(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 106980582 B
(45)授权公告日 2022.05.13

(21)申请号 201610031757.6
(22)申请日 2016.01.18
(65)同一申请的已公布的文献号
申请公布号 CN 106980582 A
(43)申请公布日 2017.07.25
(73)专利权人 中兴通讯股份有限公司
地址 518057 广东省深圳市南山区科技南路55号
(72)发明人 刘卯银 秦长鹏 戴庆军 牛克强 张翼 舒坦
(74)专利代理机构 北京天昊联合知识产权代理有限公司 11112
专利代理师 姜春咸 冯建基
(51)Int.Cl.
G06F 13/28(2006.01)
G06F 12/0831(2016.01)

(56)对比文件
CN 106980582 A,2017.07.25
CN 103440202 A,2013.12.11
US 8874680 B1,2014.10.28

审查员 丁娴子

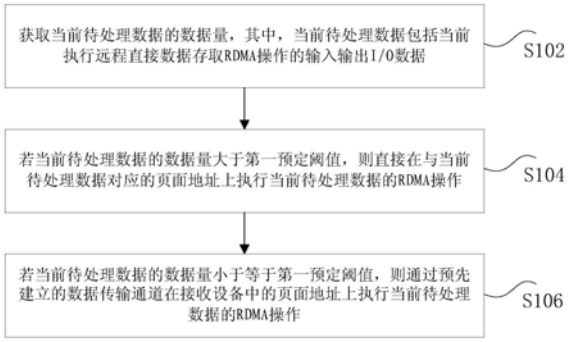
权利要求书5页 说明书21页 附图5页

(54)发明名称

数据处理方法和装置

(57)摘要

本发明提供了一种数据处理方法和装置。该方法包括:获取当前待处理数据的数据量,其中,当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据;若当前待处理数据的数据量大于第一预定阈值,则直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作;若当前待处理数据的数据量小于等于第一预定阈值,则通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。通过本发明,解决了现有技术中RDMA只能通过底层协议栈多次交互,多次拷贝来传输数据的方式执行RDMA操作所导致的数据处理效率较低的问题,进而达到了提高数据处理效率的效果。



CN 106980582 B

1.一种数据处理方法,其特征在于,包括:

获取当前待处理数据的数据量,其中,所述当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据;

若所述当前待处理数据的数据量大于第一预定阈值,则直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作;

若所述当前待处理数据的数据量小于等于所述第一预定阈值,则通过预先建立的数据传输通道在接收设备中的页面地址上执行所述当前待处理数据的所述RDMA操作;

在直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作之前,还包括:

从本地的页面地址缓存池中直接获取所述页面地址,其中,所述页面地址缓存池用于缓存一个或多个页面地址。

2.根据权利要求1所述的方法,其特征在于,在从本地的页面地址缓存池中直接获取所述页面地址之前,包括:

检测所述页面地址缓存池中缓存的页面地址数量;

若所述页面地址数量小于等于第二预定阈值,则通过所述数据传输通道获取新的页面地址。

3.根据权利要求2所述的方法,其特征在于,所述通过所述数据传输通道获取新的页面地址包括:

将用于请求获取所述页面地址的获取请求作为待传输数据通过所述数据传输通道发送给所述接收设备;

获取所述接收设备发送的所述页面地址;

添加所述页面地址到所述页面地址缓存池中。

4.根据权利要求1所述的方法,其特征在于,所述通过预先建立的数据传输通道在接收设备中的所述页面地址上执行所述当前待处理数据的所述RDMA操作包括:

将所述当前待处理数据作为待传输数据通过所述数据传输通道直接发送给所述接收设备,并保存在所述数据传输通道在所述接收设备的通道缓存器中,以使所述接收设备利用所述通道缓存器中的所述当前待处理数据在所述页面地址上执行所述RDMA操作。

5.根据权利要求4所述的方法,其特征在于,所述接收设备利用所述通道缓存器中的所述当前待处理数据在所述页面地址上执行所述RDMA操作包括:

所述接收设备将所述通道缓存器中的所述当前待处理数据拷贝到与所述页面地址对应页面内存中。

6.根据权利要求1所述的方法,其特征在于,在获取当前待处理数据的数据量之前,还包括:

在发送设备与所述接收设备之间建立所述数据传输通道,其中,所述数据传输通道连接所述发送设备中的控制队列与所述接收设备中的环形队列,其中,所述控制队列包括用于控制所述环形队列中数据状态变化的控制信息。

7.根据权利要求6所述的方法,其特征在于,所述在发送设备与所述接收设备之间建立所述数据传输通道包括:

在所述控制队列及所述环形队列中分别设置用于指示队列中数据状态变化的指针,其

中,所述控制队列与所述环形队列中的所述指针所指示的位置同步变化;

其中,所述指针包括:队列头指针、队列尾指针及队列接收指针,其中,所述队列尾指针与所述队列头指针之间的数据用于表示所述接收设备尚未确认接收的数据;所述队列接收指针与所述队列尾指针之间的数据用于表示所述接收设备已确认接收,且尚未处理的数据。

8.根据权利要求7所述的方法,其特征在于,所述控制队列与所述环形队列中的所述指针所指示的位置同步变化包括:

在所述发送设备执行发送操作时,所述控制队列的所述队列头指针将向前移动N个数据字节,并将所述队列头指针所指示的位置将同步到所述接收设备的所述环形队列中,其中,所述N为大于等于1的自然数;

在所述接收设备执行接收操作时,所述环形队列的所述队列尾指针将向前移动M个数据字节,并将所述队列尾指针所指示的位置同步到所述发送设备的所述控制队列中,其中,M≤N,M为大于等于1的自然数;

其中,在所述发送设备的所述控制队列中所述队列尾指针移动到所述队列头指针时,更新所述队列接收指针的位置。

9.一种数据处理方法,其特征在于,包括:

在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时,接收所述发送设备直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的远程直接数据存取RDMA操作,其中,所述当前待处理数据包括当前执行所述RDMA操作的输入输出I/O数据;其中,发送设备是从本地的页面地址缓存池中直接获取所述页面地址的,所述页面地址缓存池用于缓存一个或多个页面地址;

在所述当前待处理数据的数据量小于等于所述第一预定阈值时,接收通过预先建立的数据传输通道在接收设备中的页面地址上执行所述当前待处理数据的所述RDMA操作;

在接收所述发送设备直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作之前,还包括:

接收所述发送设备发送的用于请求获取所述页面地址的获取请求;

响应所述获取请求发送所述页面地址。

10.根据权利要求9所述的方法,其特征在于,所述通过预先建立的数据传输通道在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作包括:

接收所述当前待处理数据;

将所述待处理数据保存在所述数据传输通道在所述接收设备的通道缓存器中;

将所述通道缓存器中的所述当前待处理数据一次拷贝到与所述页面地址对应页面内存中。

11.根据权利要求9所述的方法,其特征在于,还包括:

预先在发送设备与所述接收设备之间建立所述数据传输通道,其中,所述数据传输通道连接所述发送设备中的控制队列与所述接收设备中的环形队列,其中,所述控制队列包括用于控制所述环形队列中数据状态变化的控制信息。

12.根据权利要求11所述的方法,其特征在于,所述在发送设备与所述接收设备之间建立所述数据传输通道包括:

在所述控制队列及所述环形队列中分别设置用于指示队列中数据状态变化的指针，其中，所述控制队列与所述环形队列中的所述指针所指示的位置同步变化；

其中，所述指针包括：队列头指针、队列尾指针及队列接收指针，其中，所述队列尾指针与所述队列头指针之间的数据用于表示所述接收设备尚未确认接收的数据；所述队列接收指针与所述队列尾指针之间的数据用于表示所述接收设备已确认接收，且尚未处理的数据。

13.根据权利要求12所述的方法，其特征在于，所述控制队列与所述环形队列中的所述指针所指示的位置同步变化包括：

在所述发送设备执行发送操作时，所述控制队列的所述队列头指针将向前移动N个数据字节，并将所述队列头指针所指示的位置将同步到所述接收设备的所述环形队列中，其中，所述N为大于等于1的自然数；

在所述接收设备执行接收操作时，所述环形队列的所述队列尾指针将向前移动M个数据字节，并将所述队列尾指针所指示的位置同步到所述发送设备的所述控制队列中，其中，M≤N，M为大于等于1的自然数；

其中，在所述发送设备的所述控制队列中所述队列尾指针移动到所述队列头指针时，更新所述队列接收指针的位置。

14.一种数据处理装置，其特征在于，包括：

第一获取单元，用于获取当前待处理数据的数据量，其中，所述当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据；

第一处理单元，用于在所述当前待处理数据的数据量大于第一预定阈值时，直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作；

第二处理单元，用于在所述当前待处理数据的数据量小于等于所述第一预定阈值时，通过预先建立的数据传输通道在接收设备中的页面地址上执行所述当前待处理数据的所述RDMA操作；

第二获取单元，用于在直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作之前，从本地的页面地址缓存池中直接获取所述页面地址，其中，所述页面地址缓存池用于缓存一个或多个页面地址。

15.根据权利要求14所述的装置，其特征在于，还包括：

检测单元，用于在从本地的页面地址缓存池中直接获取所述页面地址之前，检测所述页面地址缓存池中缓存的页面地址数量；

第三获取单元，用于在所述页面地址数量小于等于第二预定阈值时，通过所述数据传输通道获取新的页面地址。

16.根据权利要求15所述的装置，其特征在于，所述第三获取单元包括：

发送模块，用于将用于请求获取所述页面地址的获取请求作为待传输数据通过所述数据传输通道发送给所述接收设备；

获取模块，用于获取所述接收设备发送的所述页面地址；

添加模块，用于添加所述页面地址到所述页面地址缓存池中。

17.根据权利要求14所述的装置，其特征在于，所述第二处理单元包括：

处理模块，用于将所述当前待处理数据作为待传输数据通过所述数据传输通道直接发

送给所述接收设备，并保存在所述数据传输通道在所述接收设备的通道缓存器中，以使所述接收设备利用所述通道缓存器中的所述当前待处理数据在所述页面地址上执行所述RDMA操作。

18.根据权利要求14所述的装置，其特征在于，还包括：

建立单元，用于在获取当前待处理数据的数据量之前，在发送设备与所述接收设备之间建立所述数据传输通道，其中，所述数据传输通道连接所述发送设备中的控制队列与所述接收设备中的环形队列，其中，所述控制队列包括用于控制所述环形队列中数据状态变化的控制信息。

19.根据权利要求18所述的装置，其特征在于，所述建立单元包括：

设置模块，用于在所述控制队列及所述环形队列中分别设置用于指示队列中数据状态变化的指针，其中，所述控制队列与所述环形队列中的所述指针所指示的位置同步变化；

其中，所述指针包括：队列头指针、队列尾指针及队列接收指针，其中，所述队列尾指针与所述队列头指针之间的数据用于表示所述接收设备尚未确认接收的数据；所述队列接收指针与所述队列尾指针之间的数据用于表示所述接收设备已确认接收，且尚未处理的数据。

20.根据权利要求19所述的装置，其特征在于，所述设置模块通过以下方式控制所述控制队列与所述环形队列中的所述指针所指示的位置同步变化包括：

在所述发送设备执行发送操作时，所述控制队列的所述队列头指针将向前移动N个数据字节，并将所述队列头指针所指示的位置将同步到所述接收设备的所述环形队列中，其中，所述N为大于等于1的自然数；

在所述接收设备执行接收操作时，所述环形队列的所述队列尾指针将向前移动M个数据字节，并将所述队列尾指针所指示的位置同步到所述发送设备的所述控制队列中，其中，M≤N，M为大于等于1的自然数；

