控制系统接口互连的主要媒介。在银行、政府、军方等特殊的应用场合, 对网络系统的稳定性要求非常高, 为提高系统的可靠性和抗毁性, 必须采用双冗余网络技术。

目前, 市场上主要通过在每个网络节点采用双网卡, 中间用 2 个 HUB 或交换机互连, 实现网络冗余备份功能。当某个节点的一块网卡出现故障或网线损坏或其中一个 HUB 或交换机故障时, 系统马上启用另一块冗余的网卡使网络仍能正常运作。因此, 网络切换的速度对网络系统的正常工作非常重要。双网卡的网络切换主要通过应用软件来实现, 切换速度较慢。目前针对具体的网卡已开发出相应网络切换驱动程序, 提高网络切换速度, 但双冗余网络备份主要针对网络系统中容易出错的物理层设备进行冗余备份。因此, 在驱动程序里进行网络切换相对网络系统的出错点还有一定延时。

本文提出在一种在网络控制器内完成双网切换的方法, 并同过微电子 IP 技术在一款芯片里实现, 不但保证了物理层设备出现故障时能进行实时切换, 而且在网络接收过程中数据包出错时也能进行实时切换。

2 双网实时切换的以太网控制器的关键技术

双网实时切换的以太网控制器主要完成 IEEE 帧和以太帧^[1]数据包的发送、接收, 物理层器件的配置以及当某一个物理器件出现故障时的网络切换功能。其帧结构^[2-3]包括: 帧前序(Preamble), 帧起始符(SFD), MAC 目的地址与源地址, 帧长度/类型(L/T), 客户数据(Data), 填充区(Pad), 帧校验序列(FCS)。

2.1 以太网的流量控制

以太网流量控制是解决网络由于负载过重或过于失衡而导致网络拥塞甚至造成瘫痪的主要手段^[4]。当发送节点的接收缓存器超过设置的最低门限值时, 将发送具有最大延迟的暂停帧到远端发送节点, 远端发送节点在收到暂停帧后, 立即停止下一帧数据发送, 其延迟时间的大小根据收到的暂停帧的控制参数决定。当接收节点的缓存器低于设置的最低门限值时, 发送一个具有零延迟的暂停帧给对方, 发送节点在收到暂停帧后, 有重新向对方发送数据。由于暂停帧的传输具有不确定的延时, 因此缓存器的最低门限值的设置非常关键, 设置太小可能会引起丢包。

2.2 实时切换的原理

本文提出的实时切换方案解决了 3 方面的问题:

(1) 保证了在某个物理层设备出现故障后, 网络能正常工作。(2) 当 2 个网络通道在传输数据时, 当一个通道的数据在物理层传输中出现错误, 则切换到另一个通道上, 保证数据能正确接收。(3) 在网络通道切换过程中, 由于采取了双通道数据备份, 因此在实时切换过程中不会丢包。

网络切换必须与网络数据帧的接收同步。每个网络通道都设置有各自的接收缓存区。在正常接收数据帧时, 网络通

基金项目: “十一五” 船舶基金资助项目(08J3.7.4)

作者简介: 方湘艳(1975 -), 男, 博士研究生, 主研方向: 计算机容错, 微电子设计; 陈敬东、熊庭刚, 研究员; 袁由光, 研究员、博士生导师

收稿日期: 2009-02-06 **E-mail:** fangxy_whu@126.com

道 1 和网络通道 2 分别将数据存入接收缓存区 1 和接收缓存区 2，并开始设置一个默认接收缓存区，如接收缓存区 2 为上电后 PCI 读取数据的默认缓存区。上层应用程序通过 PCI 总线从这个默认接收缓存区 2 中读取以太网帧数据。当某个网络通道出现故障后，如果这个通道不是默认接收缓存区 2 对应的通道，即网络通道 1 出现故障，则正常接收数据，如果是网络通道 2 出现故障，立即将另一个缓存区 1 设置为默认接收缓存区，即将 PCI 总线从原来缓存区 2 读取数据切换到现在的缓存区 1 读取数据。在 2 个网络通道正常运行情况下，如果当前的默认接收缓存区 2 对应的通道在接收数据出现 CRC 检验错误时，则立即将另一个缓存区 1 设置为默认接收缓存区，直到原来缓存区 2 对应的通道上数据接收正确后再恢复缓存区 2 为默认接收缓存区。

