(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109617832 B
(45)授权公告日 2022.07.08

(21)申请号 201910101471.4

(22)申请日 2019.01.31

(65)同一申请的已公布的文献号
申请公布号 CN 109617832 A

(43)申请公布日 2019.04.12

(73)专利权人 新华三技术有限公司合肥分公司
地址 230000 安徽省合肥市高新区创新大道2800号创新产业园二期J1楼A座5-9层

(72)发明人 徐炽云

(74)专利代理机构 北京超凡志成知识产权代理事务所(普通合伙) 11371
专利代理师 邓超

(51)Int.Cl.
H04L 49/90(2022.01)

(56)对比文件
CN 105677580 A,2016.06.15
CN 108132889 A,2018.06.08
US 2008244231 A1,2008.10.02
US 2010146218 A1,2010.06.10
CN 103338157 A,2013.10.02
CN 105337896 A,2016.02.17
CN 108768898 A,2018.11.06

审查员 白红昌

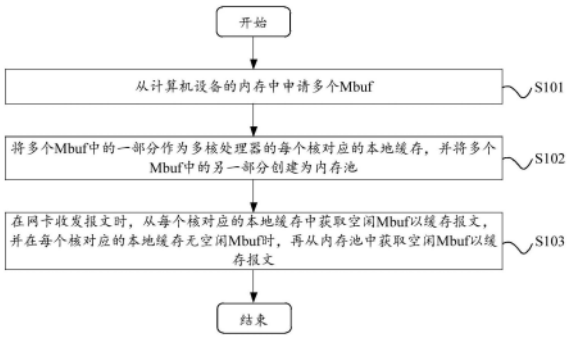
权利要求书2页 说明书9页 附图2页

(54)发明名称

报文缓存方法及装置

(57)摘要

本发明实施例涉及计算机技术领域，提供一种报文缓存方法及装置，所述方法包括：从计算机设备的存储器中申请多个Mbuf；将多个Mbuf中的一部分作为多核处理器的每个核对应的本地缓存，并将多个Mbuf中的另一部分创建为内存池；在网卡收发报文时，从每个核对应的本地缓存中获取空闲Mbuf以缓存报文，并在每个核对应的本地缓存无空闲Mbuf时，再从内存池中获取空闲Mbuf以缓存报文。与现有技术相比，本发明实施例在申请多个Mbuf时将每个核对应的本地缓存考虑在内，并且在收发报文时优先使用每个核对应的本地缓存中的Mbuf缓存报文，这样可以保证在缓存大量报文的同时，驱动仍然能正常收发包。



CN 109617832 B

1.一种报文缓存方法,其特征在于,应用于计算机设备,所述计算机设备包括网卡、多核处理器和存储器,所述方法包括:

从所述计算机设备的存储器中申请多个Mbuf;

将所述多个Mbuf中的一部分作为所述多核处理器的每个核对应的本地缓存,并将所述多个Mbuf中的另一部分创建为内存池;

在网卡收发报文时,从所述每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在所述每个核对应的本地缓存无空闲Mbuf时,再从所述内存池中获取空闲Mbuf以缓存报文;

所述从所述计算机设备的存储器中申请多个Mbuf的步骤,包括:

计算所述网卡的端口接收队列占用Mbuf总数目;

计算所述网卡的端口发送队列占用Mbuf总数目;

确定所述多核处理器的每个核对应的本地缓存占用Mbuf总数目;

将所述端口接收队列占用Mbuf总数目、所述端口发送队列占用Mbuf总数目、所述每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目之和作为待申请Mbuf数目;

依据所述待申请Mbuf数目,从所述计算机设备的大页内存中申请多个Mbuf。

2.如权利要求1所述的方法,其特征在于,所述计算所述网卡的端口接收队列占用Mbuf总数目的步骤,包括:

获取接收队列长度、所述网卡的端口数目、以及所述多核处理器的转发核数目;

确定每个端口的总的接收队列数目为所述转发核数目,其中,每个转发核在每个端口都对应一个接收队列;

依据所述端口数目、接收队列长度、以及每个端口的总的接收队列数目的乘积,计算所述端口接收队列占用Mbuf总数目。

3.如权利要求1所述的方法,其特征在于,所述计算所述网卡的端口发送队列占用Mbuf总数目的步骤,包括:

获取发送队列长度、所述网卡的端口数目、以及所述多核处理器的控制核数目和转发核数目;

确定每个端口的总的发送队列数目为所述多核处理器的转发核与控制核的总数目,其中,每个控制核和每个转发核在每个端口均对应一个发送队列;

依据所述端口数目、发送队列长度、以及每个端口的总的发送队列数目的乘积,计算所述端口发送队列占用Mbuf总数目。

4.如权利要求1所述的方法,其特征在于,所述确定所述多核处理器的每个核对应的本地缓存占用Mbuf总数目的步骤,包括:

获取所述多核处理器的控制核数目和转发核数目;