其中，在所述发送设备的所述控制队列中所述队列尾指针移动到所述队列头指针时，更新所述队列接收指针的位置。

21.一种数据处理装置，其特征在于，包括：

第一处理单元，用于在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时，接收所述发送设备直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的远程直接数据存取RDMA操作，其中，所述当前待处理数据包括当前执行所述RDMA操作的输入输出I/O数据；其中，发送设备是从本地的页面地址缓存池中直接获取所述页面地址的，所述页面地址缓存池用于缓存一个或多个页面地址；

第二处理单元，用于在所述当前待处理数据的数据量小于等于所述第一预定阈值时，接收通过预先建立的数据传输通道在接收设备中的页面地址上执行所述当前待处理数据的所述RDMA操作；

接收单元，用于在接收所述发送设备直接在与所述当前待处理数据对应的页面地址上执行所述当前待处理数据的所述RDMA操作之前，接收所述发送设备发送的用于请求获取所述页面地址的获取请求；

发送单元，用于响应所述获取请求发送所述页面地址。

22.根据权利要求21所述的装置，其特征在于，所述第二处理单元包括：

接收模块，用于接收所述当前待处理数据；

保存模块，用于将所述待处理数据保存在所述数据传输通道在所述接收设备的通道缓存器中；

拷贝模块，用于将所述通道缓存器中的所述当前待处理数据拷贝到与所述页面地址对应页面内存中。

数据处理方法和装置

技术领域

[0001] 本发明涉及通信领域,具体而言,涉及一种数据处理方法和装置。

背景技术

[0002] 随着用户数据的不断膨胀,信息技术的兴起。各种通信设备间的带宽越来越大。PCIe(PCI-Express)链路作为一种节点内主流的高速传输协议被广泛的应用。在不断提高带宽的同时,PCIe协议逐步开始从节点内中央处理器(CPU,Central Processing Unit)和外部设备之间的互联协议中走出来,利用PCIe的非透明桥(NTB,Non Transparent bridge)技术,PCIe协议可以支持节点间的高速互联,互联的节点间通过NTB进行地址域的隔离。经过NTB的地址映射后,本节点上的DMA引擎通过访问NTB映射过来的虚拟地址,就可以实现对对端的节点上的内存的访问。虽然通过NTB和存储器直接访问(DMA,Direct Memory Acess)技术,PCIe协议物理上实现了直接访问对端节点上的内存,但是这离高效的节点间的数据交互还有一些距离。

[0003] 具体来说,由于PCIe的远程直接数据存取RDMA(RDMA,Remote Direct Memory Acess)操作只在发送节点上进行,因而PCIe在与对端进行交互时,例如,通过DMA引擎访问对端,或向对端发送数据时,就需要确切的获取对端节点上的物理地址,从而实现在源端的一次RDMA操作过程中,就将数据RDMA到对端节点上的内存。也就是说,在源端向对端获取对端节点上执行RDMA操作的物理地址时,还需要在节点间进行多次交互,才能申请到对端节点上的内存。此外,PCIe交换器厂商提供的DEMO软件一般都采用PCIe去模拟一个以太网接口,但是模拟的以太网接口在数据传输时,往往需要多次的数据拷贝,不能充分利用到PCIe协议提供的高带宽。也就是说,采用现有技术中所提供的只能通过多次交互获取对端物理地址的方式执行RDMA操作时,将导致数据处理效率较低的问题。

发明内容

[0004] 本发明提供了一种数据处理方法和装置,以至少解决相关技术中RDMA只能通过底层协议栈多次交互,多次拷贝来传输数据的方式执行RDMA操作所导致的数据处理效率较低的问题。

[0005] 根据本发明的一个方面,提供了一种数据处理方法,包括:获取当前待处理数据的数据量,其中,上述当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据;若上述当前待处理数据的数据量大于第一预定阈值,则直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作;若上述当前待处理数据的数据量小于等于上述第一预定阈值,则通过预先建立的数据传输通道在接收设备中的上述页面地址上执行上述当前待处理数据的上述RDMA操作。

[0006] 可选地,在直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作之前,还包括:从本地的页面地址缓存池中直接获取上述页面地址,其中,上述页面地址缓存池用于缓存一个或多个页面地址。

[0007] 可选地，在从本地的页面地址缓存池中直接获取上述页面地址之前，包括：检测上述页面地址缓存池中缓存的页面地址数量；若上述页面地址数量小于等于第二预定阈值，则通过上述数据传输通道获取新的页面地址。

[0008] 可选地，上述通过上述数据传输通道获取新的页面地址包括：将用于请求获取上述页面地址的获取请求作为待传输数据通过上述数据传输通道发送给上述接收设备；获取上述接收设备发送的上述页面地址；添加上述页面地址到上述页面地址缓存池中。

[0009] 可选地，上述通过预先建立的数据传输通道在接收设备中的上述页面地址上执行上述当前待处理数据的上述RDMA操作包括：将上述当前待处理数据作为待传输数据通过上述数据传输通道直接发送给上述接收设备，并保存在上述数据传输通道在上述接收设备的通道缓存器中，以使上述接收设备利用上述通道缓存器中的上述当前待处理数据在上述页面地址上执行上述RDMA操作。

[0010] 可选地，上述接收设备利用上述通道缓存器中的上述当前待处理数据在上述页面地址上执行上述RDMA操作包括：上述接收设备将上述通道缓存器中的上述当前待处理数据拷贝到与上述页面地址对应页面内存中。

[0011] 可选地，在获取当前待处理数据的数据量之前，还包括：在发送设备与上述接收设备之间建立上述数据传输通道，其中，上述数据传输通道连接上述发送设备中的控制队列与上述接收设备中的环形队列，其中，上述控制队列包括用于控制上述环形队列中数据状态变化的控制信息。

[0012] 可选地，上述在发送设备与上述接收设备之间建立上述数据传输通道包括：在上述控制队列及上述环形队列中分别设置用于指示队列中数据状态变化的指针，其中，上述控制队列与上述环形队列中的上述指针所指示的位置同步变化；其中，上述指针包括：队列头指针、队列尾指针及队列接收指针，其中，上述队列尾指针与上述队列头指针之间的数据用于表示上述接收设备尚未确认接收的数据；上述队列接收指针与上述队列尾指针之间的数据用于表示上述接收设备已确认接收，且尚未处理的数据。

[0013] 可选地，上述控制队列与上述环形队列中的上述指针所指示的位置同步变化包括：在上述发送设备执行发送操作时，上述控制队列的上述队列头指针将向前移动N个数据字节，并将上述队列头指针所指示的位置将同步到上述接收设备的上述环形队列中，其中，上述N为大于等于1的自然数；在上述接收设备执行接收操作时，上述环形队列的上述队列尾指针将向前移动M个数据字节，并将上述队列尾指针所指示的位置同步到上述发送设备的上述控制队列中，其中，$M \leq N$，M为大于等于1的自然数；其中，在上述发送设备的上述控制队列中上述队列尾指针移动到上述队列头指针时，更新上述队列接收指针的位置。

[0014] 根据本发明的另一方面，提供了一种数据处理方法，包括：在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时，接收上述发送设备直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的远程直接数据存取RDMA操作，其中，上述当前待处理数据包括当前执行上述RDMA操作的输入输出I/O数据；在上述当前待处理数据的数据量小于等于上述第一预定阈值时，接收通过预先建立的数据传输通道在接收设备中的上述页面地址上执行上述当前待处理数据的上述RDMA操作。

[0015] 可选地，在接收上述发送设备直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作之前，还包括：接收上述发送设备发送的用于请求

获取上述页面地址的获取请求;响应上述获取请求发送上述页面地址。

[0016] 可选地,上述通过预先建立的数据传输通道在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作包括:接收上述当前待处理数据;将上述待处理数据保存在上述数据传输通道在上述接收设备的通道缓存器中;将上述通道缓存器中的上述当前待处理数据一次拷贝到与上述页面地址对应页面内存中。

[0017] 可选地,还包括:预先在发送设备与上述接收设备之间建立上述数据传输通道,其中,上述数据传输通道连接上述发送设备中的控制队列与上述接收设备中的环形队列,其中,上述控制队列包括用于控制上述环形队列中数据状态变化的控制信息。

[0018] 可选地,上述在发送设备与上述接收设备之间建立上述数据传输通道包括:在上述控制队列及上述环形队列中分别设置用于指示队列中数据状态变化的指针,其中,上述控制队列与上述环形队列中的上述指针所指示的位置同步变化;其中,上述指针包括:队列头指针、队列尾指针及队列接收指针,其中,上述队列尾指针与上述队列头指针之间的数据用于表示上述接收设备尚未确认接收的数据;上述队列接收指针与上述队列尾指针之间的数据用于表示上述接收设备已确认接收,且尚未处理的数据。

[0019] 可选地,上述控制队列与上述环形队列中的上述指针所指示的位置同步变化包括:在上述发送设备执行发送操作时,上述控制队列的上述队列头指针将向前移动N个数据字节,并将上述队列头指针所指示的位置将同步到上述接收设备的上述环形队列中,其中,上述N为大于等于1的自然数;在上述接收设备执行接收操作时,上述环形队列的上述队列尾指针将向前移动M个数据字节,并将上述队列尾指针所指示的位置同步到上述发送设备的上述控制队列中,其中,M≤N,M为大于等于1的自然数;其中,在上述发送设备的上述控制队列中上述队列尾指针移动到上述队列头指针时,更新上述队列接收指针的位置。

[0020] 根据本发明的又一方面,提供了一种数据处理装置,包括:第一获取单元,用于获取当前待处理数据的数据量,其中,上述当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据;第一处理单元,用于在上述当前待处理数据的数据量大于第一预定阈值时,直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作;第二处理单元,用于在上述当前待处理数据的数据量小于等于上述第一预定阈值时,通过预先建立的数据传输通道在接收设备中的上述页面地址上执行上述当前待处理数据的上述RDMA操作。

[0021] 可选地,上述装置还包括:第二获取单元,用于在直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作之前,从本地的页面地址缓存池中直接获取上述页面地址,其中,上述页面地址缓存池用于缓存一个或多个页面地址。

[0022] 可选地,上述装置还包括:检测单元,用于在从本地的页面地址缓存池中直接获取上述页面地址之前,检测上述页面地址缓存池中缓存的页面地址数量;第三获取单元,用于在上述页面地址数量小于等于第二预定阈值时,通过上述数据传输通道获取新的页面地址。

[0023] 可选地,上述第三获取单元包括:发送模块,用于将用于请求获取上述页面地址的获取请求作为待传输数据通过上述数据传输通道发送给上述接收设备;获取模块,用于获取上述接收设备发送的上述页面地址;添加模块,用于添加上述页面地址到上述页面地址缓存池中。

[0024] 可选地，上述第二处理单元包括：处理模块，用于将上述当前待处理数据作为待传输数据通过上述数据传输通道直接发送给上述接收设备，并保存在上述数据传输通道在上述接收设备的通道缓存器中，以使上述接收设备利用上述通道缓存器中的上述当前待处理数据在上述页面地址上执行上述RDMA操作。

[0025] 可选地，还包括：建立单元，用于在获取当前待处理数据的数据量之前，在发送设备与上述接收设备之间建立上述数据传输通道，其中，上述数据传输通道连接上述发送设备中的控制队列与上述接收设备中的环形队列，其中，上述控制队列包括用于控制上述环形队列中数据状态变化的控制信息。

[0026] 可选地，上述建立单元包括：设置模块，用于在上述控制队列及上述环形队列中分别设置用于指示队列中数据状态变化的指针，其中，上述控制队列与上述环形队列中的上述指针所指示的位置同步变化；其中，上述指针包括：队列头指针、队列尾指针及队列接收指针，其中，上述队列尾指针与上述队列头指针之间的数据用于表示上述接收设备尚未确认接收的数据；上述队列接收指针与上述队列尾指针之间的数据用于表示上述接收设备已确认接收，且尚未处理的数据。