3 双网实时切换的以太网控制器 IP 核的实现

电子系统的设计一般有自下而上和自上而下 2 种方法。自下而上的设计方法是在系统划分和分解的基础上先进行基础单元设计，逐步向上进行功能设计，然后进行系统设计。常用于原理图的设计中。自上而下的设计方法是首先进行行为设计，接着进行结构设计，完成后将结构转换为逻辑图，即进行逻辑设计，最后完成整个电路设计，一般适用于 HDL 语言的设计中。本设计采用自上而下的设计方法。整个芯片分为 4 个模块：PCI 总线控制模块，以太网控制模块 1，以太网控制模块 2，实时切换模块。其系统结构如图 1 所示。

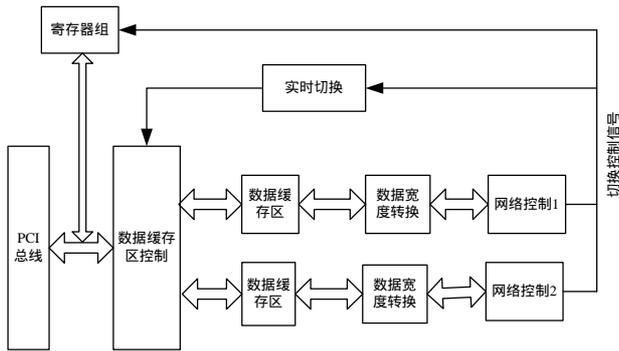


图 1 双网实时切换系统结构

3.1 PCI 总线控制模块

PCI 总线控制模块完成 PCI 协议的功能，主要包括 PCI 配置空间设置、PCI 总线的交易控制、地址译码与解码、命令译码以及 PCI 突发数据传输功能。

PCI 总线的控制交易分为读写 2 种操作，每种操作含有 6 种状态：idle、dev_resp、object_wait1、object_wait2、object_data、stop_resp。其中，idle 是系统的缺省状态，表示总线当前空闲。在 idle 状态时，要检测来自 PCI 总线的信号，如果是访问本 PCI 设备则自动进入 dev_resp 状态；在 dev_resp 状态时，如果检测到 IRDY 信号有效则驱动 DEVSEL 为低电平，表示 PCI 设备已经成为当前访问的目标设备，并自动进入 object_wait1 状态，插入等待周期，一直等到数据读写交易完成才进入 object_wait2 状态；在 object_wait2 状态，检测本次数据交易是单周期交易还是多周期交易，如果是单周期交易则进入 object_data 状态，如果是多周期交易则进入 stop_resp 状态；在 stop_resp 状态产生有效的 STOP 信号，向主设备申请交易断开，同时进入 object_data 状态，产生有效的 TRDY 信号，这时 IRDY 和 TRDY2 2 个信号同时有效，数据交易完成，最后又回到 idle 状态。从模式下 PCI 状态机结构见图 2。

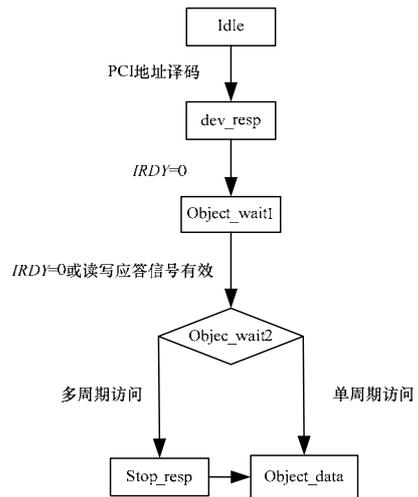


图 2 从模式下 PCI 交易的状态机

当 CPU 进行网络通信时，通过 PCI 总线进行数据突发传输。在缺省状态下，网络控制器处于空闲状态(idle)，当检测到突发传输使能寄存器有效时，发出突发请求；在 CPU 完成请求响应后，PCI 总线由 MAC 控制器控制，CPU 只是作为从设备响应 PCI 的读写请求。状态机进入数据传输准备状态(wr_trans1)，接着进入突发传输状态(wr_trans2)，如果传输的是最后一个数据，则转入状态 lasttran，如果数据缓存区已经为空，则进入状态 trandone，如果数据大小计数器不为 0，则继续保持在状态 wr_trans2 下传输数据。PCI 突发数据传输状态机如图 3 所示。