依据所述多核处理器中控制核和转发核的总数目与预设占用数目的乘积,计算所述多核处理器的每个核对应的本地缓存占用Mbuf总数目,其中,所述预设占用数目是所述多核处理器中单个核占用的Mbuf数目。

5.一种报文缓存装置,其特征在于,应用于计算机设备,所述计算机设备包括网卡、多核处理器和存储器,所述装置包括:

申请模块,用于从所述计算机设备的存储器中申请多个Mbuf;

执行模块,用于将所述多个Mbuf中的一部分作为所述多核处理器的每个核对应的本地

缓存,并将所述多个Mbuf中的另一部分创建为内存池;

报文缓存模块,用于在网卡收发报文时,从所述每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在所述每个核对应的本地缓存无空闲Mbuf时,再从所述内存池中获取空闲Mbuf以缓存报文;

所述申请模块具体用于:

计算所述网卡的端口接收队列占用Mbuf总数目;

计算所述网卡的端口发送队列占用Mbuf总数目;

确定所述多核处理器的每个核对应的本地缓存占用Mbuf总数目;

将所述端口接收队列占用Mbuf总数目、所述端口发送队列占用Mbuf总数目、所述每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目之和作为待申请Mbuf数目;

依据所述待申请Mbuf数目,从所述计算机设备的大页内存中申请多个Mbuf。

6.如权利要求5所述的装置,其特征在于,所述申请模块执行所述计算所述网卡的端口接收队列占用Mbuf总数目的方式,包括:

获取接收队列长度、所述网卡的端口数目、以及所述多核处理器的转发核数目;

确定每个端口的总的接收队列数目为所述转发核数目,其中,每个转发核在每个端口都对应一个接收队列;

依据所述端口数目、接收队列长度、以及每个端口的总的接收队列数目的乘积,确定出所述端口接收队列占用Mbuf总数目。

7.如权利要求5所述的装置,其特征在于,所述申请模块执行所述计算所述网卡的端口发送队列占用Mbuf总数目的方式,包括:

获取发送队列长度、所述网卡的端口数目、以及所述多核处理器的控制核数目和转发核数目;

确定每个端口的总的发送队列数目为所述多核处理器的转发核与控制核的总数目,其中,每个控制核和每个转发核在每个端口均对应一个发送队列;

依据所述端口数目、发送队列长度、以及每个端口的总的发送队列数目的乘积,计算所述端口发送队列占用Mbuf总数目。

8.如权利要求5所述的装置,其特征在于,所述申请模块执行所述确定所述多核处理器的每个核对应的本地缓存占用Mbuf总数目的方式,包括:

获取所述多核处理器的控制核数目和转发核数目;

依据所述多核处理器中控制核和转发核的总数目与预设占用数目的乘积,计算所述多核处理器的每个核对应的本地缓存占用Mbuf总数目,其中,所述预设占用数目是所述多核处理器中单个核占用的Mbuf数目。

报文缓存方法及装置

技术领域

[0001] 本发明实施例涉及计算机技术领域,具体而言,涉及一种报文缓存方法及装置。

背景技术

[0002] 数据面开发套件(Data Plan Develop Kit,DPDK)技术是Intel公司开发的基于数据面的报文处理框架,DPDK可以支持数据的快速转发,是X86平台报文快速数据包处理的函数库和驱动集。

[0003] 为了高效访问数据,DPDK将内存封装在Mbuf(Memory buffer,存储器缓存)结构体内,即通过Mbuf来封装存放收到的报文。为了避免频繁收发包申请Mbuf内存带来的性能开销,通常将Mbuf存放在存储器的内存池中,内存池是由N个Mbuf组成的环形阵列,正常情况下,网卡的每个端口收发包时都可以从内存池中获取Mbuf。但是,网络中收到的报文包括待转发报文和送本机的报文,送本机的报文需要交给协议栈处理,在收到待转发报文和送本机的报文时都需要从内存池中获取Mbuf来承载报文,报文发出去后再将Mbuf还给内存池,如果内存池中的Mbuf缓存大量送本机的报文,会导致内存池中无空闲Mbuf,此时驱动无法再收包。

发明内容

[0004] 本发明实施例的目的在于提供一种报文缓存方法及装置,用以在缓存大量报文时保证驱动正常收包。

[0005] 为了实现上述目的,本发明实施例采用的技术方案如下:

[0006] 第一方面,本发明实施例提供了一种报文缓存方法,应用于计算机设备,所述计算机设备包括网卡、多核处理器和存储器,所述方法包括:从所述计算机设备的存储器中申请多个Mbuf;将所述多个Mbuf中的一部分作为所述多核处理器的每个核对应的本地缓存,并将所述多个Mbuf中的另一部分创建为内存池;在网卡收发报文时,从所述每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在所述每个核对应的本地缓存无空闲Mbuf时,再从所述内存池中获取空闲Mbuf以缓存报文。

