

(12) 发明专利申请

(10) 申请公布号 CN 102185770 A

(43) 申请公布日 2011. 09. 14

(21) 申请号 201110115194. 6

(22) 申请日 2011. 05. 05

(71) 申请人 汉柏科技有限公司

地址 300384 天津市华苑产业区海泰西路
18号西3楼104室

(72) 发明人 殷建儒 王琳 王博

(74) 专利代理机构 北京路浩知识产权代理有限
公司 11002

代理人 王莹

(51) Int. Cl.

H04L 12/56(2006. 01)

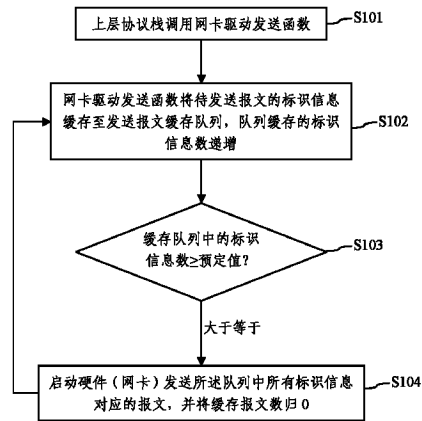
权利要求书 1 页 说明书 4 页 附图 2 页

(54) 发明名称

基于多核架构的批量式报文收发方法

(57) 摘要

本发明公开了一种基于多核架构的批量式报文收发方法,涉及数据通信技术领域,包括:S1:上层协议栈调用网卡驱动发送函数;S2:该函数将待发送报文的标识信息缓存至发送报文缓存队列,队列中的标识信息数递增,并启动定时器;S3:网卡驱动发送函数检测所述队列中的标识信息数,若标识信息数≥预定值,则转步骤S4,否则退出网卡驱动发送函数,转步骤S5;S4:发送队列中缓存的标识信息对应的报文,并将缓存的标识信息数归0,停止定时器;S5:判断定时器是否超时,若是,则转步骤S4,否则循环等待。本发明通过批量发送报文,减少了网卡发送争夺总线的次数,从而增加了网卡接收对总线的使用几率,提高了系统的报文接收能力和系统吞吐量。



1. 一种基于多核架构的批量式报文收发方法,其特征在于,包括以下步骤:

S1:上层协议栈调用网卡驱动发送函数;

S2:所述网卡驱动发送函数将待发送报文的标识信息缓存至发送报文缓存队列,所述发送报文缓存队列缓存的标识信息数递增,所述标识信息包括:待发送报文的地址和报文描述信息的指针;

S3:所述网卡驱动发送函数检测所述发送报文缓存队列中的标识信息数,若缓存的标识信息数大于等于预定值,则转步骤 S4, 否则不发送;

S4:发送所述发送报文缓存队列中缓存的所有标识信息对应的待发送报文,并将缓存的标识信息数归 0, 转步骤 S2。

2. 如权利要求 1 所述的基于多核架构的批量式报文收发方法,其特征在于,所述步骤 S2 中当接收到第一个报文的标识信息时,还包括启动定时器的步骤;

所述步骤 S3 中若发送报文缓存队列中缓存的标识信息数小于所述预定值,且定时器超时,则转步骤 S4, 否则循环等待;

所述步骤 S4 发送完报文缓存队列中缓存的标识信息对应的报文后,复位定时器。

3. 如权利要求 1 或 2 所述的基于多核架构的批量式报文收发方法,其特征在于,执行步骤 S2 时,若不存在发送报文缓存队列,则先建立发送报文缓存队列,并将缓存的标识信息数设置为 0。

4. 如权利要求 1 ~ 3 中任一项所述的基于多核架构的批量式报文收发方法,其特征在于,所述发送报文缓存队列大小为 1 ~ n 个待发送报文的标识信息的总大小,所述预定值小于等于发送报文缓存队列大小。

5. 如权利要求 4 所述的基于多核架构的批量式报文收发方法,其特征在于,所述 n 为 1 到 256 中任意为 2 的幂的值。

基于多核架构的批量式报文收发方法

技术领域

[0001] 本发明涉及数据通信技术领域,特别涉及一种基于多核架构的批量式报文收发方法。

背景技术

[0002] 数据通信设备以数据报文的接收、处理、发送为主要处理流程,报文的接收能力、处理能力、发送能力综合决定了系统的吞吐量,而且具备木桶效应,最短的板决定了系统吞吐量。在单核系统下,通常处理能力成为短板。多核系统与单核系统相比提高了整个系统的处理能力,并且同时可以支持多个网卡。