[0027] 可选地，上述设置模块通过以下方式控制上述控制队列与上述环形队列中的上述指针所指示的位置同步变化包括：在上述发送设备执行发送操作时，上述控制队列的上述队列头指针将向前移动N个数据字节，并将上述队列头指针所指示的位置将同步到上述接收设备的上述环形队列中，其中，上述N为大于等于1的自然数；在上述接收设备执行接收操作时，上述环形队列的上述队列尾指针将向前移动M个数据字节，并将上述队列尾指针所指示的位置同步到上述发送设备的上述控制队列中，其中，M≤N，M为大于等于1的自然数；其中，在上述发送设备的上述控制队列中上述队列尾指针移动到上述队列头指针时，更新上述队列接收指针的位置。

[0028] 根据本发明的又一方面，提供了一种数据处理装置，包括：第一处理单元，用于在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时，接收上述发送设备直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的远程直接数据存取RDMA操作，其中，上述当前待处理数据包括当前执行上述RDMA操作的输入输出I/O数据；第二处理单元，用于在上述当前待处理数据的数据量小于等于上述第一预定阈值时，接收通过预先建立的数据传输通道在接收设备中的上述页面地址上执行上述当前待处理数据的上述RDMA操作。

[0029] 可选地，上述装置还包括：接收单元，用于在接收上述发送设备直接在与上述当前待处理数据对应的页面地址上执行上述当前待处理数据的上述RDMA操作之前，接收上述发送设备发送的用于请求获取上述页面地址的获取请求；发送单元，用于响应上述获取请求发送上述页面地址。

[0030] 可选地，上述第二处理单元包括：接收模块，用于接收上述当前待处理数据；保存模块，用于将上述待处理数据保存在上述数据传输通道在上述接收设备的通道缓存器中；拷贝模块，用于将上述通道缓存器中的上述当前待处理数据拷贝到与上述页面地址对应页面内存中。

[0031] 通过本发明，通过根据当前待处理数据的数据量选择合理的数据处理方式：在当前待处理数据的数据量较大时，采用直接在页面地址上执行当前待处理数据的RDMA操作，

而无需每次都通过交互获取对应的页面地址，从而达到减少数据交互的目的；在当前待处理数据的数据量较小时，直接通过数据传输通道在接收设备中通过内存拷贝完成对当前待处理数据的RDMA操作，从而实现对数据处理过程的硬件加速的效果。通过根据不同的数据开销，选择合理的数据处理方式，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低，从而实现提高数据处理效率的效果。

附图说明

[0032] 此处所说明的附图用来提供对本发明的进一步理解，构成本申请的一部分，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。在附图中：

[0033] 图1是根据本发明实施例的一种可选的数据处理方法的流程图；

[0034] 图2是根据本发明实施例的一种可选的数据处理方法的示意图；

[0035] 图3是根据本发明实施例的另一种可选的数据处理方法的示意图；

[0036] 图4是根据本发明实施例的一种可选的数据处理方法的应用示意图；

[0037] 图5是根据本发明实施例的另一种可选的数据处理方法的应用示意图；

[0038] 图6是根据本发明实施例的又一种可选的数据处理方法的应用示意图；

[0039] 图7是根据本发明实施例的另一种可选的数据处理方法的流程图；

[0040] 图8是根据本发明实施例的一种可选的数据处理装置的示意图；以及

[0041] 图9是根据本发明实施例的另一种可选的数据处理装置的示意图。

具体实施方式

[0042] 下文中将参考附图并结合实施例来详细说明本发明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

[0043] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。

[0044] 实施例1

[0045] 在本实施例中提供了一种数据处理方法，图1是根据本发明实施例的数据处理方法的流程图，如图1所示，该流程包括如下步骤：

[0046] 步骤S102，获取当前待处理数据的数据量，其中，当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据；

[0047] 步骤S104，若当前待处理数据的数据量大于第一预定阈值，则直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作；

[0048] 步骤S106，若当前待处理数据的数据量小于等于第一预定阈值，则通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。

[0049] 可选地，在本实施例中，上述数据处理方法可以但不限于应用于PCIe(PCI-Express)链路的节点通信过程中，发送设备在获取到当前待处理数据的数据量后，将根据上述数据量选择不同的数据处理方式：在数据量小于等于第一预定阈值时，由于数据开销较小，则可以将当前待处理数据通过预先建立的数据传输通道发送给接收设备，以使接收设备进行内存拷贝，以完成在页面地址对应的通道缓存上对当前待处理数据的远程直接数据存取RDMA操作；在数据量大于第一预定阈值时，由于数据开销较大，则可以在获取目的节

点上的页面地址后，直接在该页面地址上执行当前待处理数据的RDMA操作。其中，当前待处理数据包括当前执行RDMA操作的输入输出I/O数据。通过根据当前待处理数据的数据量选择合理的数据处理方式，从而实现提高数据处理效率的效果，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低。

[0050] 需要说明的是，由于一次发送请求并接收到请求响应的交互时长固定，如果一次传输的数据量较小时，显然将使得数据传输通道的利用率变低，因而，在本实施例中，当数据量(即开销)较小时，就可以直接通过上述数据传输通道将当前待处理数据传输给接收设备，以使接收设备通过一次内存拷贝完成数据处理，从而实现对数据处理过程的硬件加速，而无需预先获取执行RDMA操作的页面地址，也避免了对数据的多次拷贝过程。也就是说，当数据量较小时，数据的处理时长将根据数据的拷贝时长决定。进一步，当数据量(即开销)较大时，执行拷贝所需的时间很长，因而，则采用直接在页面地址上执行当前待处理数据的RDMA操作。

[0051] 可选地，在本实施例中，在直接在页面地址上执行当前待处理数据的RDMA操作之前，还需获取页面地址，其中，上述页面地址的获取方式可以包括但不限于以下至少之一：通过数据传输通道向接收设备发送用于获取页面地址的获取请求、从本地的页面地址缓存池中直接获取已缓存的页面地址。

[0052] 可选地，在本实施例中，上述页面地址缓存池中的页面地址可以通过以下方式获取：检测页面地址缓存池中缓存的页面地址数量；若页面地址数量小于等于第二预定阈值，则通过数据传输通道获取新的页面地址。

[0053] 需要说明的是，在本实施例中，上述页面地址缓存池将根据当前的缓存量，对池中的页面地址进行及时添加更新。从而实现在数据量较大时，避免每次都向接收设备请求获取页面地址所造成的处理延时的问题。进一步，在获取新的页面地址添加到上述页面地址缓存池时，并不影响对当前待处理数据正常的RDMA操作，也就是说，二者可以异步同时进行，从而进一步实现提高数据处理效率的效果。

[0054] 可选地，在本实施例中，上述数据传输通道可以但不限于是基于直接访问对端内存的消息，在节点之间(如发送设备和接收设备之间)建立的跨节点通信通道。其中，上述跨节点建立的上述数据传输通道可以但不限于是基于两侧的数据队列建立。具体而言，在接收设备设置一个环形队列，在发送设备设置一个用于控制环形队列控制队列，其中，控制队列包括用于控制环形队列中数据状态变化的控制信息。也就是说，通过直接控制发送和接收两侧的数据队列，实现对当前待处理数据的传输控制。例如，同步更新发送和接收两侧的数据队列中的数据指针所指示的位置，以达到准确控制当前待处理数据的传输状态。

[0055] 通过本申请提供的实施例，通过根据当前待处理数据的数据量选择合理的数据处理方式：在当前待处理数据的数据量较大时，采用直接在页面地址上执行当前待处理数据的RDMA操作，而无需每次都通过交互获取对应的页面地址，从而达到减少数据交互的目的；在当前待处理数据的数据量较小时，直接通过数据传输通道在接收设备中通过内存拷贝完成对当前待处理数据的RDMA操作，从而实现对数据处理过程的硬件加速的效果。通过根据不同的数据开销，选择合理的数据处理方式，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低，从而实现提高数据处理效率的效果。

[0056] 作为一种可选的方案，在直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作之前，还包括：

[0057] S1，从本地的页面地址缓存池中直接获取页面地址，其中，页面地址缓存池用于缓存一个或多个页面地址。

[0058] 可选地，在本实施例中，上述页面地址缓存池可以但不限于根据不同的业务(也称之为应用)设置多个不同的页面地址缓存池。如图2所示，可以划分为页面地址缓存池202-1至页面地址缓存池202-N。

[0059] 可选地，在本实施例中，上述页面地址缓存池中的页面地址可以但不限于在由对端(即接收设备)获取后，添加更新到本地页面地址缓存池中。其中，页面地址缓存池获取页面地址的方式可以包括但不限于以下至少之一：节点1检测到缓存量低于预定阈值时，通过数据传输通道向节点2请求补充页面地址；节点2通过数据传输通道主动按照预定周期为节点1补充新的页面地址。

[0060] 具体结合以下示例进行说明，如图2所示，发送设备以节点1为例，接收设备以节点2为例，如步骤S206-S208，当节点1向节点2发送数据时，可以直接向页面地址缓存池申请获取页面地址而无需向节点2发送申请获取的请求，从而减少节点1每次获取页面地址的时间，进而达到减少处理延时的效果。

[0061] 进一步，如步骤S202-S204，上述节点1为对端节点2可以设置多个页面地址缓存池，如页面地址缓存池202-1至页面地址缓存池202-N。当检测到一个缓存池中的缓存量低于预定阈值时，则可以通过数据传输通道向对端节点2发送获取请求，以实现由节点2为节点1补充新的页面地址。

[0062] 需要说明的是，在本示例中，上述节点1获取页面地址的过程(即步骤S202-S204)与页面地址缓存池获取页面地址的过程(即步骤S206-S208)并不限于如图所示的顺序，上述两个过程可以但不限异步同时进行，本实施例中对此不作任何限定。

[0063] 通过本申请提供的实施例，通过在本地设置页面地址缓存池，以使发送设备可以直接从本地获取页面地址，并直接在页面地址上进行RDMA操作，从而达到减少对待处理数据的处理延时。

[0064] 作为一种可选的方案，在从本地的页面地址缓存池中直接获取所述页面地址之前，包括：

[0065] S1，检测页面地址缓存池中缓存的页面地址数量；

[0066] S2，若页面地址数量小于等于第二预定阈值，则通过数据传输通道获取新的页面地址。

[0067] 可选地，在本实施例中，上述第二预定阈值可以但不限于根据不同的应用场景设置为不同取值。其中，节点1(即发送设备)可以为节点2设置多个页面地址缓存池，可以设置一个第二预定阈值，即检测所有页面地址缓存池中页面地址的数量的总量是否满足第二预定阈值；也可以为每个页面地址缓存池设置不同取值的第二预定阈值，即分别检测各个页面地址缓存池中页面地址的数量是否满足对应的第二预定阈值，本实施例中对此不作任何限定。

[0068] 可选地，在本实施例中，通过数据传输通道获取新的页面地址包括：S22，将用于请求获取页面地址的获取请求作为待传输数据通过数据传输通道发送给接收设备；S24，获取

接收设备发送的页面地址;S26,添加页面地址到页面地址缓存池中。

[0069] 具体如图2所示,当检测到页面地址缓存池中的页面地址数量较小时,可以执行步骤S202-S204,向对端节点获取页面地址进行补充。具体过程可以参见上述示例,本示例在此不再赘述。

[0070] 通过本申请提供的实施例,通过实时检测页面地址缓存池中的页面地址的数量,实现对页面地址缓存池中的页面地址的及时补充,从而保证在当前待处理数据正常执行RDMA操作的同时,还可以及时缓存新的页面地址,进一步实现提高数据处理效率的效果。