[0007] 第二方面,本发明实施例还提供了一种报文缓存装置,应用于计算机设备,所述计算机设备包括网卡、多核处理器和存储器,所述装置包括申请模块、执行模块及报文缓存模块。其中,申请模块用于从所述计算机设备的存储器中申请多个Mbuf;执行模块用于将所述多个Mbuf中的一部分作为所述多核处理器的每个核对应的本地缓存,并将所述多个Mbuf中的另一部分创建为内存池;报文缓存模块用于在网卡收发报文时,从所述每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在所述每个核对应的本地缓存无空闲Mbuf时,再从所述内存池中获取空闲Mbuf以缓存报文。

[0008] 相对现有技术,本发明实施例提供的一种报文缓存方法及装置,从计算机设备的存储器中申请多个Mbuf,并将该多个Mbuf中的一部分作为计算机设备的本地缓存、将该多个Mbuf中的另一部分创建为内存池,在网卡收发报文时,优先从本地缓存中获取空闲Mbuf

以缓存报文，并在本地缓存无空闲Mbuf时，再从内存池中获取空闲Mbuf以缓存报文。与现有技术相比，本发明实施例在申请多个Mbuf时将每个核对应的本地缓存考虑在内，并且在收发报文时优先使用每个核对应的本地缓存中的Mbuf缓存报文，这样即使每个核对应的本地缓存中缓存了大量送本机的报文，也能保证内存池中存在空闲Mbuf，如此，在收到报文时仍然可以从内存池中获取空闲Mbuf来缓存报文，也就是说，可以保证在缓存大量送本机的报文的同时，驱动仍然能正常收发包。

[0009] 为使本发明的上述目的、特征和优点能更明显易懂，下文特举较佳实施例，并配合所附附图，作详细说明如下。

附图说明

[0010] 为了更清楚地说明本发明实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本发明的某些实施例，因此不应被看作是对范围的限定，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他相关的附图。

[0011] 图1示出了本发明实施例提供的计算机设备的方框示意图。

[0012] 图2示出了本发明实施例提供的报文缓存方法流程图。

[0013] 图3为图2示出的步骤S101的子步骤流程图。

[0014] 图4示出了本发明实施例提供的报文缓存装置的方框示意图。

[0015] 图标：100-计算机设备；101-多核处理器；102-存储器；103-总线；104-网卡；200-报文缓存装置；201-申请模块；202-执行模块；203-报文缓存模块。

具体实施方式

[0016] 下面将结合本发明实施例中附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本发明实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围，而是仅仅表示本发明的选定实施例。基于本发明的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0017] 应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步定义和解释。同时，在本发明的描述中，术语“第一”、“第二”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

[0018] 请参照图1，图1示出了本发明实施例提供的计算机设备100的方框示意图。计算机设备100包括多核处理器101、存储器102、总线103和网卡104，多核处理器101、存储器102和网卡104通过总线103互相通信。

[0019] 其中，多核处理器101用于执行存储器102中存储的可执行模块，例如计算机程序。本发明实施例所指的多核处理器101可以是多核CPU(Central Processing Unit，中央处理器)，例如，四核CPU等，多核处理器101中的每个核可以是CPU核。

[0020] 存储器102，主要用于存储计算机设备100中的各种程序和数据。存储器102可以是一个存储装置，也可以是多个存储元件的统称，存储器102可以包括随机存储器(random

access memory,RAM),也可以包括非易失性存储器(non-volatile memory),例如磁盘存储器,闪存(Flash)等。

[0021] 总线103可以是工业标准体系结构(Industry Standard Architecture,ISA)总线、外部设备互连(Peripheral Component,PCI)总线或扩展工业标准体系结构(Extended Industry Standard Architecture,EISA)总线等。该总线103可以分为地址总线、数据总线、控制总线等。为便于表示,图1中仅用一个双向箭头表示,但并不表示仅有一根总线或一种类型的总线。

[0022] 网卡104,为网络接口卡(Network Interface Card,NIC)的简称,为主要工作在链路层的网络组件,是局域网中连接计算机和传输介质的接口,不仅能实现与局域网传输介质之间的物理连接和电信号匹配,还涉及帧的发送与接收、帧的封装与拆封、介质访问控制、数据的编码与解码以及数据缓存的功能等。

[0023] 存储器102用于存储程序,所述多核处理器101在接收到执行指令后,执行所述程序以实现发明实施例揭示的报文缓存方法。

[0024] 本发明实施例提供的报文缓存方法可以应用于上述的计算机设备100,计算机设备100可以是服务器、个人计算机、网络设备等,其操作系统可以是Windows操作系统、Linux操作系统等。

[0025] 为了适应数据高速转发的需要,在计算机设备100的操作系统中配置DPDK,DPDK用于快速数据包处理的函数库和驱动集合,是一种可以极大提高数据处理性能和吞吐量、以及数据平台应用程序的工作效率的软件开发套件。DPDK通过Mbuf来封装存放收到的报文,为了避免频繁收发包申请Mbuf内存带来的性能开销,通常将Mbuf存放在内存池中,具体来说,DPDK初始化时,创建一个包含多个Mbuf的内存池,计算机设备100的网卡104收发报文时,从内存池中获取Mbuf以缓存报文。

