

## 说明书

## 通信方法和通信装置

本申请要求于 2023 年 11 月 27 日提交中国专利局、申请号为 202311603388.X、申请名称为“通信方法和通信装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

## 技术领域

本申请涉及计算机技术领域，特别涉及一种通信方法和通信装置。

## 背景技术

外围器件互联总线(peripheral component interconnect express, PCIe)是一种高速串行计算机扩展总线标准，用于实现处理器和外部设备的连接，这些外部设备被称为 PCIe 设备或者简称为设备(Device)。

在当前 PCIe 体系中，任何 2 个设备之间通信，只能是 2 个接口之间的单路径通信，若通信的两个设备之间需要实现报文保序传输，可以通过点对点的保序(order)机制，来保证传输路径上的报文顺序，进而达到端到端的保序。

但是目前的 PCIe 体系中报文保序方式仅支持单路径。因此，如何设计一种可以在多路径传输场景下实现报文保序的方案成为亟待解决的问题。

## 发明内容

本申请提供一种通信方法，以期在多路径传输的场景下，实现报文保序。

第一方面，提供了一种通信方法，应用于第一主机。该方法可以由第一主机执行，或者，也可以由配置于第一主机中电路或芯片执行，本申请对此不作限定。为了便于描述，下文中以第一主机执行为例进行说明。

该通信方法，包括：第一主机获取第二主机中第一资源空间的信息，所述第一资源空间的信息包括所述第一资源空间的起始地址、所述第一资源空间的长度信息和所述第一资源空间对应的保序属性信息，所述第一资源空间对应的保序属性信息用于指示对所述第一资源空间进行访问的第一访问报文对应的保序要求；所述第一主机保存所述第一资源空间对应的保序属性信息和第一路径信息，所述第一路径信息用于指示访问所述第一资源空间的第一访问报文对应的传输路径，其中，所述第一路径信息所指示的传输路径为所述第一主机和所述第二主机之间包括多条传输路径中的一条。

基于上述技术方案，第一主机作为源设备，第二主机作为目标设备。在通用总线系统中，源设备和目标设备之间进行报文传输时，报文可以通过多条传输路径进行传输。具体地，该技术方案中第二主机上的第一资源空间(如，内存片段(memory segment)和/或功能实体(function entity))可以通过注册的方式，提供给第一主机使用。第一主机获取第二主机的第一资源空间相关信息时，除了获取第一资源空间的大小和地址(如，第一资源空间的起始地址、第一资源空间的长度信息)信息之外，还获取第一资源空间对应的保序属性信息，并且保存第一资源空间对应的信息(如，保存第一资源空间对应的保序属性信息和第一路径信息)，以使得第一主机在有向第二主机发送访问报文访问第一资源空间需求的情况下，可以基于已保存的信息获知访问第一资源空间的访问报文的保序要求，以及访问报文的传输路径，即具有相同保序要求的访问报文可以通过某条传输路径传输，而不同保序要求的访问报文通过不同的路径传输，实现在多路径传输的场景下，报文的保序传输。

结合第一方面，在第一方面的某些实现方式中，所述第一主机保存所述第一资源空间对应的保序属性信息和第一路径信息，包括：所述第一主机将所述第一资源空间的信息配置到译码器表项中，所述译码器表项中包括第一条目，所述第一条目包括所述第一路径信息和所述第一资源空间对应的保序属性信息。

基于上述技术方案，第一主机可以通过将第一资源空间的信息配置到译码器表项，基于译码器表项中的第一条目记录该第一资源空间的信息，简化了第一主机保存第一资源空间的信息的方式。

结合第一方面，在第一方面的某些实现方式中，在所述第一主机确定向所述第二主机发送第一访问

报文访问所述第一资源空间时，所述方法包括：所述第一主机查询所述译码器表项确定所述第一路径信息和所述第一访问报文对应的保序要求；所述第一主机根据所述第一路径信息确定所述第一访问报文对应的输出端口和第一传输路径；其中，所述第一传输路径为所述多条传输路径中用于传输所述第一访问报文的传输路径，所述第一传输路径用于传输相同路径信息对应的访问报文。

基于上述技术方案，第一主机确定向第二主机发送第一访问报文访问第一资源空间的情况下，第一主机可以通过查译码器表项，获取到第一访问报文对应的序属性信息，以保证在满足第一访问报文的保序要求的前提下，进行第一访问报文的传输。