[0071] 作为一种可选的方案,通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作包括:

[0072] S1,将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备,并保存在数据传输通道在接收设备的通道缓存器中,以使接收设备利用通道缓存器中的当前待处理数据在页面地址上执行RDMA操作。

[0073] 可选地,在本实施例中,在执行当前待处理数据的RDMA操作时,还可以不获取确切的页面地址,将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备,并保存在接收设备的通道缓存器中,接收设备的应用会直接从上述通道缓存器中将当前待处理数据读取出来并以此拷贝到对应的页面内存中,进行RDMA操作。

[0074] 可选地,在本实施例中,接收设备利用通道缓存器中的当前待处理数据在页面地址上执行RDMA操作包括:S12,接收设备将通道缓存器中的当前待处理数据拷贝到与页面地址对应页面内存中。

[0075] 通过本申请提供的实施例,在数据量较小时,利用数据传输通道的通道缓存,实现直接在接收设备中完成对当前待处理数据的RDMA操作,从而达到提高数据处理效率的效果。

[0076] 作为一种可选的方案,在获取当前待处理数据的数据量之前,还包括:

[0077] S1,在发送设备与接收设备之间建立数据传输通道,其中,数据传输通道连接发送设备中的控制队列与接收设备中的环形队列,其中,控制队列包括用于控制环形队列中数据状态变化的控制信息。

[0078] 可选地,在本实施例中,在发送设备与接收设备之间建立数据传输通道包括:

[0079] S12,在控制队列及环形队列中分别设置用于指示队列中数据状态变化的指针,其中,控制队列与环形队列中的指针所指示的位置同步变化;

[0080] 其中,上述指针包括:队列头指针、队列尾指针及队列接收指针,其中,队列尾指针与队列头指针之间的数据用于表示接收设备尚未确认接收的数据;队列接收指针与队列尾指针之间的数据用于表示接收设备已确认接收,且尚未处理的数据。

[0081] 可选地,在本实施例中,在控制队列及环形队列中分别设置队列头指针HEAD、队列尾指针TAIL及队列接收指针RECV_TAIL,通过上述指针所指示的位置的变化,实现对待传输数据的传输控制。

[0082] 需要说明的是,发送设备的新消息采用加入到头部的方式进行,接收设备从尾部开始接收。发送设备判定当HEAD等于TAIL时,则队列为空,当HEAD+1等于RECV_TAIL时,则队列为满。

[0083] 通过本申请提供的实施例,通过在发送和接收两侧分别建立对应的数列,实现基

于数列建立数据传输通道,以实现对待处理数据的灵活传输控制。

[0084] 作为一种可选的方案,控制队列与环形队列中的指针所指示的位置同步变化包括:

[0085] S1,在发送设备执行发送操作时,控制队列的队列头指针将向前移动N个数据字节,并将队列头指针所指示的位置将同步到接收设备的环形队列中,其中,N为大于等于1的自然数;

[0086] S2,在接收设备执行接收操作时,环形队列的队列尾指针将向前移动M个数据字节,并将队列尾指针所指示的位置同步到发送设备的控制队列中,其中,M≤N,M为大于等于1的自然数;

[0087] 其中,在发送设备的控制队列中队列尾指针移动到队列头指针时,更新队列接收指针的位置。

[0088] 具体结合以下示例进行说明,如图3所示,图3所示的节点1和节点2之间建立的跨节点数据传输通道是基于对数据队列的控制,具体的交互流程如下:

[0089] S1,节点1向节点2发送消息时,将消息数据写入控制队列后,将控制队列的指针HEAD增加,例如,指针HEAD向前移动5个字节,并将指针HEAD所指示的位置同步给节点2;

[0090] S2,节点2收到节点1的中断信号或者轮询到指针HEAD变化时,进行消息处理,将消息交给节点2中的应用模块处理。消息处理后,节点2将环形队列的指针TAIL增加,例如,指针TAIL向前移动2个数据字节,则将指针TAIL所指示的位置同步给节点1;

[0091] S3,节点1收到节点2的中断信号或者轮询到指针TAIL变化时,说明已发送待确认的队列里有数据,节点1处理已发送待确认队列里的消息,通知节点1中的应用模块消息处理完成,则节点1将更新指针RECV_TAIL的位置,即释放指针RECV_TAIL的当前位置。

[0092] 需要说明的是,图3所示的实线表示在本侧执行处理后得到的位置,虚线表示在对侧执行处理后同步得到的位置。

[0093] 通过本申请提供的实施例,同步两侧队列中指针所指示的位置,从而实现准确控制所传输的数据的传输状态,达到作为数据传输通道准确完成数据传输的目的。

[0094] 可选地,在本实施例中,上述数据处理方法可以但不限于应用于如图4所示的系统中,系统中位于发送设备的通信模块402分别与对应的业务模块1至业务模块N相连,位于接收设备的通信模块404也分别与对应的业务模块1至业务模块N相连,发送设备和接收设备通过网络406实现RDMA操作。

[0095] 作为一种可选的实施方式,如图5所示的双节点系统,节点1和节点2通过PCIe链路直连。节点间通过NTB(Non-Transport)非透明桥进行PCIe的地址域隔离。使用节点上的CPU带的DMA引擎,这个DMA引擎可以通过PCIe交换网络直接访问其它节点上的内存。

[0096] 在系统上电时,节点1和节点2使用事先约定好的固定地址的访问进行访问,获取对端的跨节点消息队列的状态。各个业务模块会向通信模块注册回调函数,注册的回调函数包含:

[0097] 1)业务模块对应的消息处理函数将执行以下步骤:

[0098] S1,当通信模块收到目的地是该业务模块(通过模块号或者端口号来区分)时,调用业务模块注册的消息处理函数处理消息接收。

[0099] 2)业务模块的页面内存申请、释放函数将执行以下步骤:

[0100] S1,通信模块会调用业务模块的页面内存申请函数申请业务模块的内存,并填充到对端节点上的页面地址缓存池中。当对端节点故障或者离线时,通信模块会调用业务模块的页面释放函数释放分配到目的节点上的内存。

[0101] 3)消息发送结果通知函数将执行以下步骤:

[0102] S1,消息发送给对端,对端回了响应后,通信模块会调用业务模块注册的消息发送结果通知函数来通知业务模块,消息已经送达。

[0103] 业务模块注册成功回调函数,通信模块完成节点间的信息交互后,一次消息的交互过程如下:

[0104] S1,业务模块1调用通信模块的接口向对端节点上的业务模块2发送消息;

[0105] S2,通信模块接收到业务模块1的消息,根据本端保存的对端的队列信息,将消息发送到对端节点上的消息内存中;

[0106] S3,累计发送1个或者多个消息后,待发送队列为空时,节点1上的通信模块更新节点2上队列的HEAD指针;

[0107] S4,节点2检测到HEAD指针变化,将队列中的消息拷贝出来,根据消息的目的模块,交给业务模块2进行处理;

[0108] S5,节点2上的接收队列处理为空的时候,节点2更新节点1上的TAIL指针;

[0109] S6,节点1检测到TAIL指针变化时,处理节点1上的已发送待确认队列,直到队列为空,更新RECV_TAIL指针。

[0110] 作为另一种可选的实施方式,如图6所示的多节点系统,节点1和节点7通过PCIe交换机进行连接。为了保持冗余,防止单点故障,PCIe交换机通常采用冗余的双星架构进行配置。每个节点出两条链路分别连接到两台PCIe交换机。本实施例中以单链路的情况为例进行具体描述。节点间通过NTB(Non-Transport)非透明桥进行PCIe的地址域隔离,PCIe交换机会为每个节点分配不同的地址域。使用PCIe交换器上带的DMA引擎进行RDMA操作,这个DMA引擎可以通过PCIe交换网络直接访问其它节点上的内存。

[0111] 节点间使用带外(以太网)链路和集群内的其它节点进行信息交互,获取各个节点上跨节点的生产者消费者队列的地址信息,以及节点的状态。各个业务模块会向通信模块注册回调函数,注册的回调函数包含:

[0112] 1)业务模块对应的消息处理函数将执行以下步骤:

[0113] S1,当通信模块收到目的地是该业务模块(通过模块号或者端口号来区分)时,调用业务模块注册的消息处理函数处理消息接收。

[0114] 2)业务模块的页面内存申请、释放函数执行以下步骤:

[0115] S1,通信模块会调用业务模块的页面内存申请函数申请业务模块的内存,并填充到对端节点上的页面内存缓冲区,图3中的页面池。当对端节点故障或者离线时,通信模块会调用业务模块的页面释放函数释放分配到目的节点上的内存。

[0116] 3)消息发送结果通知函数执行以下步骤:

[0117] S1,消息发送给对端,对端回了响应后,通信模块会调用业务模块注册的消息发送结果通知函数来通知业务模块,消息已经送达。

[0118] 业务模块注册成功回调函数,通信模块完成节点间的信息交互后,一次带页面的消息交互过程如下:

[0119] S1,节点1上的业务模块1调用通信模块的接口向节点7上的业务模块2发送带页面数据的消息;

[0120] S2,通信模块接收到业务模块1的消息,向本地保存的节点7上的业务模块2的页面地址缓存池申请页面地址;

[0121] S3,将页面数据写到对端的页面内存,并将页面地址,消息发送到对端节点上的消息内存中;

[0122] S4,累计发送1个或者多个消息后,待发送队列为空时,节点1上的通信模块更新节点7上队列的HEAD指针;

[0123] S5,节点7检测到HEAD指针变化,将队列中的消息拷贝出来,根据消息的目的模块,交给业务模块2进行处理;

[0124] S6,节点7上的接收队列处理为空的时候,节点7更新节点1上的TAIL指针;

[0125] S7,节点1检测到TAIL指针变化时,处理节点1上的已发送待确认队列,直到队列为空,更新RECV_TAIL指针。

[0126] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。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(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,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的方法。

[0127] 实施例2

[0128] 在本实施例中提供了一种数据处理方法,图7是根据本发明实施例的数据处理方法的流程图,如图7所示,该流程包括如下步骤:

[0129] S702,在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时,接收发送设备直接在与当前待处理数据对应的页面地址上执行当前待处理数据的远程直接数据存取RDMA操作,其中,当前待处理数据包括当前执行RDMA操作的输入输出I/O数据;

[0130] S704,在当前待处理数据的数据量小于等于第一预定阈值时,接收通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。

[0131] 可选地,在本实施例中,上述数据处理方法可以但不限于应用于PCIe(PCI-Express)链路的节点通信过程中,发送设备在获取到当前待处理数据的数据量后,将根据上述数据量选择不同的数据处理方式:在数据量小于等于第一预定阈值时,由于数据开销较小,则可以将当前待处理数据通过预先建立的数据传输通道发送给接收设备,以使接收设备进行内存拷贝,以完成在页面地址对应的通道缓存上对当前待处理数据的远程直接数据存取RDMA操作;在数据量大于第一预定阈值时,由于数据开销较大,则可以在获取目的节点上的页面地址后,直接在该页面地址上执行当前待处理数据的RDMA操作。其中,当前待处理数据包括当前执行RDMA操作的输入输出I/O数据。通过根据当前待处理数据的数据量选择合理的数据处理方式,从而实现提高数据处理效率的效果,以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低。

[0132] 需要说明的是,由于一次发送请求并接收到请求响应的交互时长固定,如果一次传输的数据量较小时,显然将使得数据传输通道的利用率变低,因而,在本实施例中,当数

