(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104753813 B
(45)授权公告日 2018.03.16

(21)申请号 201310740954.1

(22)申请日 2013.12.27

(65)同一申请的已公布的文献号
申请公布号 CN 104753813 A

(43)申请公布日 2015.07.01

(73)专利权人 国家计算机网络与信息安全管理中心
地址 100029 北京市朝阳区裕民路甲3号
专利权人 杭州迪普科技股份有限公司

(72)发明人 邹昕 周立 何清林 王维晟 闫攀 任晓瑶 秦德楼 于林涛 杜建明 原万万

(74)专利代理机构 北京博思佳知识产权代理有限公司 11415
代理人 林祥

(51)Int.Cl.
H04L 12/861(2013.01)

(56)对比文件
CN 102420749 A,2012.04.18,
CN 102244579 A,2011.11.16,
CN 102185770 A,2011.09.14,
CN 101650698 A,2010.02.17,
CN 101645832 A,2010.02.10,
CN 1801806 A,2006.07.12,
US 7764676 B1,2010.07.27,
苏绥平.一种零拷贝报文捕获技术及其性能分析.《数字技术与应用》.2011,(第7期),正文第199-203页.
刘小威等.零拷贝技术在网络分析工具中的应用.《计算机系统应用》.2012,第21卷(第4期),正文第169-173页.

审查员 刘畅

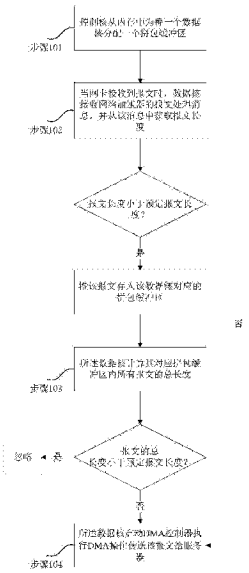
权利要求书1页 说明书3页 附图3页

(54)发明名称

DMA传送报文的方法

(57)摘要

本发明提供一种DMA传送报文的方法，应用于网卡，包括：控制核从内存中为每一个数据核分配一个拼包缓冲区；当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度信息，若该报文长度小于预定报文长度，则将该报文存入该数据核对应的拼包缓冲区；若该报文长度大于或等于预定报文长度，则启动DMA控制器执行DMA操作传送该报文给服务器；若数据核对应的拼包缓冲区内所有报文的总长度大于或等于预定报文长度，则启动DMA控制器执行DMA操作传送该报文给服务器；若报文总长度小于预定报文长度，则不作处理。本发明能够有效降低DMA的操作次数，提高PCIE通道带宽的有效利用率，从而提升了网卡小包的收包速率。



CN 104753813 B

1.一种DMA传送报文的方法，应用于网卡，该网卡包括多核CPU、内存以及DMA控制器，其中，该多核CPU包括控制核、网络加速器以及多个数据核，其特征在于，该方法包括以下步骤：

步骤A，控制核从内存中为每一个数据核分配一个拼包缓冲区，所述拼包缓冲区大小不小于2L-2字节，其中，L为预定报文长度；

步骤B，当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度信息，若该报文长度小于预定报文长度，则将该报文存入该数据核对应的拼包缓冲区，转至步骤C；若该报文长度大于或等于预定报文长度，则转至步骤D；

步骤C，所述数据核计算其对应拼包缓冲区内所有报文的总长度，若报文总长度大于或等于预定报文长度，则转至步骤D；若报文总长度小于预定报文长度，则不作处理；

步骤D，所述数据核启动DMA控制器执行DMA操作传送该报文给服务器。

2.如权利要求1所述的方法，其特征在于，所述预定报文长度为128字节。

3.如权利要求1所述的方法，其特征在于，还包括：

步骤E，控制核定时轮询所有拼包缓冲区，若轮询到的当前拼包缓冲区有报文，则启动DMA控制器执行DMA操作传送该拼包缓冲区内报文给服务器。

4.如权利要求3所述的方法，其特征在于：所述定时的时长大于控制核对所有拼包缓冲区进行一次轮询的总时间长度。

DMA传送报文的方法

技术领域

[0001] 本发明涉及网络数据处理领域，尤其涉及一种DMA传送报文的方法。

背景技术

[0002] 随着网络技术的快速发展，网络带宽每年以接近3倍的速度迅猛发展，目前10G网络已开始部署到部分端系统，给网络流量分析应用带来了较大压力。随着下一代互联网(NGI)核心技术，如IPV6/MPLS、路由协议、QOS技术等不断成熟，互联网进入向下一代演进的关键阶段。各种网络发展呈现高带宽、高流量化趋势，因此，大幅度提高网络设备的性能及处理能力显得尤为重要。