[0026] 但是,网络中收到的报文包括待转发报文和送本机的报文。DPDK在收到送本机的报文时需要交给协议栈处理,而DPDK在收到送本机的报文时也是从内存池中获取Mbuf进行缓存的,在协议栈处理过程中,缓存这部分报文的Mbuf不会被释放。因此,如果DPDK收到大量送本机的报文,会从内存池中获取大量Mbuf进行缓存,导致内存池中无空闲Mbuf,这样网卡104在收发报文时会由于无法获取到空闲Mbuf而导致报文被丢弃。为了解决这一问题,本发明实施例在申请多个Mbuf时将每个核对应的本地缓存考虑在内,并且在收发报文时优先使用每个核对应的本地缓存中的Mbuf缓存报文,这样即使每个核对应的本地缓存中的Mbuf缓存了大量送本机的报文,也能保证内存池中存在空闲Mbuf,如此,在收到报文时仍然可以从内存池中获取空闲Mbuf来缓存报文,从而保证在缓存大量报文的同时,驱动仍然能正常收发包,下面进行详细描述。

[0027] 第一实施例

[0028] 请参照图2,图2示出了本发明实施例提供的报文缓存方法流程图。该报文缓存方法应用于上述的计算机设备100,该报文缓存方法包括以下步骤:

[0029] 步骤S101,从计算机设备的存储器中申请多个Mbuf。

[0030] 在相关技术中,在DPDK基于多核架构的情况下,即,计算机设备100包括多核处理器101,多核处理器101的多个核访问同一个内存池时,每个核进行数据读写都要进行比较并交换(Compare and Swap,CAS)操作来保证数据未被其他核修改,这样就导致报文转发效

率极低。

[0031] 为了解决这一问题，本发明实施例将一部分Mbuf作为多核处理器101的每个核对应的本地缓存，这样多核处理器101的每个核可以优先从其对应的本地缓存中获取Mbuf，从而减少多核处理器101中多个核竞争内存池带来的开销。也就是说，本发明实施例在DPDK初始化申请多个Mbuf时，将多核处理器101的每个核对应的本地缓存考虑在内，来提高报文转发效率。

[0032] 需要说明的是，本实施例中的本地缓存是指存储器102中的缓存，本地缓存是为了与内存池进行区别。

[0033] 在DPDK中，如果多核处理器101的控制核和转发核需要同时访问网卡104，则利用网卡多队列技术，在进行队列分配时，设置每个转发核在网卡104的每个端口上都负责一个接收队列，避免控制核和转发核对同一队列进行并发操作。具体来说，由于每个核在网卡104的每个端口上都有可能发包，则设置每个核在每个端口都对应一个发送队列，即，网卡104每个端口的总的发送队列数目等于启动计算机设备100时占用控制核和转发核的数目；同时，由于多核处理器101的控制核仅负责发送队列，则设置每个转发核在每个端口对应一个接收队列，即，网卡104每个端口的总的接收队列数目等于启动计算机设备100时占用转发核的数目。

[0034] 例如，计算机设备100启动时，网卡104包括两个端口port0和port1，并占用四个核core0、core1、core2和core3，其中，core0为控制核，core1、core2和core3为转发核，则可以设置计算机设备100中两个端口、四个核、接收队列RxQ及发送队列TxQ之间的对应关系如下表所示：

[0035]

	Core0	Core1	Core2	Core3
Port0	TxQ：0	TxQ：1	TxQ：2	TxQ：3

[0036]

	RxQ：NA	RxQ：0	RxQ：1	RxQ：2
Port1	TxQ：0 RxQ：NA	TxQ：1 RxQ：0	TxQ：2 RxQ：1	TxQ：3 RxQ：2

[0037] 其中，0、1、2、3表示队列标识，例如TxQ:0表示发送队列0，RxQ:0表示接收队列0，RxQ:NA表示不负责接收队列。

[0038] 基于以上队列分配方式，为了提高内存分配性能，在DPDK初始化申请多个Mbuf时，将多核处理器101中每个核对应的本地缓存考虑在内，因此，待申请Mbuf数目需要考虑网卡104的端口接收队列占用Mbuf总数目、网卡104的端口发送队列占用Mbuf总数目、多核处理器101的每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目，也就是说，待申请Mbuf数目等于网卡104的端口接收队列占用Mbuf总数目、网卡104的端口发送队列占用Mbuf总数目、每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目的总和。

[0039] 其中，网卡104的端口接收队列占用Mbuf总数目等于网卡104的端口数目、接收队列长度、以及每个端口的总的接收队列数目的乘积，由于每个转发核在每个端口都对应一个接收队列，故每个端口的总的接收队列数目等于多核处理器101的转发核数目，网卡104的端口接收队列可以是网卡104的所有端口的接收队列的集合，例如，网卡104包括两个端

口port0和port1，端口port0包括0#接收队列和1#接收队列，端口port1包括2#接收队列和3#接收队列，则网卡104的端口接收队列包括0#接收队列、1#接收队列、2#接收队列和3#接收队列。