据量(即开销)较小时,就可以直接通过上述数据传输通道将当前待处理数据传输给接收设备,以使接收设备通过一次内存拷贝完成数据处理,从而实现对数据处理过程的硬件加速,而无需预先获取执行RDMA操作的页面地址,也避免了对数据的多次拷贝过程。也就是说,当数据量较小时,数据的处理时长将根据数据的拷贝时长决定。进一步,当数据量(即开销)较大时,执行拷贝所需的时间很长,因而,则采用直接在页面地址上执行当前待处理数据的RDMA操作。

[0133] 可选地,在本实施例中,上述数据传输通道可以但不限于是基于直接访问对端内存的消息,在节点之间(如发送设备和接收设备之间)建立的跨节点通信通道。其中,上述跨节点建立的上述数据传输通道可以但不限于是基于两侧的数据队列建立。具体而言,在接收设备设置一个环形队列,在发送设备设置一个用于控制环形队列控制队列,其中,控制队列包括用于控制环形队列中数据状态变化的控制信息。也就是说,通过直接控制发送和接收两侧的数据队列,实现对当前待处理数据的传输控制。例如,同步更新发送和接收两侧的数据队列中的数据指针所指示的位置,以达到准确控制当前待处理数据的传输状态。

[0134] 通过本申请提供的实施例,通过根据当前待处理数据的数据量选择合理的数据处理方式:在当前待处理数据的数据量较大时,采用直接在页面地址上执行当前待处理数据的RDMA操作,而无需每次都通过交互获取对应的页面地址,从而达到减少数据交互的目的;在当前待处理数据的数据量较小时,直接通过数据传输通道在接收设备中通过内存拷贝完成对当前待处理数据的RDMA操作,从而实现对数据处理过程的硬件加速的效果。通过根据不同的数据开销,选择合理的数据处理方式,以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低,从而实现提高数据处理效率的效果。

[0135] 作为一种可选的方案,在接收发送设备直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作之前,还包括:

[0136] S1,接收发送设备发送的用于请求获取页面地址的获取请求;

[0137] S2,响应获取请求发送页面地址。

[0138] 通过本申请提供的实施例,响应发送设备发送的获取请求,向发送设备发送页面地址,从而保证在当前待处理数据正常执行RDMA操作的同时,还可以及时向发送设备缓存新的页面地址,进一步实现提高数据处理效率的效果。

[0139] 作为一种可选的方案,通过预先建立的数据传输通道在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作包括:

[0140] S1,接收当前待处理数据;

[0141] S2,将待处理数据保存在数据传输通道在接收设备的通道缓存器中;

[0142] S3,将通道缓存器中的当前待处理数据一次拷贝到与页面地址对应页面内存中。

[0143] 可选地,在本实施例中,在执行当前待处理数据的RDMA操作时,还可以不获取确切的页面地址,将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备,并保存在接收设备的通道缓存器中,接收设备的应用会直接从上述通道缓存器中将当前待处理数据读取出来并以此拷贝到对应的页面内存中,进行RDMA操作。

[0144] 通过本申请提供的实施例,在数据量较小时,利用数据传输通道的通道缓存,实现直接在接收设备中完成对当前待处理数据的RDMA操作,从而达到提高数据处理效率的效

果。

[0145] 作为一种可选的方案,还包括:

[0146] S1,预先在发送设备与接收设备之间建立数据传输通道,其中,数据传输通道连接发送设备中的控制队列与接收设备中的环形队列,其中,控制队列包括用于控制环形队列中数据状态变化的控制信息。

[0147] 可选地,在本实施例中,在发送设备与接收设备之间建立数据传输通道包括:

[0148] S1,在控制队列及环形队列中分别设置用于指示队列中数据状态变化的指针,其中,控制队列与环形队列中的指针所指示的位置同步变化;

[0149] 其中,指针包括:队列头指针、队列尾指针及队列接收指针,其中,队列尾指针与队列头指针之间的数据用于表示接收设备尚未确认接收的数据;队列接收指针与队列尾指针之间的数据用于表示接收设备已确认接收,且尚未处理的数据。

[0150] 可选地,在本实施例中,在控制队列及环形队列中分别设置队列头指针HEAD、队列尾指针TAIL及队列接收指针RECV_TAIL,通过上述指针所指示的位置的变化,实现对待传输数据的传输控制。

[0151] 需要说明的是,发送设备的新消息采用加入到头部的方式进行,接收设备从尾部开始接收。发送设备判定当HEAD等于TAIL时,则队列为空,当HEAD+1等于RECV_TAIL时,则队列为满。

[0152] 通过本申请提供的实施例,通过在发送和接收两侧分别建立对应的数列,实现基于数列建立数据传输通道,以实现对待处理数据的灵活传输控制。

[0153] 作为一种可选的方案,控制队列与环形队列中的指针所指示的位置同步变化包括:

[0154] S1,在发送设备执行发送操作时,控制队列的队列头指针将向前移动N个数据字节,并将队列头指针所指示的位置将同步到接收设备的环形队列中,其中,N为大于等于1的自然数;

[0155] S2,在接收设备执行接收操作时,环形队列的队列尾指针将向前移动M个数据字节,并将队列尾指针所指示的位置同步到发送设备的控制队列中,其中,M≤N,M为大于等于1的自然数;

[0156] 其中,在发送设备的控制队列中队列尾指针移动到队列头指针时,更新队列接收指针的位置。

[0157] 具体结合以下示例进行说明,如图3所示,图3所示的节点1和节点2之间建立的跨节点数据传输通道是基于对数据队列的控制,具体的交互流程如下:

[0158] S1,节点1向节点2发送消息时,将消息数据写入控制队列后,将控制队列的指针HEAD增加,例如,指针HEAD向前移动5个字节,并将指针HEAD所指示的位置同步给节点2;

[0159] S2,节点2收到节点1的中断信号或者轮询到指针HEAD变化时,进行消息处理,将消息交给节点2中的应用模块处理。消息处理后,节点2将环形队列的指针TAIL增加,例如,指针TAIL向前移动2个数据字节,则将指针TAIL所指示的位置同步给节点1;

[0160] S3,节点1收到节点2的中断信号或者轮询到指针TAIL变化时,说明已发送待确认的队列里有数据,节点1处理已发送待确认队列里的消息,通知节点1中的应用模块消息处理完成,则节点1将更新指针RECV_TAIL的位置,即释放指针RECV_TAIL的当前位置。

[0161] 需要说明的是，图3所示的实线表示在本侧执行处理后得到的位置，虚线表示在对侧执行处理后同步得到的位置。

[0162] 通过本申请提供的实施例，同步两侧队列中指针所指示的位置，从而实现准确控制所传输的数据的传输状态，达到作为数据传输通道准确完成数据传输的目的。

[0163] 实施例3

[0164] 在本实施例中提供了一种数据处理装置，图8是根据本发明实施例的数据处理装置的示意图，如图8所示，该装置包括：

[0165] 1)第一获取单元802，用于获取当前待处理数据的数据量，其中，当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据；

[0166] 2)第一处理单元804，用于在当前待处理数据的数据量大于第一预定阈值时，直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作；

[0167] 3)第二处理单元806，用于在当前待处理数据的数据量小于等于第一预定阈值时，通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。

[0168] 可选地，在本实施例中，上述数据处理方法可以但不限于应用于PCIe(PCI-Express)链路的节点通信过程中，发送设备在获取到当前待处理数据的数据量后，将根据上述数据量选择不同的数据处理方式：在数据量小于等于第一预定阈值时，由于数据开销较小，则可以将当前待处理数据通过预先建立的数据传输通道发送给接收设备，以使接收设备进行内存拷贝，以完成在页面地址对应的通道缓存上对当前待处理数据的远程直接数据存取RDMA操作；在数据量大于第一预定阈值时，由于数据开销较大，则可以在获取目的节点上的页面地址后，直接在该页面地址上执行当前待处理数据的RDMA操作。其中，当前待处理数据包括当前执行RDMA操作的输入输出I/O数据。通过根据当前待处理数据的数据量选择合理的数据处理方式，从而实现提高数据处理效率的效果，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低。

[0169] 需要说明的是，由于一次发送请求并接收到请求响应的交互时长固定，如果一次传输的数据量较小时，显然将使得数据传输通道的利用率变低，因而，在本实施例中，当数据量(即开销)较小时，就可以直接通过上述数据传输通道将当前待处理数据传输给接收设备，以使接收设备通过一次内存拷贝完成数据处理，从而实现对数据处理过程的硬件加速，而无需预先获取执行RDMA操作的页面地址，也避免了对数据的多次拷贝过程。也就是说，当数据量较小时，数据的处理时长将根据数据的拷贝时长决定。进一步，当数据量(即开销)较大时，执行拷贝所需的时间很长，因而，则采用直接在页面地址上执行当前待处理数据的RDMA操作。

[0170] 可选地，在本实施例中，在直接在页面地址上执行当前待处理数据的RDMA操作之前，还需获取页面地址，其中，上述页面地址的获取方式可以包括但不限于以下至少之一：通过数据传输通道向接收设备发送用于获取页面地址的获取请求、从本地的页面地址缓存池中直接获取已缓存的页面地址。

[0171] 可选地，在本实施例中，上述页面地址缓存池中的页面地址可以通过以下方式获取：检测页面地址缓存池中缓存的页面地址数量；若页面地址数量小于等于第二预定阈值，则通过数据传输通道获取新的页面地址。

[0172] 需要说明的是，在本实施例中，上述页面地址缓存池将根据当前的缓存量，对池中的页面地址进行及时添加更新。从而实现在数据量较大时，避免每次都向接收设备请求获取页面地址所造成的处理延时的问题。进一步，在获取新的页面地址添加到上述页面地址缓存池时，并不影响对当前待处理数据正常的RDMA操作，也就是说，二者可以异步同时进行，从而进一步实现提高数据处理效率的效果。

[0173] 可选地，在本实施例中，上述数据传输通道可以但不限于是基于直接访问对端内存的消息，在节点之间(如发送设备和接收设备之间)建立的跨节点通信通道。其中，上述跨节点建立的上述数据传输通道可以但不限于是基于两侧的数据队列建立。具体而言，在接收设备设置一个环形队列，在发送设备设置一个用于控制环形队列控制队列，其中，控制队列包括用于控制环形队列中数据状态变化的控制信息。也就是说，通过直接控制发送和接收两侧的数据队列，实现对当前待处理数据的传输控制。例如，同步更新发送和接收两侧的数据队列中的数据指针所指示的位置，以达到准确控制当前待处理数据的传输状态。

[0174] 通过本申请提供的实施例，通过根据当前待处理数据的数据量选择合理的数据处理方式：在当前待处理数据的数据量较大时，采用直接在页面地址上执行当前待处理数据的RDMA操作，而无需每次都通过交互获取对应的页面地址，从而达到减少数据交互的目的；在当前待处理数据的数据量较小时，直接通过数据传输通道在接收设备中通过内存拷贝完成对当前待处理数据的RDMA操作，从而实现对数据处理过程的硬件加速的效果。通过根据不同的数据开销，选择合理的数据处理方式，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低，从而实现提高数据处理效率的效果。

[0175] 作为一种可选的方案，还包括：

[0176] 1)第二获取单元，用于在直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作之前，从本地的页面地址缓存池中直接获取页面地址，其中，页面地址缓存池用于缓存一个或多个页面地址。

[0177] 可选地，在本实施例中，上述页面地址缓存池可以但不限于根据不同的业务(也称之为应用)设置多个不同的页面地址缓存池。如图2所示，可以划分为页面地址缓存池202-1至页面地址缓存池202-N。

[0178] 可选地，在本实施例中，上述页面地址缓存池中的页面地址可以但不限于在由对端(即接收设备)获取后，添加更新到本地页面地址缓存池中。其中，页面地址缓存池获取页面地址的方式可以包括但不限于以下至少之一：节点1检测到缓存量低于预定阈值时，通过数据传输通道向节点2请求补充页面地址；节点2通过数据传输通道主动按照预定周期为节点1补充新的页面地址。