[0003] 服务器作为一种为用户提供共享信息资源和各种服务的高性能网络设备，其处理网络数据的能力很大程度上取决于网卡的性能。在实际使用中，网卡通过PCI-E插槽与服务器连接，当报文从网卡光口捕获并做前期处理后，通过PCI-E通道DMA(Direct Memory Access，直接内存存取)到服务器侧缓冲区，完成网卡侧任务处理。

[0004] 目前，网卡的捕包采用每接收到一个报文做一次DMA的策略。当报文长度越小时，携带的协议数据所占的比例就越大，占用的PCI-E通道带宽就越大。另外，在流量一定的情况下，报文长度越小，发起DMA操作的次数越多。

[0005] 综上所述，当报文长度越小时，由于其携带的控制协议增多，从而导致PCI-E通道带宽有效利用率越低。

发明内容

[0006] 有鉴于此，本发明提供一种DMA传送报文的方法，应用于网卡，该网卡包括多核CPU、内存以及DMA控制器，其中，该多核CPU包括控制核、网络加速器以及多个数据核，其特征在于，该方法包括以下步骤：

[0007] 步骤A，控制核从内存中为每一个数据核分配一个拼包缓冲区；

[0008] 步骤B，当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度信息，若该报文长度小于预定报文长度，则将该报文存入该数据核对应的拼包缓冲区，转至步骤C；若该报文长度大于或等于预定报文长度，则转至步骤D；

[0009] 步骤C，所述数据核计算其对应拼包缓冲区内所有报文的总长度，若报文总长度大于或等于预定报文长度，则转至步骤D；若报文总长度小于预定报文长度，则不作处理；

[0010] 步骤D，所述数据核启动DMA控制器执行DMA操作传送该报文给服务器。

[0011] 本发明能够有效降低DMA的操作次数，提高PCIE通道带宽的有效利用率，从而提升了网卡面向设备内部的实际数据速率。

附图说明

[0012] 图1是本发明一种实施方式中网卡内部基础硬件环境的示意图。

[0013] 图2是本发明一种实施方式中DMA传送报文方法的流程图。

[0014] 图3是本发明一种实施方式中DMA传送报文方法的详细流程图。

具体实施方式

[0015] 以下结合附图对本发明进行详细描述。

[0016] 本发明是针对网卡做出的改进方案，网卡作为服务器的重要组成部分，其网络数据处理能力决定了服务器的性能。网卡通过PCI-E插槽与服务器连接，当报文从网卡光口或者千兆电口被捕获并做前期处理后，通过PCI-E通道DMA到服务器侧缓冲区，完成网卡侧任务处理。如图1所示，该网卡包括多核CPU、内存、DMA控制器以及其他硬件，其中，该多核CPU包括控制核、网络加速器以及多个数据核。该DMA传送报文的方法通过在上述网卡硬件的基础上运行网卡驱动程序来实现。请参考图2，该方法的实现包括以下步骤：

[0017] 步骤101，控制核从内存中为每一个数据核分配一个拼包缓冲区；

[0018] 步骤102，当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度信息，若该报文长度小于预定报文长度，则将该报文存入该数据核对应的拼包缓冲区，转至步骤103；若该报文长度大于或等于预定报文长度，则转至步骤104；

[0019] 步骤103，所述数据核计算其对应拼包缓冲区内所有报文的总长度，若报文总长度大于或等于预定报文长度，则转至步骤104；若报文总长度小于预定报文长度，则不作处理；

[0020] 步骤104，所述数据核启动DMA控制器执行DMA操作传送该报文给服务器。

[0021] 现以具体实施例来说明该DMA传送报文方法的实现过程，请参考图3。首先，控制核需要从网卡内存中为每一个数据核分配一个拼包缓冲区，该拼包缓冲区大小不小于2L-2字节，其中，L为预定报文长度。在本实施例中，优选地，预定报文长度为128字节，则每一个拼包缓冲区的大小不小于254字节。这是因为对于大于或等于128字节的报文，数据核直接启动DMA操作将该报文传送给服务器，而对于小于128字节的报文，数据核则将该报文存入其对应的拼包缓冲区，因此，存入拼包缓冲区的最大报文长度为127字节，而拼包缓冲区中至少要存入2个报文才可以进行一次DMA操作，若2次存入的报文长度均为最大可存入报文长度，即127字节，则拼包缓冲区大小至少需要254字节。