[0040] 网卡104的端口发送队列占用Mbuf总数目等于网卡104的端口数目、发送队列长度、以及每个端口的总的发送队列数目的乘积，由于每个控制核和每个转发核在每个端口均对应一个发送队列，故每个端口的总的发送队列数目为所述多核处理器101的转发核与控制核的总数目，网卡104的端口发送队列可以是网卡104的所有端口的发送队列的集合，例如，网卡104包括两个端口port0和port1，端口port0包括0#发送队列和1#发送队列，端口port1包括2#发送队列和3#发送队列，则网卡104的端口发送队列包括0#发送队列、1#发送队列、2#发送队列和3#发送队列。

[0041] 多核处理器101的每个核对应的本地缓存占用Mbuf总数目包括多核处理器101中每个核对应的预设占用数目的总和，也就是，每个核对应的本地缓存占用Mbuf总数目包括多核处理器101中控制核和转发核的总数目与每个核对应的预设占用数目的乘积，预设占用数目是多核处理器101中单个核占用的Mbuf数目。可选地，预设占用数目可以为DPDK支持的多核处理器101中单个核最大可占用的Mbuf个数。在一种可选的实施方式中，多核处理器101中每个控制核对应的预设占用数目可以是512个，每个转发核对应的预设占用数目也可以是512个。预设的Mbuf数目可以是根据协议栈可以缓存的由DPDK驱动收上来的报文数目的最大值确定的，以在申请Mbuf时，为协议栈缓存的报文预留出可申请的Mbuf，在一种示例中，预设的Mbuf数目可以是16K。

[0042] 请参照图3，步骤S101还可以包括以下子步骤：

[0043] 子步骤S1011，计算网卡的端口接收队列占用Mbuf总数目。

[0044] 在本发明实施例中，多核处理器101的控制核仅负责发送队列，故网卡104每个端口的总的接收队列数目等于启动计算机设备100时占用转发核的数目，则网卡104的端口接收队列占用Mbuf总数目的计算过程可以包括：

[0045] 获取接收队列长度、网卡104的端口数目、以及多核处理器101的转发核数目；

[0046] 确定每个端口的总的接收队列数目为多核处理器101的转发核数目，其中，每个转发核在每个端口都对应一个接收队列；

[0047] 依据端口数目、接收队列长度、以及每个端口的总的接收队列数目的乘积，计算端口接收队列占用Mbuf总数目，即，端口接收队列占用Mbuf总数目等于p*(M-1)*L1，其中，p表示网卡104的端口数目，M表示多核处理器101的转发核与控制核的总数目，(M-1)表示多核处理器101的转发核数目即每个端口的总的接收队列数目，L1表示接收队列长度。

[0048] 子步骤S1012，计算网卡的端口发送队列占用Mbuf总数目。

[0049] 在本发明实施例中，网卡104每个端口的发送队列数目等于启动计算机设备100时占用转发核和控制核的数目，则网卡104的端口发送队列占用Mbuf总数目计算过程可以包括：

[0050] 获取发送队列长度、网卡104的端口数目、以及多核处理器101的控制核数目和转发核数目；

[0051] 确定每个端口的总的发送队列数目为多核处理器101的转发核与控制核的总数目，其中，每个控制核和每个转发核在每个端口均对应一个发送队列；

[0052] 依据端口数目、发送队列长度、以及每个端口的总的发送队列数目的乘积，计算端口发送队列占用Mbuf总数目，即，端口发送队列占用Mbuf总数目等于p*M*L2，其中，p表示网卡104的端口数目，M表示多核处理器101的转发核与控制核的总数目即每个端口的总的发送队列数目，L2表示发送队列长度。

[0053] 子步骤S1013，确定多核处理器的每个核对应的本地缓存占用Mbuf总数目。

[0054] 在本发明实施例中，为了减少多核处理器101中控制核和转发核竞争内存池带来的开销，可以让多核处理器101的每个核均占用一部分Mbuf，则确定多核处理器101的每个核对应的本地缓存占用Mbuf总数目的过程可以包括：

[0055] 获取多核处理器101的控制核数目和转发核数目；

[0056] 依据多核处理器101中控制核和转发核的总数目与预设占用数目的乘积，计算多核处理器101的每个核对应的本地缓存占用Mbuf总数目，预设占用数目是多核处理器101中单个核占用的Mbuf数目；

[0057] 其中，考虑到缓存报文的情况，多核处理器101中单个核对应的预设占用数目均可以是512个，即，多核处理器101的每个核对应的本地缓存占用Mbuf总数目等于M*512，其中，M表示多核处理器101的转发核与控制核的总数目即每个端口的总的发送队列数目。

[0058] 子步骤S1014，将端口接收队列占用Mbuf总数目、端口发送队列占用Mbuf总数目、每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目之和作为待申请Mbuf数目。

[0059] 在本发明实施例中，预设的Mbuf数目可以是16K，则待申请Mbuf数目的计算过程可以用公式N＝p*(M-1)*L1+p*M*L2+M*512+16K表示，其中，N表示待申请Mbuf数目。