[0003] 多核系统的吞吐量也由接收能力、处理能力、发送能力综合决定,其短板为报文的接收能力,即系统吞吐量由接收能力决定。目前,多核系统中报文逐个发送,每个发送请求到达网卡,都会增加网卡试图争夺总线的意图。以下是影响报文接收的因素:

[0004] 1、报文从网卡硬件缓存区拷贝到内存;

[0005] 2、报文发送对内存总线带宽的争夺;

[0006] 3、对内存缓存区报文的及时分发;

[0007] 4、从内核态内存拷贝到用户态应用程序内存。

[0008] 因此,要提高多核系统的系统吞吐量必须提高多核系统的报文接收能力。

[0009] 如何提高报文接收能力成为提升系统吞吐量的关键。

[0010] 主要分析 1~3 点,第 4 点在不同的操作系统和系统实现上有不同表现,有些操作系统根本不区分内核态内存和用户态内存,即使在区分内核态和用户态的系统上,也有方法避免报文从内核态拷贝到用户态,而且对数通产品协议栈通常都是自定义实现,不必依赖于操作系统协议栈,所以第 4 点不是问题,不影响分析。

[0011] 1、进入网卡硬件缓存区的报文只有经过 DMA (Direct Memory Access) 接口存入内存后才能被软件处理,所以这是提高接收能力的第一个关键点,影响该点的因素主要是其他元素对内存总线带宽的争夺。报文发送是对内存总线带宽的最有力争夺者。所以降低报文发送对内存总线带宽的争夺是提高接收能力的第一个任务。

[0012] 2、和第 1 点的描述类似。

[0013] 3、只有对内存区已缓存的报文及时清理掉,后续报文才能进入。影响该点的因素涉及到软件效率,线程调度,缓存区大小等等。实验证明,在缓存区一定,软件效率一定情况下,尽可能多的调度接收线程,让接收线程占有更多的 cpu 时间仍然无法提高接收能力,因为报文是批量进入内存的,线程空闲时没有报文,报文来了又可能无法及时处理(没有及时调度,或者调度了但是批量太大),造成报文累积,更大的缓存区也有溢出的时候。所以提高接收软件的效率是必须的,但是当效率提升总有尽头,关键是一次进入内存的报文批量相对于接收能力太大。所以适量减少网卡一次进入内存的报文批量是提高接收能力的第二个任务。

发明内容

[0014] (一) 要解决的技术问题

[0015] 本发明要解决的技术问题是：如何提高报文接收能力以提升系统吞吐量。

[0016] (二) 技术方案

[0017] 为解决上述技术问题，本发明提供了一种基于多核架构的批量式报文收发方法，包括以下步骤：

[0018] S1：上层协议栈调用网卡驱动发送函数；

[0019] S2：所述网卡驱动发送函数将待发送报文的标识信息缓存至发送报文缓存队列，所述发送报文缓存队列缓存的标识信息数递增，所述标识信息包括：待发送报文的地址和报文描述信息的指针；

[0020] S3：所述网卡驱动发送函数检测所述发送报文缓存队列中的标识信息数，若缓存的标识信息数大于等于预定值，则转步骤 S4，否则不发送；

[0021] S4：发送所述发送报文缓存队列中缓存的所有标识信息对应的待发送报文，并将缓存的标识信息数归 0，转步骤 S2。

[0022] 其中，所述步骤 S2 中当接收到第一个报文的标识信息时，还包括启动定时器的步骤；

[0023] 所述步骤 S3 中若发送报文缓存队列中缓存的标识信息数小于所述预定值，且定时器超时，则转步骤 S4，否则循环等待；

[0024] 所述步骤 S4 发送完报文缓存队列中缓存的标识信息对应的报文后，复位定时器。

[0025] 其中，执行步骤 S2 时，若不存在发送报文缓存队列，则先建立发送报文缓存队列，并将缓存的标识信息数设置为 0。

[0026] 其中，所述发送报文缓存队列大小为 1 ~ n 个待发送报文的标识信息的总大小，所述预定值小于等于发送报文缓存队列大小。

[0027] 其中，所述 n 为 1 到 256 中任意为 2 的幂的值。

[0028] (三) 有益效果

[0029] 本发明通过对多核系统中的发送报文采取批量发送的方式，一次性批量映射到网卡，与每次发送一个报文相比，批量发送报文减少了网卡发送争夺总线的次数，从而增加了网卡接收对总线的使用几率，提高了系统的报文接收能力和系统吞吐量。