[0022] 在控制核完成拼包缓冲区的分配后，网卡即可接收报文。当网卡从外部接收到报文时，先由网卡多核CPU内部的网络加速器模块对报文进行预处理。该网络加速器模块是多核CPU内部的硬件模块，在网卡启动时，由控制核对网络加速器模块进行初始化。当网络加速器模块接收到报文时，通过提取该报文中的报文特征，计算出应该由哪一个数据核对该报文进行处理，并向该数据核发送报文处理消息。网络加速器模块对报文的预处理过程是由硬件自动完成的，无需软件干预。

[0023] 数据核在接收到网络加速器的报文处理消息后，从该消息中获取报文长度信息。若该报文长度大于或等于预定报文长度，则该数据核启动DMA操作传送该报文给服务器；若该报文长度小于预定报文长度，则该数据核将该报文存入其对应的拼包缓冲区，并计算该拼包缓冲区内所有报文的总长度，若该报文总长度大于或等于预定报文长度，则启动一次DMA操作传送该报文给服务器；否则，不作处理。优选地，在本实施例中，预定报文长度为128字节，当接收到的报文长度或拼包后的报文总长度大于或等于128字节时，数据核启动DMA操作传送该报文给服务器。

[0024] 在上述对小报文进行拼包处理的过程中，为了避免拼包缓冲区内长时间没有新报

文存入，而导致拼包缓冲区内已有报文无法被DMA传送，控制核通过定时查询拼包缓冲区的方式，将拼包缓冲区内的报文DMA传送给服务器。具体实现方式为，控制核启动定时器，当定时时间到时，控制核顺序查询每一个拼包缓冲区，若当前查询的拼包缓冲区内有报文，则控制核启动DMA操作传送该报文给服务器，然后继续查询下一个拼包缓冲区，直到所有拼包缓冲区内的报文均已被DMA传送给服务器。定时器的定时时长取决于控制核的处理速度以及拼包缓冲区的数量，为了保证控制核有足够的时间处理所有拼包缓冲区内的报文，定时时长应大于控制核对所有拼包缓冲区轮询一次的总时间长度。

[0025] 经DMA操作的报文需通过PCI-E通道传送给服务器。PCI-E的传输性能取决于多种因素，其中包括协议开销和负载大小。当TLP协议帧头长度为5个双字时，忽略DLLP和PLP，如果报文长度为64字节，则PCI-E带宽有效利用率为64/(64+20)，约为76.2%；如果报文长度为128字节，此时的带宽有效利用率为128/(128+20)，约为86.5%。可见，报文长度越大，PCI-E带宽有效利用率越高。但是，在实际应用中，为了提高带宽利用率，无限制的增加拼包报文长度并不可取。这是因为，拼包报文长度的增加必然导致多个小报文滞留在拼包缓冲区中，报文传输的实时性变差，因此，要综合考虑上述因素，确定符合实际需求的拼包报文长度。

[0026] 综上所述，通过对小报文进行拼包处理，使得单次传送的有效报文长度增加，减少了DMA操作的次数，提高了PCI-E通道带宽的有效利用率，从而提升了网卡小包的收包速率。