[0060] 子步骤S1015，依据待申请Mbuf数目，从计算机设备的大页内存中申请多个Mbuf。

[0061] 在本发明实施例中，计算出待申请Mbuf数目之后，则按照待申请Mbuf数目，从计算机设备100的大页内存中申请多个Mbuf，一般的常规页大小为4K字节，使用大页时页大小设置为2M或1G字节。

[0062] 步骤S102，将多个Mbuf中的一部分作为多核处理器的每个核对应的本地缓存，并将多个Mbuf中的另一部分创建为内存池。

[0063] 在本发明实施例中，按照待申请Mbuf数目，从计算机设备100的大页内存中申请多个Mbuf之后，分出一定数目的Mbuf作为多核处理器101的每个核对应的本地缓存，具体来说，首先，依次将预设占用数目的Mbuf作为多核处理器101中每个核对应的本地缓存，也就是说，CPU中每个核对应的本地缓存均包括预设占用数目的Mbuf，预设占用数目可以是512个；然后，将多个Mbuf中除多核处理器101的每个核对应的本地缓存之外的其它Mbuf创建为内存池，即，将(N-M*512)个Mbuf创建为内存池。

[0064] 步骤S103，在网卡收发报文时，从每个核对应的本地缓存中获取空闲Mbuf以缓存报文，并在每个核对应的本地缓存无空闲Mbuf时，再从内存池中获取空闲Mbuf以缓存报文。

[0065] 在本发明实施例中，空闲Mbuf是指未缓存报文的Mbuf，DPDK中普遍采用纯轮询模式进行报文收发，与报文收发有关的中断在网卡104端口初始化时会关闭。在网卡104收发报文时，多核处理器101的每个核优先从其对应的本地缓存中申请Mbuf来缓存报文，如果其对应的本地缓存中无空闲Mbuf，再从内存池中申请空闲Mbuf来缓存报文。同样的，报文收发完成需要释放Mbuf时，优先将Mbuf释放到每个核对应的本地缓存中，如果每个核对应的本地缓存中Mbuf均达到预设占用数目，再将Mbuf释放到内存池中。

[0066] 与现有技术相比，本发明实施例具有以下有益效果：

[0067] 首先，现有技术中，DPDK中接收队列的实现需要在计算机设备100启动时，用户手动配置网卡104端口、接收队列和多核处理器101中每个核的对应关系，本发明实施例通过设置，即，网卡104每个端口的总的接收队列数目等于启动计算机设备100时占用转发核的数目，无需用户手动配置，使得计算机设备100的启动更为方便。另外，每个转发核在每个端口对应一个接收队列的方式，可以避免控制核和转发核对同一队列进行并发操作，从而系统整体吞吐能力得到较大提升。

[0068] 其次，将从计算机设备100的存储器102中申请的多个Mbuf中的一部分Mbuf作为多核处理器101的每个核对应的本地缓存，这样多核处理器101的每个核可以优先从其对应的本地缓存中获取Mbuf，即使每个核对应的本地缓存中缓存了大量送本机的报文，也能保证内存池中存在空闲Mbuf，如此，在收到报文时仍然可以从内存池中获取空闲Mbuf来缓存报文，从而保证在缓存大量报文的同时，驱动仍然能正常收发包。

[0069] 最后，为多核处理器101中每个核对应的本地缓存分配预设占用数目(例如，512个)的Mbuf，这样即使协议栈缓存了大量报文，也能保证系统仍然可以具有较高的吞吐能力。

[0070] 第二实施例

[0071] 请参照图4，图4示出了本发明实施例提供的报文缓存装置200的方框示意图。报文缓存装置200包括申请模块201、执行模块202及报文缓存模块203。

[0072] 申请模块201，用于从计算机设备的存储器中申请多个Mbuf。

[0073] 在本发明实施例中，申请模块201具体用于计算网卡的端口接收队列占用Mbuf总数目；计算网卡的端口发送队列占用Mbuf总数目；确定多核处理器的每个核对应的本地缓存占用Mbuf总数目；将端口接收队列占用Mbuf总数目、端口发送队列占用Mbuf总数目、每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目之和作为待申请Mbuf数目；依据待申请Mbuf数目，从计算机设备的大页内存中申请多个Mbuf。

[0074] 在本发明实施例中，申请模块201执行计算网卡的端口接收队列占用Mbuf总数目的方式，包括：获取接收队列长度、网卡的端口数目、以及多核处理器的转发核数目；确定每个端口的总的接收队列数目为转发核数目，其中，每个转发核在每个端口都对应一个接收队列；依据端口数目、接收队列长度、以及每个端口的总的接收队列数目的乘积，确定出端口接收队列占用Mbuf总数目。

[0075] 在本发明实施例中，申请模块201执行计算网卡的端口发送队列占用Mbuf总数目的方式，包括：获取发送队列长度、网卡的端口数目、以及多核处理器的控制核数目和转发核数目；确定每个端口的总的发送队列数目为多核处理器的转发核与控制核的总数目，其中，每个控制核和每个转发核在每个端口均对应一个发送队列；依据端口数目、发送队列长度、以及每个端口的总的发送队列数目的乘积，计算所述发送队列占用Mbuf总数目。