附图说明

[0030] 图 1 是本发明实施例的一种基于多核架构的批量式报文收发方法流程图；

[0031] 图 2 是本发明实施例的另一种基于多核架构的批量式报文收发方法流程图。

具体实施方式

[0032] 下面结合附图和实施例，对本发明的具体实施方式作进一步详细描述。以下实施例用于说明本发明，但不用来限制本发明的范围。

[0033] 实施例 1

[0034] 本发明方法的实现位于以太网卡驱动程序，需要修改网卡驱动程序。首先本文方法在网卡驱动初始化流程末，建立一个发送报文缓存队列，用于缓存协议栈请求发送报文

的标识信息（具体指：待发送报文的地址和报文描述信息的指针）。修改网卡驱动程序发送函数，将直接发送改为有条件发送，条件 1：发送报文缓存队列中已经缓存了预定量的待发送报文的标识信息，这个预定量根据具体系统的试验确定，通常可以选择 16, 64, 128 等。

[0035] 如图 1 所示，本发明的基于多核架构的批量式报文收发方法包括：

[0036] 步骤 S201，上层协议栈调用网卡驱动发送函数。

[0037] 步骤 S202，网卡驱动发送函数将待发送报文的标识信息缓存至发送报文缓存队列，发送报文缓存队列缓存的标识信息数递增。在此之前，若未创建发送报文缓存队列，则创建发送报文缓存队列，其大小为 1 ~ n 个待发送报文的标识信息的总大小，并将缓存的标识信息数设置为 0，n 优选取值为：1 到 256 中任意为 2 的幂的值。

[0038] 步骤 S103，网卡驱动发送函数检测所述发送报文缓存队列中的标识信息数，若缓存的标识信息数大于等于预定值，则转步骤 S4，否则不发送；

[0039] 步骤 S104，发送所述发送报文缓存队列中缓存的所有标识信息对应的待发送报文，并将缓存的标识信息数归 0，转步骤 S2。

[0040] 本发明增加发送批量的大小，将上层软件请求发送的报文的标识信息入队，积攒到一定数目的待发报文，一次性批量映射到网卡，因为几乎所有常用网卡都支持批量发送。与每次发送一个报文相比，批量发送减少了网卡发送争夺总线的次数，当然这是建立在网卡发送事实上的高效率上。这样自然增加了网卡接收对总线的使用几率。网卡接收是从线路上将报文比特流缓存到卡上缓存中，然后等待总线空闲时机将缓存的报文批量 DMA 到内存。

[0041] 实施例 2

[0042] 进一步地，发送时满足以下条件 2：

[0043] 条件 2：定时器超时，该定时器的作用是，避免长时间没有发送请求，条件 1 长时间无法满足，导致已经缓存在队列中的标识信息对应的报文无法发送。

[0044] 如图 2 所示，本发明的基于多核架构的批量式报文收发方法包括：

[0045] 步骤 S201，上层协议栈调用网卡驱动发送函数。

[0046] 步骤 S202，网卡驱动发送函数将待发送报文的标识信息缓存至发送报文缓存队列，发送报文缓存队列缓存的标识信息数递增，并启动定时器。在此之前，若未创建发送报文缓存队列，则创建发送报文缓存队列，其大小为 1 ~ n 个待发送报文的标识信息的总大小，并将缓存的标识信息数设置为 0，n 优选取值为：1 到 256 中任意为 2 的幂的值。

[0047] 步骤 S203，所述网卡驱动发送函数检测发送报文缓存队列中缓存的标识信息数，若缓存的标识信息数大于预定值，则转步骤 S204，否则退出网卡驱动发送函数，转步骤 S205。这个预定值根据具体系统的试验确定，通常可以选择 16, 64, 128 等，但小于等于发送报文缓存队列的大小。

[0048] 步骤 S204，启动硬件（网卡）发送队列中缓存的所有标识信息对应的待发送报文，并将缓存的标识信息数归 0，并复位定时器；

[0049] 步骤 S205，判断定时器是否超时，若超时，则转步骤 S204，否则循环等待。定时器设置时间可根据当前收发报文的任务量及网络状况确定。

[0050] 在实施例 1 的基础上，步骤 2 中增加了定时器，避免长时间没有发送请求，条件 1 长时间无法满足，导致已经缓存在队列中的标识信息对应的报文无法发送的情况。

[0051] 以上实施方式仅用于说明本发明,而并非对本发明的限制,有关技术领域的普通技术人员,在不脱离本发明的精神和范围的情况下,还可以做出各种变化和变型,因此所有等同的技术方案也属于本发明的范畴,本发明的专利保护范围应由权利要求限定。