[0179] 具体结合以下示例进行说明，如图2所示，发送设备以节点1为例，接收设备以节点2为例，如步骤S206-S208，当节点1向节点2发送数据时，可以直接向页面地址缓存池申请获取页面地址而无需向节点2发送申请获取的请求，从而减少节点1每次获取页面地址的时间，进而达到减少处理延时的效果。

[0180] 进一步，如步骤S202-S204，上述节点1为对端节点2可以设置多个页面地址缓存池，如页面地址缓存池202-1至页面地址缓存池202-N。当检测到一个缓存池中的缓存量低于预定阈值时，则可以通过数据传输通道向对端节点2发送获取请求，以实现由节点2为节

点1补充新的页面地址。

[0181] 需要说明的是，在本示例中，上述节点1获取页面地址的过程(即步骤S202-S204)与页面地址缓存池获取页面地址的过程(即步骤S206-S208)并不限于如图所示的顺序，上述两个过程可以但不限异步同时进行，本实施例中对此不作任何限定。

[0182] 通过本申请提供的实施例，通过在本地设置页面地址缓存池，以使发送设备可以直接从本地获取页面地址，并直接在页面地址上进行RDMA操作，从而达到减少对待处理数据的处理延时。

[0183] 作为一种可选的方案，还包括：

[0184] 1)检测单元，用于在从本地的页面地址缓存池中直接获取页面地址之前，检测页面地址缓存池中缓存的页面地址数量；

[0185] 2)第三获取单元，用于在页面地址数量小于等于第二预定阈值时，通过数据传输通道获取新的页面地址。

[0186] 可选地，在本实施例中，上述第二预定阈值可以但不限于根据不同的应用场景设置为不同取值。其中，节点1(即发送设备)可以为节点2设置多个页面地址缓存池，可以设置一个第二预定阈值，即检测所有页面地址缓存池中页面地址的数量的总量是否满足第二预定阈值；也可以为每个页面地址缓存池设置不同取值的第二预定阈值，即分别检测各个页面地址缓存池中页面地址的数量是否满足对应的第二预定阈值，本实施例中对此不作任何限定。

[0187] 可选地，在本实施例中，第三获取单元包括：1)发送模块，用于将用于请求获取页面地址的获取请求作为待传输数据通过数据传输通道发送给接收设备；2)获取模块，用于获取接收设备发送的页面地址；3)添加模块，用于添加页面地址到页面地址缓存池中。

[0188] 具体如图2所示，当检测到页面地址缓存池中的页面地址数量较小时，可以执行步骤S202-S204，向对端节点获取页面地址进行补充。具体过程可以参见上述示例，本示例在此不再赘述。

[0189] 通过本申请提供的实施例，通过实时检测页面地址缓存池中的页面地址的数量，实现对页面地址缓存池中的页面地址的及时补充，从而保证在当前待处理数据正常执行RDMA操作的同时，还可以及时缓存新的页面地址，进一步实现提高数据处理效率的效果。

[0190] 作为一种可选的方案，第二处理单元包括：

[0191] 1)处理模块，用于将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备，并保存在数据传输通道在接收设备的通道缓存器中，以使接收设备利用通道缓存器中的当前待处理数据在页面地址上执行RDMA操作。

[0192] 可选地，在本实施例中，在执行当前待处理数据的RDMA操作时，还可以不获取确切的页面地址，将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备，并保存在接收设备的通道缓存器中，接收设备的应用会直接从上述通道缓存器中将当前待处理数据读取出来并以此拷贝到对应的页面内存中，进行RDMA操作。

[0193] 通过本申请提供的实施例，在数据量较小时，利用数据传输通道的通道缓存，实现直接在接收设备中完成对当前待处理数据的RDMA操作，从而达到提高数据处理效率的效果。

[0194] 作为一种可选的方案，还包括：

[0195] 1)建立单元,用于在获取当前待处理数据的数据量之前,在发送设备与接收设备之间建立数据传输通道,其中,数据传输通道连接发送设备中的控制队列与接收设备中的环形队列,其中,控制队列包括用于控制环形队列中数据状态变化的控制信息。

[0196] 可选地,在本实施例中,建立单元包括:

[0197] (1)设置模块,用于在控制队列及环形队列中分别设置用于指示队列中数据状态变化的指针,其中,控制队列与环形队列中的指针所指示的位置同步变化;

[0198] 其中,指针包括:队列头指针、队列尾指针及队列接收指针,其中,队列尾指针与队列头指针之间的数据用于表示接收设备尚未确认接收的数据;队列接收指针与队列尾指针之间的数据用于表示接收设备已确认接收,且尚未处理的数据。

[0199] 可选地,在本实施例中,在控制队列及环形队列中分别设置队列头指针HEAD、队列尾指针TAIL及队列接收指针RECV_TAIL,通过上述指针所指示的位置的变化,实现对待传输数据的传输控制。

[0200] 需要说明的是,发送设备的新消息采用加入到头部的方式进行,接收设备从尾部开始接收。发送设备判定当HEAD等于TAIL时,则队列为空,当HEAD+1等于RECV_TAIL时,则队列为满。

[0201] 通过本申请提供的实施例,通过在发送和接收两侧分别建立对应的数列,实现基于数列建立数据传输通道,以实现对待处理数据的灵活传输控制。

[0202] 作为一种可选的方案,设置模块通过以下方式控制控制队列与环形队列中的指针所指示的位置同步变化包括:

[0203] 1)在发送设备执行发送操作时,控制队列的队列头指针将向前移动N个数据字节,并将队列头指针所指示的位置将同步到接收设备的环形队列中,其中,N为大于等于1的自然数;

[0204] 2)在接收设备执行接收操作时,环形队列的队列尾指针将向前移动M个数据字节,并将队列尾指针所指示的位置同步到发送设备的控制队列中,其中,M≤N,M为大于等于1的自然数;

[0205] 其中,在发送设备的控制队列中队列尾指针移动到队列头指针时,更新队列接收指针的位置。

[0206] 具体结合以下示例进行说明,如图3所示,图3所示的节点1和节点2之间建立的跨节点数据传输通道是基于对数据队列的控制,具体的交互流程如下:

[0207] S1,节点1向节点2发送消息时,将消息数据写入控制队列后,将控制队列的指针HEAD增加,例如,指针HEAD向前移动5个字节,并将指针HEAD所指示的位置同步给节点2;

[0208] S2,节点2收到节点1的中断信号或者轮询到指针HEAD变化时,进行消息处理,将消息交给节点2中的应用模块处理。消息处理后,节点2将环形队列的指针TAIL增加,例如,指针TAIL向前移动2个数据字节,则将指针TAIL所指示的位置同步给节点1;

[0209] S3,节点1收到节点2的中断信号或者轮询到指针TAIL变化时,说明已发送待确认的队列里有数据,节点1处理已发送待确认队列里的消息,通知节点1中的应用模块消息处理完成,则节点1将更新指针RECV_TAIL的位置,即释放指针RECV_TAIL的当前位置。

[0210] 需要说明的是,图3所示的实线表示在本侧执行处理后得到的位置,虚线表示在对侧执行处理后同步得到的位置。

[0211] 通过本申请提供的实施例，同步两侧队列中指针所指示的位置，从而实现准确控制所传输的数据的传输状态，达到作为数据传输通道准确完成数据传输的目的。

[0212] 实施例4

[0213] 在本实施例中提供了一种数据处理装置，图9是根据本发明实施例的数据处理装置的示意图，如图9所示，该装置包括：

[0214] 1)第一处理单元902，用于在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时，接收发送设备直接在与当前待处理数据对应的页面地址上执行当前待处理数据的远程直接数据存取RDMA操作，其中，当前待处理数据包括当前执行RDMA操作的输入输出I/O数据；

[0215] 2)第二处理单元904，用于在当前待处理数据的数据量小于等于第一预定阈值时，接收通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。

[0216] 可选地，在本实施例中，上述数据处理方法可以但不限于应用于PCIe(PCI-Express)链路的节点通信过程中，发送设备在获取到当前待处理数据的数据量后，将根据上述数据量选择不同的数据处理方式：在数据量小于等于第一预定阈值时，由于数据开销较小，则可以将当前待处理数据通过预先建立的数据传输通道发送给接收设备，以使接收设备进行内存拷贝，以完成在页面地址对应的通道缓存上对当前待处理数据的远程直接数据存取RDMA操作；在数据量大于第一预定阈值时，由于数据开销较大，则可以在获取目的节点上的页面地址后，直接在该页面地址上执行当前待处理数据的RDMA操作。其中，当前待处理数据包括当前执行RDMA操作的输入输出I/O数据。通过根据当前待处理数据的数据量选择合理的数据处理方式，从而实现提高数据处理效率的效果，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低。

[0217] 需要说明的是，由于一次发送请求并接收到请求响应的交互时长固定，如果一次传输的数据量较小时，显然将使得数据传输通道的利用率变低，因而，在本实施例中，当数据量(即开销)较小时，就可以直接通过上述数据传输通道将当前待处理数据传输给接收设备，以使接收设备通过一次内存拷贝完成数据处理，从而实现对数据处理过程的硬件加速，而无需预先获取执行RDMA操作的页面地址，也避免了对数据的多次拷贝过程。也就是说，当数据量较小时，数据的处理时长将根据数据的拷贝时长决定。进一步，当数据量(即开销)较大时，执行拷贝所需的时间很长，因而，则采用直接在页面地址上执行当前待处理数据的RDMA操作。

[0218] 可选地，在本实施例中，上述数据传输通道可以但不限于是基于直接访问对端内存的消息，在节点之间(如发送设备和接收设备之间)建立的跨节点通信通道。其中，上述跨节点建立的上述数据传输通道可以但不限于是基于两侧的数据队列建立。具体而言，在接收设备设置一个环形队列，在发送设备设置一个用于控制环形队列控制队列，其中，控制队列包括用于控制环形队列中数据状态变化的控制信息。也就是说，通过直接控制发送和接收两侧的数据队列，实现对当前待处理数据的传输控制。例如，同步更新发送和接收两侧的数据队列中的数据指针所指示的位置，以达到准确控制当前待处理数据的传输状态。

[0219] 通过本申请提供的实施例，通过根据当前待处理数据的数据量选择合理的数据处理方式：在当前待处理数据的数据量较大时，采用直接在页面地址上执行当前待处理数据

的RDMA操作，而无需每次都通过交互获取对应的页面地址，从而达到减少数据交互的目的；在当前待处理数据的数据量较小时，直接通过数据传输通道在接收设备中通过内存拷贝完成对当前待处理数据的RDMA操作，从而实现对数据处理过程的硬件加速的效果。通过根据不同的数据开销，选择合理的数据处理方式，以克服现有技术中只能通过多次交互获取对端物理地址的方式执行RDMA操作所所导致的数据处理效率较低，从而实现提高数据处理效率的效果。

[0220] 作为一种可选的方案，还包括：

[0221] 1)接收单元，用于在接收发送设备直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作之前，接收发送设备发送的用于请求获取页面地址的获取请求；

[0222] 2)发送单元，用于响应获取请求发送页面地址。

[0223] 通过本申请提供的实施例，响应发送设备发送的获取请求，向发送设备发送页面地址，从而保证在当前待处理数据正常执行RDMA操作的同时，还可以及时向发送设备缓存新的页面地址，进一步实现提高数据处理效率的效果。