[0076] 在本发明实施例中，申请模块201执行确定多核处理器的每个核对应的本地缓存占用Mbuf总数目的方式，包括：获取多核处理器的控制核数目和转发核数目；依据多核处理器中控制核和转发核的总数目与预设占用数目的乘积，计算多核处理器的每个核对应的本地缓存占用Mbuf总数目，其中，预设占用数目是多核处理器中单个核占用的Mbuf数目。

[0077] 执行模块202，用于将多个Mbuf中的一部分作为多核处理器的每个核对应的本地

缓存,并将多个Mbuf中的另一部分创建为内存池。

[0078] 报文缓存模块203,用于在网卡收发报文时,从每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在每个核对应的本地缓存无空闲Mbuf时,再从内存池中获取空闲Mbuf以缓存报文。

[0079] 本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被多核处理器101执行时实现本发明实施例揭示的报文缓存方法。

[0080] 综上所述,本发明实施例提供的一种报文缓存方法及装置,应用于计算机设备,计算机设备包括网卡、多核处理器和存储器,所述方法包括:从计算机设备的存储器中申请多个Mbuf;将多个Mbuf中的一部分作为多核处理器的每个核对应的本地缓存,并将多个Mbuf中的另一部分创建为内存池;在网卡收发报文时,从每个核对应的本地缓存中获取空闲Mbuf以缓存报文,并在每个核对应的本地缓存无空闲Mbuf时,再从内存池中获取空闲Mbuf以缓存报文。与现有技术相比,本发明实施例在申请多个Mbuf时将每个核对应的本地缓存考虑在内,并且在收发报文时优先使用每个核对应的本地缓存中的Mbuf缓存报文,这样即使每个核对应的本地缓存中缓存了大量送本机的报文,也能保证内存池中存在空闲Mbuf,如此,在收到报文时仍然可以从内存池中获取空闲Mbuf来缓存报文,也就是说,可以保证在缓存大量报文的同时,驱动仍然能正常收发包。

[0081] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本发明的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0082] 另外,在本发明各个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是各个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0083] 所述功能如果以软件功能模块的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括

没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0084] 以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步定义和解释。