[0027] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。

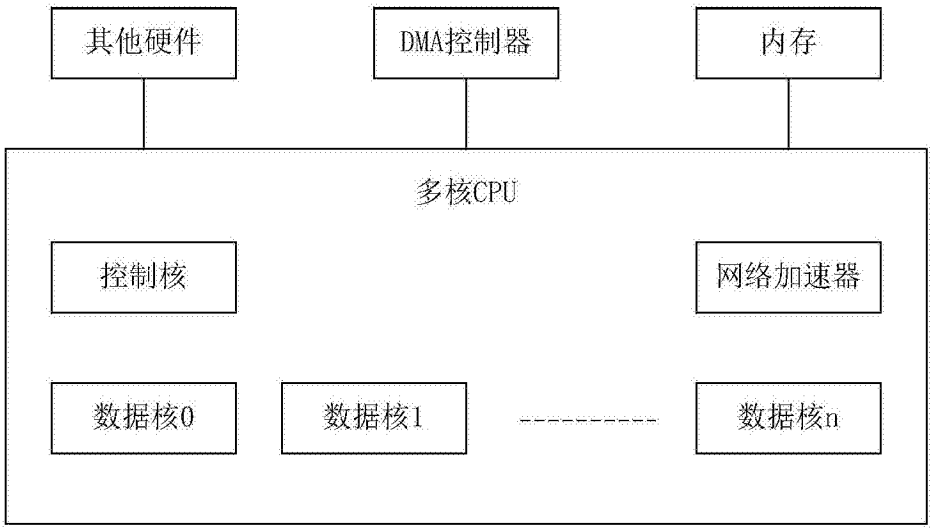


图1

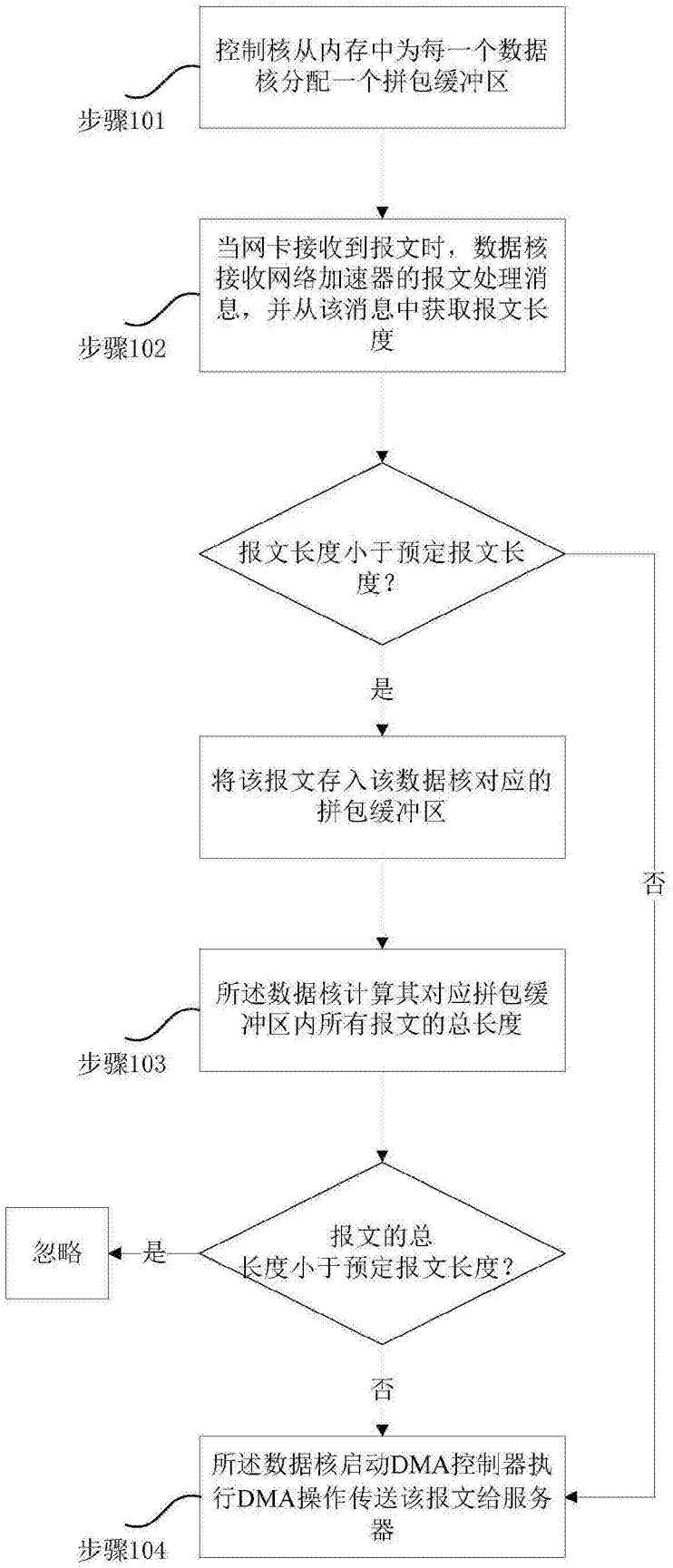

控制核从内存中为每一个数据核分配一个拼包缓冲区
步骤101
当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度
步骤102
报文长度小于预定报文长度？
是
否
将该报文存入该数据核对应的拼包缓冲区
所述数据核计算其对应拼包缓冲区内所有报文的总长度
步骤103
报文的总长度小于预定报文长度？
是
忽略
否
所述数据核启动DMA控制器执行DMA操作传送该报文给服务器
步骤104