[0224] 作为一种可选的方案，第二处理单元包括：

[0225] 1)接收模块，用于接收当前待处理数据；

[0226] 2)保存模块，用于将待处理数据保存在数据传输通道在接收设备的通道缓存器中；

[0227] 3)拷贝模块，用于将通道缓存器中的当前待处理数据拷贝到与页面地址对应页面内存中。

[0228] 可选地，在本实施例中，在执行当前待处理数据的RDMA操作时，还可以不获取确切的页面地址，将当前待处理数据作为待传输数据通过数据传输通道直接发送给接收设备，并保存在接收设备的通道缓存器中，接收设备的应用会直接从上述通道缓存器中将当前待处理数据读取出来并以此拷贝到对应的页面内存中，进行RDMA操作。

[0229] 通过本申请提供的实施例，在数据量较小时，利用数据传输通道的通道缓存，实现直接在接收设备中完成对当前待处理数据的RDMA操作，从而达到提高数据处理效率的效果。

[0230] 作为一种可选的方案，还包括：

[0231] 1)建立单元，用于预先在发送设备与接收设备之间建立数据传输通道，其中，数据传输通道连接发送设备中的控制队列与接收设备中的环形队列，其中，控制队列包括用于控制环形队列中数据状态变化的控制信息。

[0232] 可选地，在本实施例中，建立单元包括：

[0233] 1)设置模块，用于在控制队列及环形队列中分别设置用于指示队列中数据状态变化的指针，其中，控制队列与环形队列中的指针所指示的位置同步变化；

[0234] 其中，指针包括：队列头指针、队列尾指针及队列接收指针，其中，队列尾指针与队列头指针之间的数据用于表示接收设备尚未确认接收的数据；队列接收指针与队列尾指针之间的数据用于表示接收设备已确认接收，且尚未处理的数据。

[0235] 可选地，在本实施例中，在控制队列及环形队列中分别设置队列头指针HEAD、队列尾指针TAIL及队列接收指针RECV_TAIL，通过上述指针所指示的位置的变化，实现对待传输

数据的传输控制。

[0236] 需要说明的是，发送设备的新消息采用加入到头部的方式进行，接收设备从尾部开始接收。发送设备判定当HEAD等于TAIL时，则队列为空，当HEAD+1等于RECV_TAIL时，则队列为满。

[0237] 通过本申请提供的实施例，通过在发送和接收两侧分别建立对应的数列，实现基于数列建立数据传输通道，以实现对待处理数据的灵活传输控制。

[0238] 作为一种可选的方案，设置模块通过以下方式实现控制队列与环形队列中的指针所指示的位置同步变化包括：

[0239] 1)在发送设备执行发送操作时，控制队列的队列头指针将向前移动N个数据字节，并将队列头指针所指示的位置将同步到接收设备的环形队列中，其中，N为大于等于1的自然数；

[0240] 2)在接收设备执行接收操作时，环形队列的队列尾指针将向前移动M个数据字节，并将队列尾指针所指示的位置同步到发送设备的控制队列中，其中，M≤N，M为大于等于1的自然数；

[0241] 其中，在发送设备的控制队列中队列尾指针移动到队列头指针时，更新队列接收指针的位置。

[0242] 具体结合以下示例进行说明，如图3所示，图3所示的节点1和节点2之间建立的跨节点数据传输通道是基于对数据队列的控制，具体的交互流程如下：

[0243] S1，节点1向节点2发送消息时，将消息数据写入控制队列后，将控制队列的指针HEAD增加，例如，指针HEAD向前移动5个字节，并将指针HEAD所指示的位置同步给节点2；

[0244] S2，节点2收到节点1的中断信号或者轮询到指针HEAD变化时，进行消息处理，将消息交给节点2中的应用模块处理。消息处理后，节点2将环形队列的指针TAIL增加，例如，指针TAIL向前移动2个数据字节，则将指针TAIL所指示的位置同步给节点1；

[0245] S3，节点1收到节点2的中断信号或者轮询到指针TAIL变化时，说明已发送待确认的队列里有数据，节点1处理已发送待确认队列里的消息，通知节点1中的应用模块消息处理完成，则节点1将更新指针RECV_TAIL的位置，即释放指针RECV_TAIL的当前位置。

[0246] 需要说明的是，图3所示的实线表示在本侧执行处理后得到的位置，虚线表示在对侧执行处理后同步得到的位置。

[0247] 通过本申请提供的实施例，同步两侧队列中指针所指示的位置，从而实现准确控制所传输的数据的传输状态，达到作为数据传输通道准确完成数据传输的目的。

[0248] 实施例5

[0249] 本发明的实施例还提供了一种存储介质。可选地，在本实施例中，上述存储介质可以被设置为存储用于执行以下步骤的程序代码：

[0250] S1，获取当前待处理数据的数据量，其中，当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据；

[0251] S2，若当前待处理数据的数据量大于第一预定阈值，则直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作；

[0252] S3，若当前待处理数据的数据量小于等于第一预定阈值，则通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作。

[0253] 可选地，在本实施例中，上述存储介质可以包括但不限于：U盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0254] 可选地，本实施例中的具体示例可以参考上述实施例及可选实施方式中所描述的示例，本实施例在此不再赘述。

[0255] 显然，本领域的技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件结合。

[0256] 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。

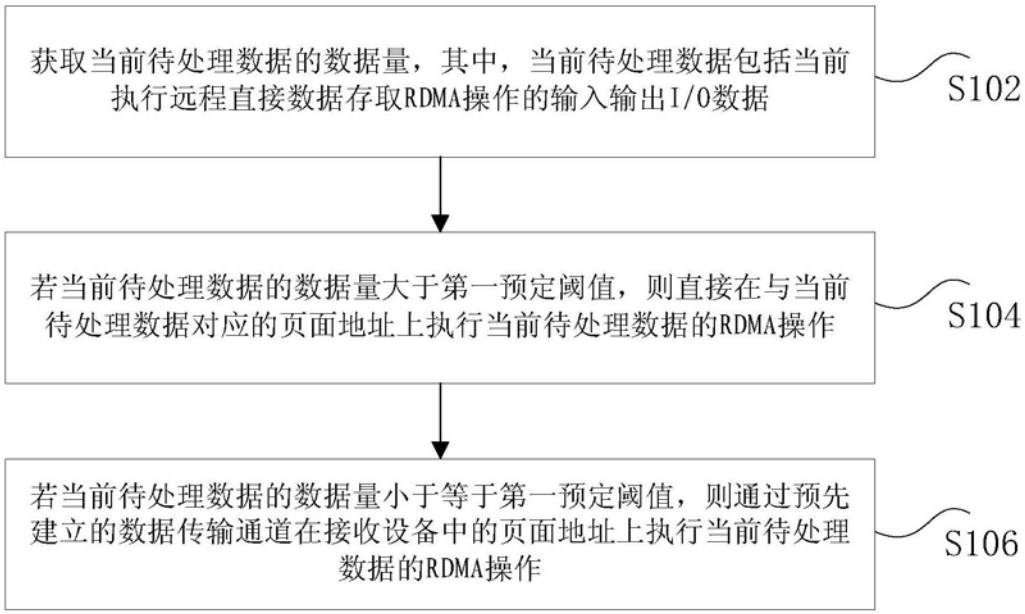

获取当前待处理数据的数据量，其中，当前待处理数据包括当前执行远程直接数据存取RDMA操作的输入输出I/O数据
S102
若当前待处理数据的数据量大于第一预定阈值，则直接在与当前待处理数据对应的页面地址上执行当前待处理数据的RDMA操作
S104
若当前待处理数据的数据量小于等于第一预定阈值，则通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作
S106

图1

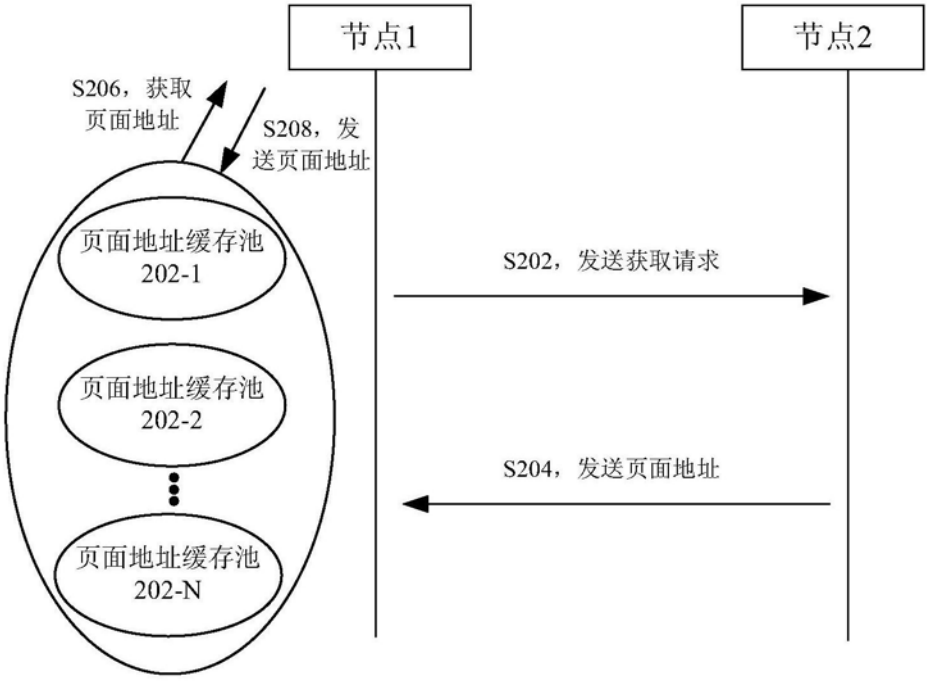

节点1
节点2
S206，获取页面地址
S208，发送页面地址
页面地址缓存池202-1
页面地址缓存池202-2
页面地址缓存池202-N
S202，发送获取请求
S204，发送页面地址