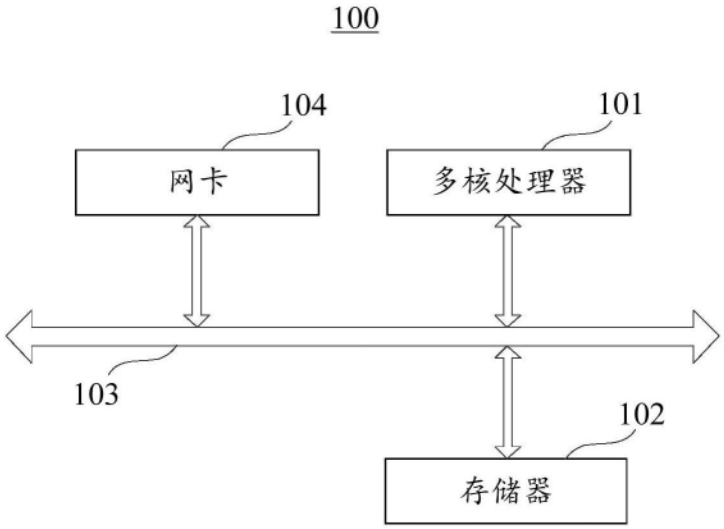

100
104
网卡
101
多核处理器
103
102
存储器

图1

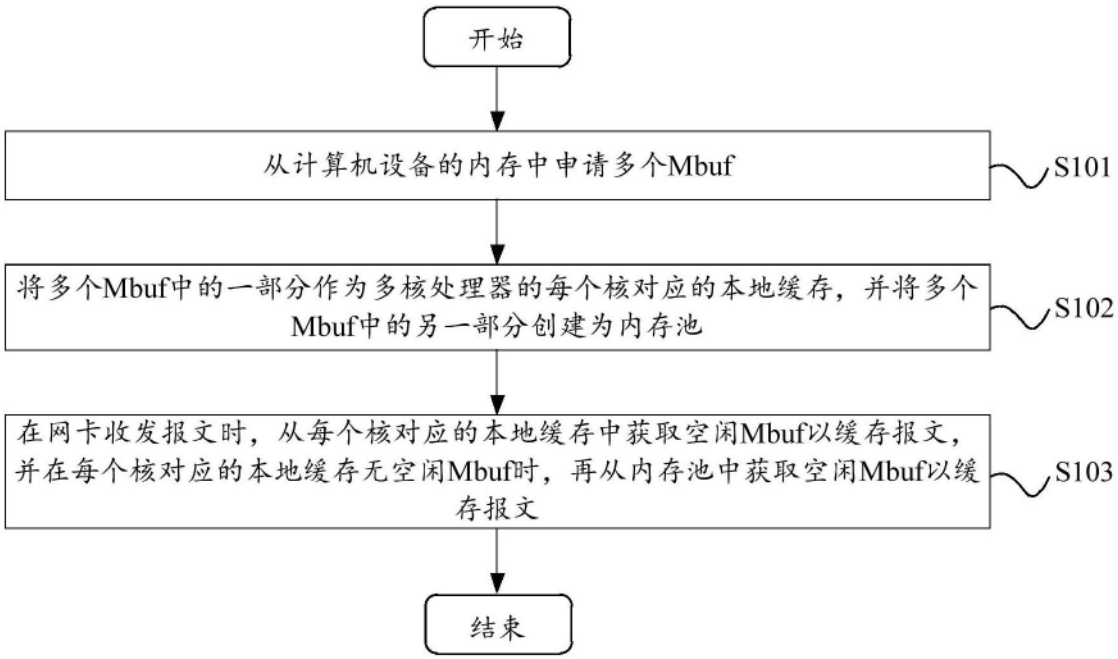

开始
从计算机设备的内存中申请多个Mbuf
S101
将多个Mbuf中的一部分作为多核处理器的每个核对应的本地缓存，并将多个Mbuf中的另一部分创建为内存池
S102
在网卡收发报文时，从每个核对应的本地缓存中获取空闲Mbuf以缓存报文，并在每个核对应的本地缓存无空闲Mbuf时，再从内存池中获取空闲Mbuf以缓存报文
S103
结束

图2

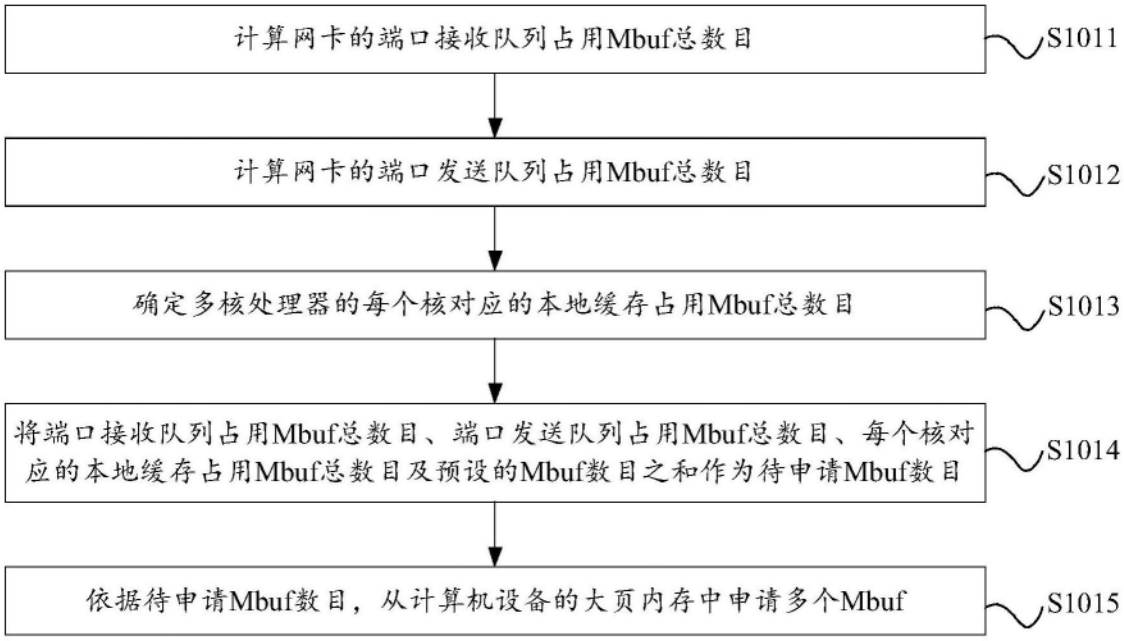

计算网卡的端口接收队列占用Mbuf总数目
S1011
计算网卡的端口发送队列占用Mbuf总数目
S1012
确定多核处理器的每个核对应的本地缓存占用Mbuf总数目
S1013
将端口接收队列占用Mbuf总数目、端口发送队列占用Mbuf总数目、每个核对应的本地缓存占用Mbuf总数目及预设的Mbuf数目之和作为待申请Mbuf数目
S1014
依据待申请Mbuf数目，从计算机设备的大页内存中申请多个Mbuf
S1015

图3

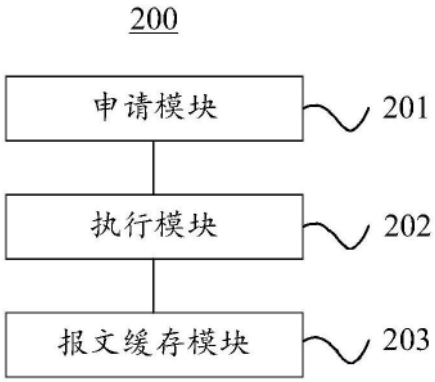

200
申请模块
201
执行模块
202
报文缓存模块
203

图4