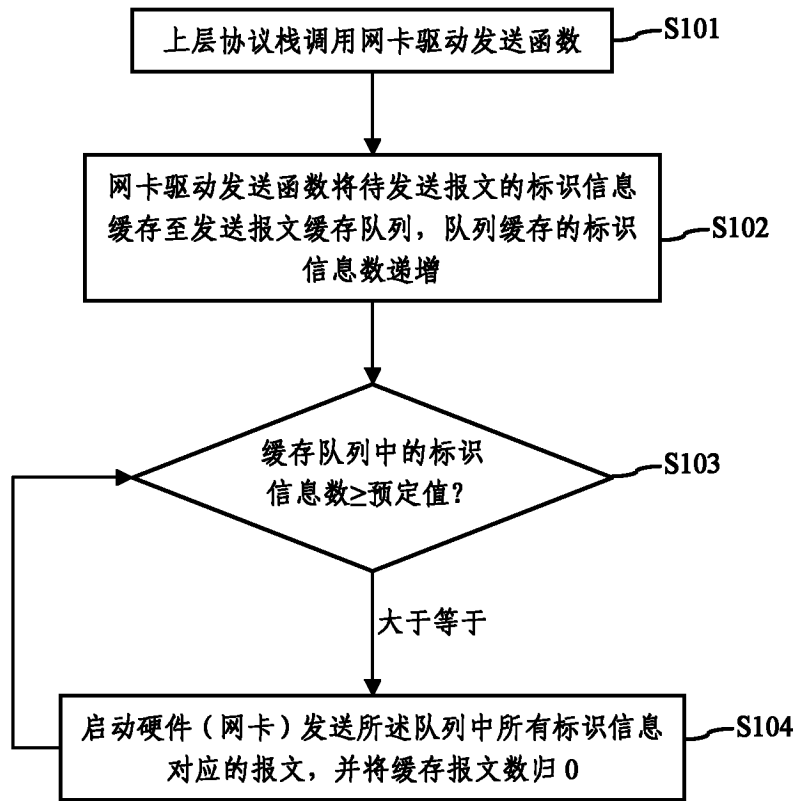


图 1

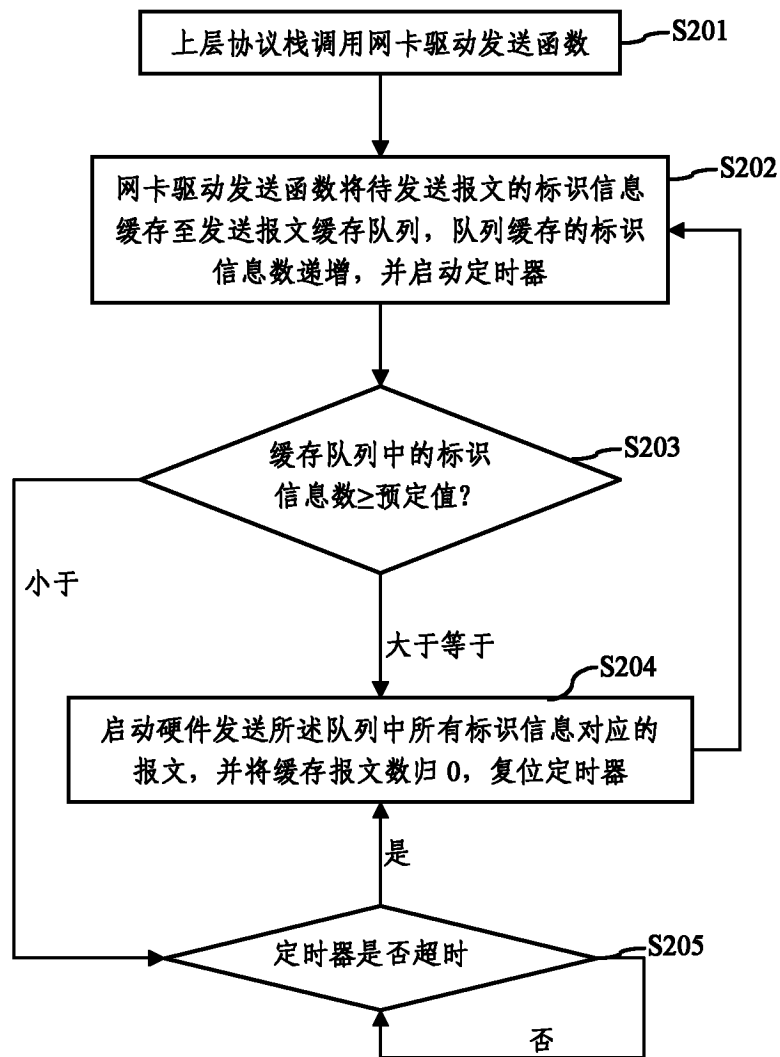


图 2