图2

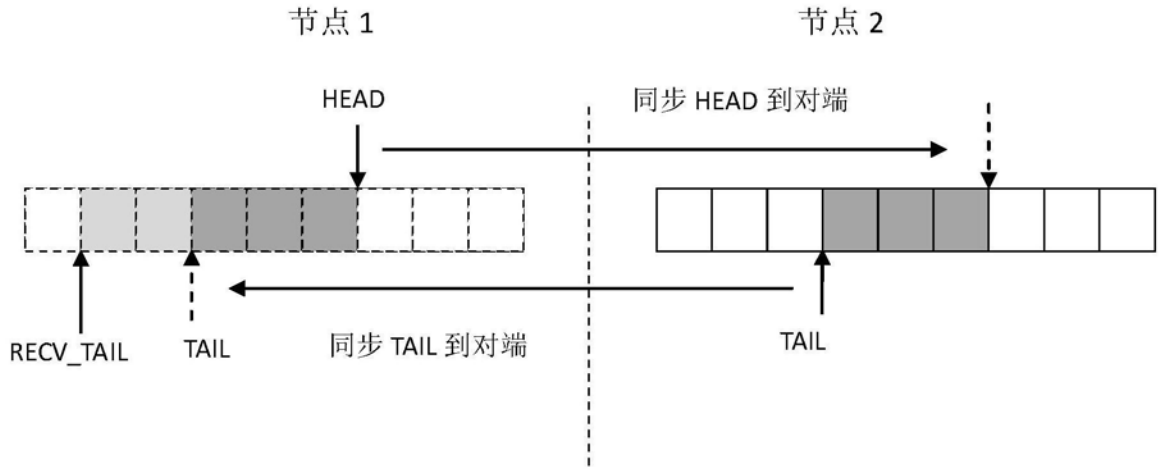

节点 1
节点 2
HEAD
同步 HEAD 到对端
RECV_TAIL
TAIL
同步 TAIL 到对端
TAIL

图3

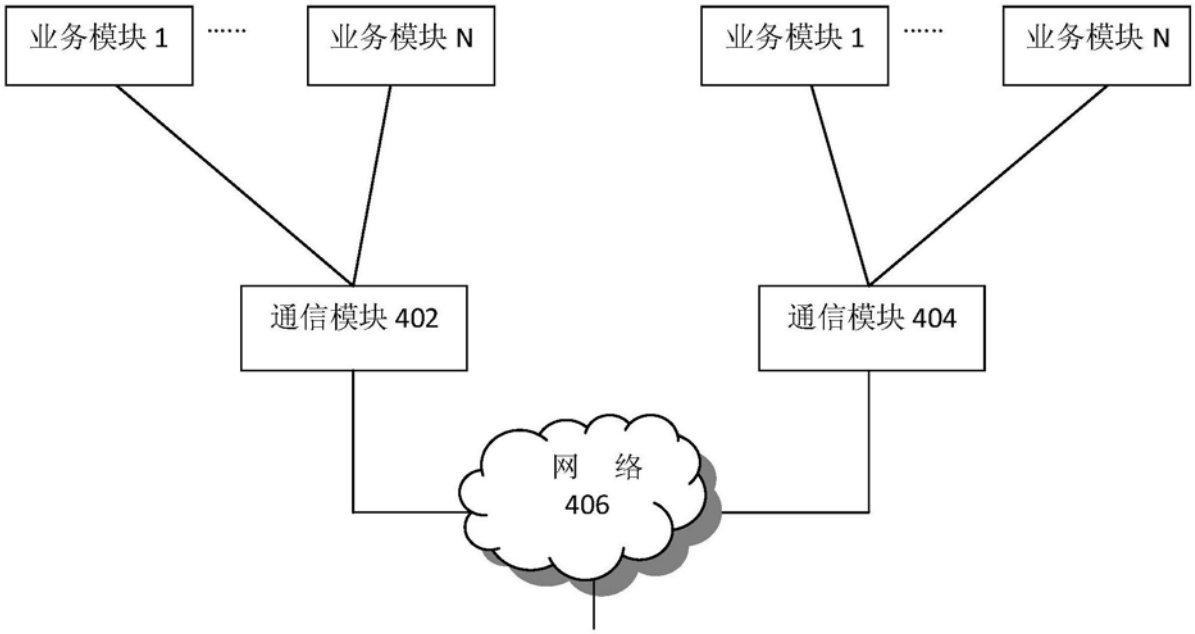

业务模块 1
……
业务模块 N
业务模块 1
……
业务模块 N
通信模块 402
通信模块 404
网 络
406

图4

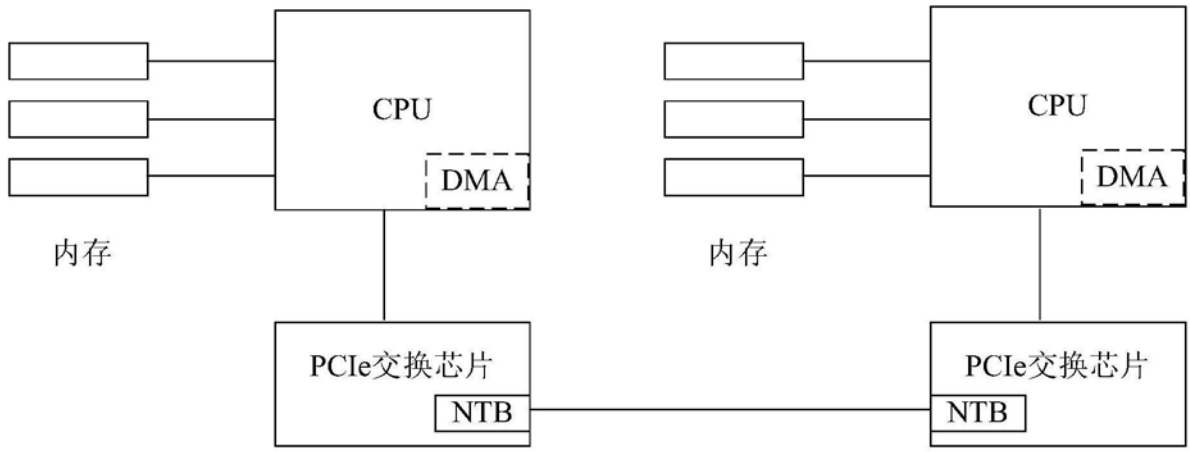

CPU
DMA
内存
CPU
DMA
内存
PCIe交换芯片
NTB
PCIe交换芯片
NTB

图5

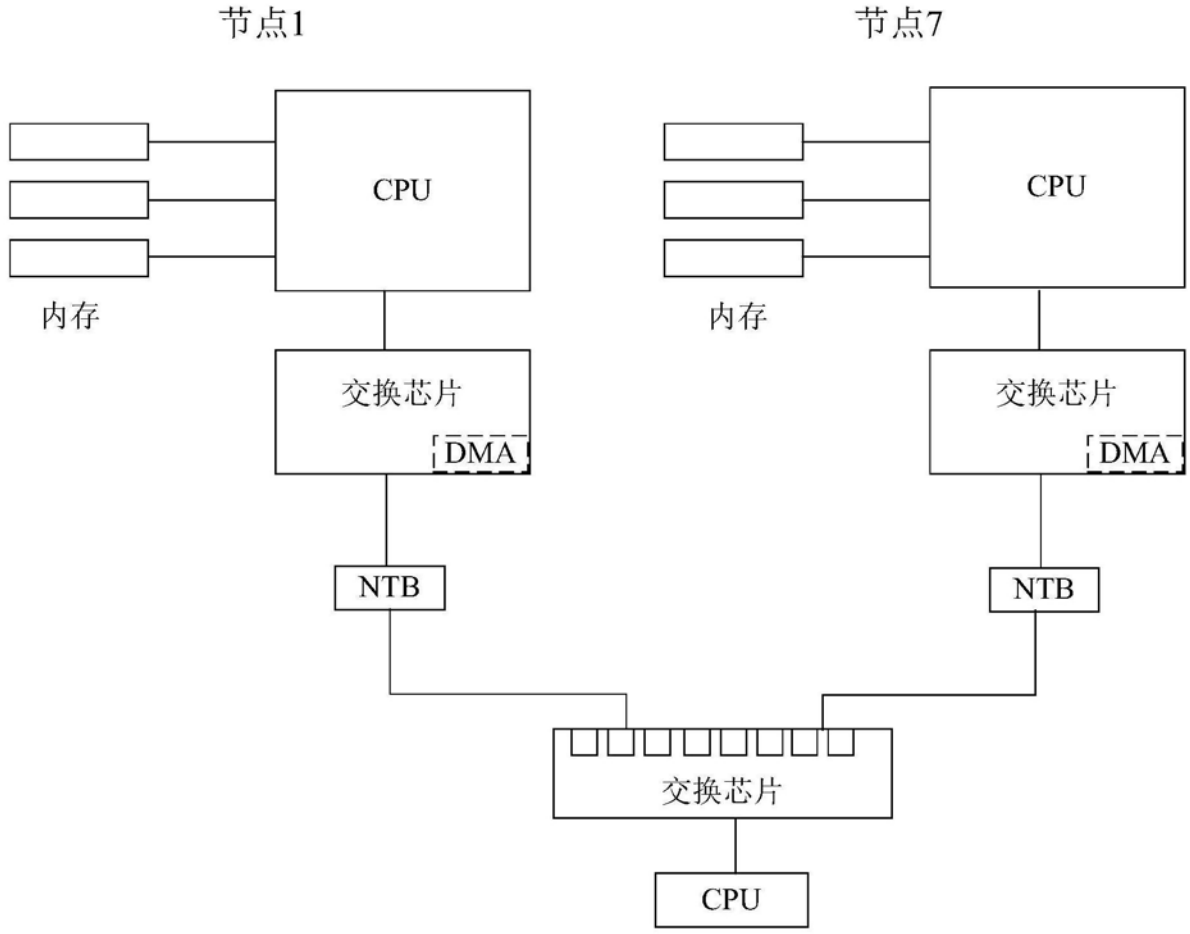

节点1
节点7
CPU
内存
交换芯片
DMA
CPU
内存
交换芯片
DMA
NTB
NTB
交换芯片
CPU

图6

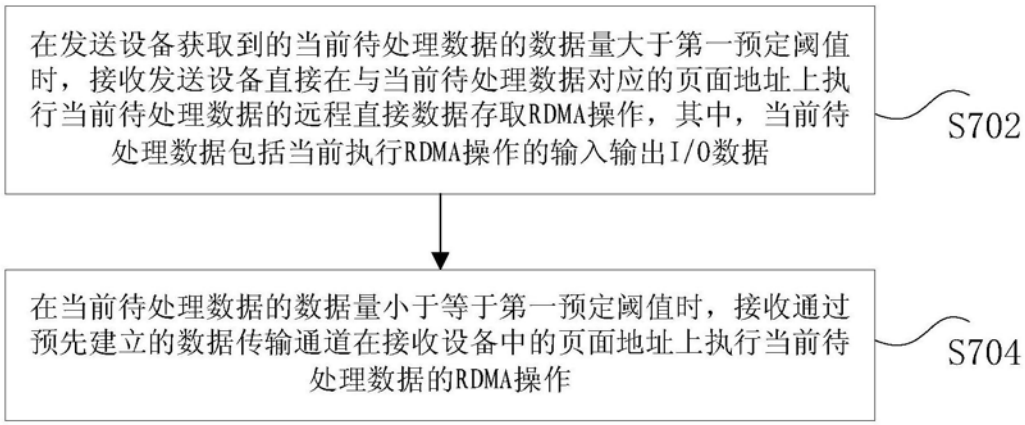

在发送设备获取到的当前待处理数据的数据量大于第一预定阈值时，接收发送设备直接在与当前待处理数据对应的页面地址上执行当前待处理数据的远程直接数据存取RDMA操作，其中，当前待处理数据包括当前执行RDMA操作的输入输出I/O数据
S702
在当前待处理数据的数据量小于等于第一预定阈值时，接收通过预先建立的数据传输通道在接收设备中的页面地址上执行当前待处理数据的RDMA操作
S704

图7

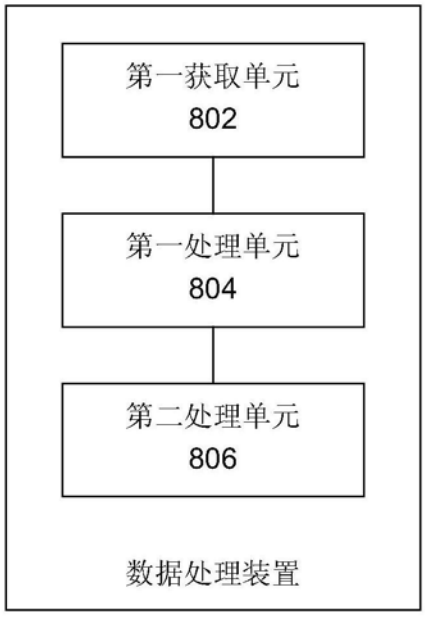

第一获取单元
802
第一处理单元
804
第二处理单元
806
数据处理装置

图8

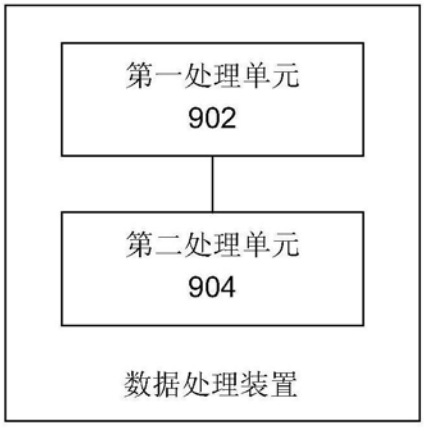


图9