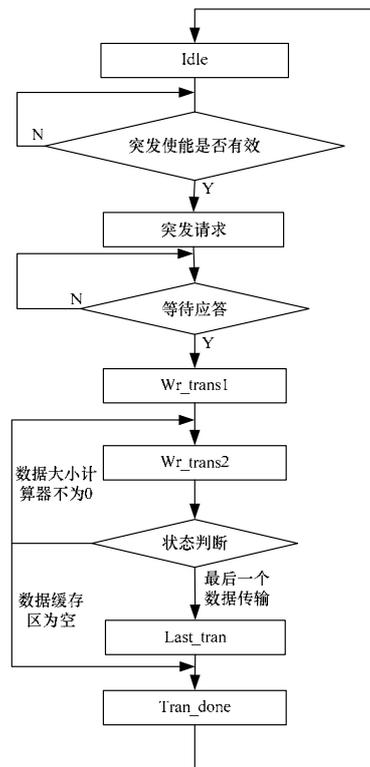


图 3 PCI 突发数据传输状态机

3.2 实时切换模块

本模块完成 2 个功能：(1)将高速数据和低速数据处理模块进行隔离；(2)对出现故障的网络通道进行实时切换，保证网络的正常工作。

PCI 总线的数据传输速度为 33 MHz，数据最大宽度为 32 bit，而以太网控制器的处理速度只有 25 MHz，数据宽度只有 4 bit。因此，采用异步 FIFO 来将 2 个不同处理数据的模块隔离，PCI 模块和控制器模块分别通过 FIFO 的满、空标志位来存取数据。当发送数据时，由于 PCI 的速率比控制器快，因此不会影响网络发送数据的速率，PCI 在将一帧数据写入到 FIFO 后，就可以释放总线，在控制器处理完后再次申请 PCI 总线传送下一帧数据。因此，发送 FIFO 不用设计太大，但不能低于一帧数据的最大数据量，即不能低于 1.5 K。在接收数据时，为了防止接收缓存区溢出，接收 FIFO 应尽量设计大些，对于判断是否发送暂停帧的 FIFO 最低门限值也应设计大些，当接收 FIFO 为空时再发送暂停帧，一定会导致 FIFO 溢出，从而出现丢包。

实时切换主要在 PCI 读取数据时，根据接收模块产生的控制信号对 PCI 读取接收 FIFO 的数据进行切换。如图 3 所示，因为接收模块输出的数据经过接收 FIFO 缓存，然后再由切换模块送入 PCI 总线，而接收模块产生的切换控制信号直接进入切换模块。因此，控制信号比数据信号提前进入切换模块，实现实时切换的功能。