图2

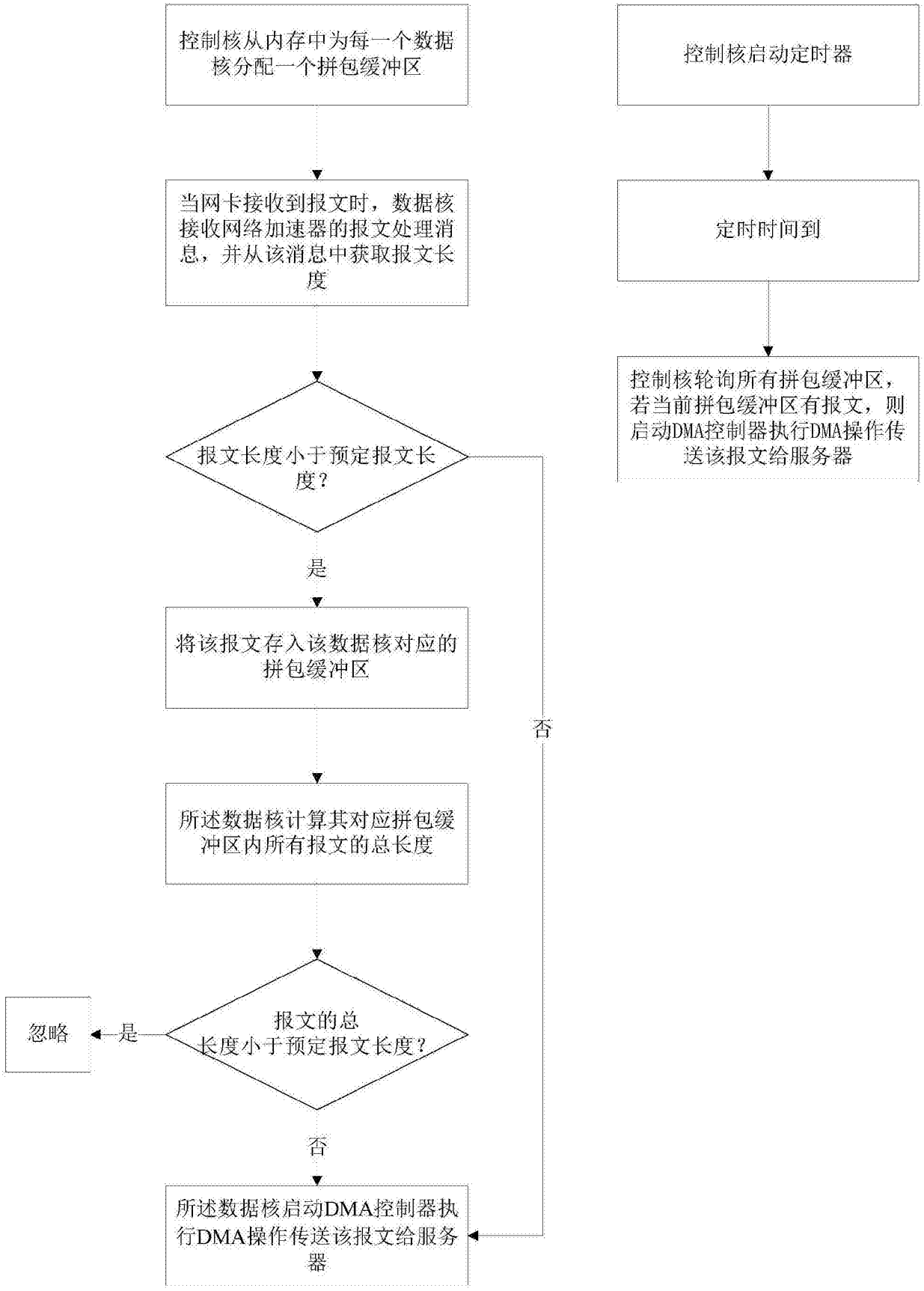

控制核从内存中为每一个数据核分配一个拼包缓冲区
当网卡接收到报文时，数据核接收网络加速器的报文处理消息，并从该消息中获取报文长度
报文长度小于预定报文长度？
是
否
将该报文存入该数据核对应的拼包缓冲区
所述数据核计算其对应拼包缓冲区内所有报文的总长度
报文的总长度小于预定报文长度？
是
忽略
否
所述数据核启动DMA控制器执行DMA操作传送该报文给服务器
控制核启动定时器
定时时间到
控制核轮询所有拼包缓冲区，若当前拼包缓冲区有报文，则启动DMA控制器执行DMA操作传送该报文给服务器

图3