3.3 以太网控制器模块

本模块通过采用 2 个状态机来分别完成以太网收发控制器的功能，如图 4 所示。

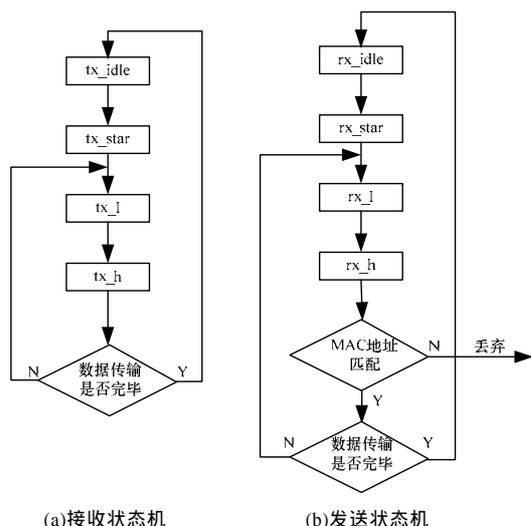


图 4 以太网控制器收发状态机

发送模块从发送 FIFO 中读取数据，将 12 字节的目标地址、源地址和 2 字节的长度/类型数据保存到寄存器中，并根据长度/类型数据来判断该帧是以太网帧还是 IEEE 帧，以及数据填充区是否需要填充数据。由于从发送 FIFO 每读出一个字节的数据，需要 2 个时钟周期发送出去。因此，在读出一帧数据的前 14 个字节数据的同时，完成 7 个字节的帧前序发送。在发送第 8 个字节帧起始符时，长度/类型数据已经从 FIFO 中读出，可以判断出该帧类型，并根据其帧类型确定帧起始符为“10101010”或“10101011”。接收状态机如图 4(a)所示。其中，tx_idle 为空闲状态，当从发送 FIFO 中有数据读出，进入 tx_star 状态。在 tx_star 状态，完成 7 字节的帧前序发送，同时 12 字节 MAC 地址和 2 字节的长度/类型数据已经保存在寄存器中，然后进入帧数据的发送状态。tx_l 状态完成 1 个字节的低 4 位发送。tx_h 状态完成 1 个字的高 4 位发送，并完成数据填充区的数据发送和 CRC 帧数据校

验的发送。当一帧数据发送完后，等待不少于 96 ns 的帧间隙时间后，返回到 tx_idle 空闲状态等待下一帧数据的发送。

接收模块从物理层器件接收数据，同发送模块一样，将 12 字节的目标地址、源地址和 2 字节的长度/类型数据保存到寄存器中，根据长度/类型数据来判断帧类型和长度。并根据帧长度过滤掉填充区的数据和 CRC 校验的数据，并将过滤后的数据写入到接收 FIFO 中，同时对接收帧数据进行 CRC 校验，判断接收的数据是否正确，将其判断结果和接收通道是否出现故障作为切换的控制信号送给实时切换模块。其状态机如图 4(b)所示。在 rx_idle 空闲状态下，如果接收数据使能信号有效，则接收帧起始符的低 4 位数据，在没有发生错误的情况下进入 rx_sfd 状态接收帧起始符的高 4 位数据，如果接收数据正确，则准备接收以太网帧数据。一个字节的两个状态接收，rx_l 状态接收字节的低 4 位数据，rx_h 状态接收字节的高 4 位数据，同时根据长度信息过滤掉填充数据和 CRC 校验数据。一帧数据接收完后返回 rx_idle 状态，准备接收下一帧数据。

4 时序仿真波形

下文通过 Modesim 工具对实时切换的以太网控制器 IP 核进行功能仿真。其中，RX_CLK1 为接收时钟信号；RXFIFO_DATA 为接收 FIFO 输出到 PCI 模块的数据；RXFIFO_DATA1 和 RXFIFO_DATA2 分别为通道 1 和通道 2 缓存区的输出数据；redundance 网络通道的实时切换控制信号，高电平有效，默认接收网络通道 2 的数据，当网络通道 2 出现故障时切换到通道 1；RXDV 为通道上数据使能信号，高电平有效；RX_ER 为通道上数据出错信号，高电平表示数据出错。RXD 为通道数据信号。

图 5 是对网络物理层设备出现故障时的实时切换仿真波形。以波形中一根竖线为标志，左边是正常情况下的接收状态，这时 2 个通道上的 redundance 都为高电平，默认接收通道 2 的数据。在竖线右边，通道 2 出现故障，使 RXDV 信号无效，以太网控制器无法接收通道 2 上的数据，因此，通道 2 缓存区输出的数据为出错前的数据。redundance2 变为低电平，控制器自动实时切换到从通道 1 的缓存区获取数据。

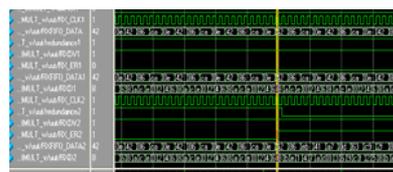


图 5 接收过程中物理层设备出现故障的实时切换

图 6 是对网络通道数据出现错误时的实时切换仿真波形。以波形中一根竖线为标志，左边是正常情况下的接收状态，这时 2 个通道上的 redundance 都为高电平，默认接收通道 2 的数据。在竖线右边，RX_ER2 为高电平，标志通道 2 上传输的数据出现错误，redundance2 变为低电平，控制器自动实时切换到从通道 1 的缓存区获取数据。

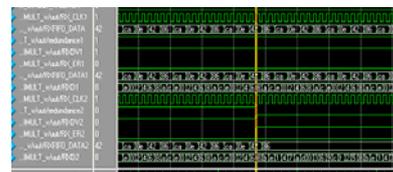


图 6 接收过程中数据出现错误的实时切换

(下转第 236 页)