结合第一方面，在第一方面的某些实现方式中，所述方法还包括：所述第一主机根据所述第一传输路径和所述第一访问报文对应的保序要求生成所述第一访问报文，其中，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述第一访问报文的负载分担因子字段承载指示所述第一传输路径的信息。

基于上述技术方案，第一主机生成第一访问报文时，根据查询译码器表项得到的第一访问报文对应的序属性信息和第一路径信息，填充报文的相关域端，从而生成的第一访问报文中携带有指示保序属性的信息以及传输路径的信息。

结合第一方面，在第一方面的某些实现方式中，所述第一访问报文的路由指示字段设置为 0，其中，所述路由指示字段设置为 0 用于指示采用哈希路由机制确定所述第一访问报文的传输路径。

结合第一方面，在第一方面的某些实现方式中，所述第一主机获取所述第一资源空间对应的保序属性信息，包括：所述第一主机接收来自所述第二主机的所述第一资源空间对应的保序属性信息；或者，所述第一主机根据所述第一资源空间的用途确定所述资源空间对应的保序属性信息。

基于上述技术方案，第一主机获取第一资源空间对应的保序属性信息的方式可以是：从第二主机接收到，例如，第二主机将第一资源空间通过注册的方式，提供给第一主机使用的过程中可以根据不同资源空间的功能设置不同的保序需求，并将相应的保序属性信息提供给第一主机。或者，第一主机可以自行根据不同资源空间的用途来决定确定不同的保序需求，提高方案的灵活性。

结合第一方面，在第一方面的某些实现方式中，所述方法还包括：所述第一主机获取所述第二主机中第二资源空间的信息，所述第二资源空间的信息包括所述第二资源空间的起始地址、所述第二资源空间的长度信息和所述第二资源空间对应的保序属性信息，所述第二资源空间对应的保序属性信息用于指示对所述第二资源空间进行访问的第二访问报文对应的保序要求；所述第一主机保存所述第二资源空间对应的保序属性信息和第二路径信息，所述第二路径信息用于指示访问所述第二资源空间的第二访问报文对应的传输路径，其中，所述第二路径信息和所述第一路径信息不同，所述第二资源空间对应的保序属性信息和所述第一资源空间对应的保序属性信息不同。

基于上述技术方案，第一主机可以获取第二主机上的不同资源空间的信息，另外，不同保序需求的资源空间对应的访问报文可以通过不同路径进行传输，从而在保证多路径传输的前提下，实现报文保序传输。

结合第一方面，在第一方面的某些实现方式中，所述第一主机保存所述第二资源空间对应的保序属性信息和第二路径信息，包括：所述第一主机将所述第二资源空间的信息配置到所述译码器表项中，所述译码器表项中包括第二条目，所述第二条目包括所述第二路径信息和所述第二资源空间对应的保序属性信息。

结合第一方面，在第一方面的某些实现方式中，所述第一访问报文对应的保序要求包括以下任意一种：强保序（strongly ordered, SO）、无需保序（non-ordered, NO）、或灵活保序（relaxed ordering, RO）。其中，强保序表征第一访问报文为保序报文，需要遵循保序方式传输，依赖其他报文的传输结果进行传输，如，第一访问报文与之前的标识为 RO 的报文有保序要求。无需保序表征第一访问报文没有保序需求，不需要依赖其他报文的传输结果进行传输；灵活保序表征第一访问报文后续的标识为 SO 的报文之间有保序要求，如，标识为 SO 的报文不能早于标识为 RO 的报文传输。

第二方面，提供了一种通信方法，应用于第二主机。该方法可以由第二主机执行，或者，也可以由配置于第二主机中电路或芯片执行，本申请对此不作限定。为了便于描述，下文中以第二主机执行为例进行说明。

该通信方法，包括：第二主机接收来自第一主机的第一访问报文，所述第一访问报文用于访问所述第二主机中的第一资源空间，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求

的信息，所述访问报文的负载分担因子字段承载指示第一传输路径的信息；所述第二主机根据所述第一访问报文的保序域确定所述第一访问报文的保序要求；其中，所述第二主机接收所述第一传输路径上传输的至少一个访问报文的物理端口相同，所述第一访问报文为所述至少一个访问报文中的一个。

以上第二方面及其可能的设计所示方法的技术效果可参照第一方面及其可能的设计中的技术效果。

第三方面，提供一种通信装置，该装置包括：存储模块，用于存储程序；处理模块，用于执行存储模块存储的程序，当存储模块存储的程序被执行时，处理模块用于执行上述各方面提供的方法。

第四方面，提供一种计算机可读存储介质，该计算机可读介质存储用于设备执行的程序代码，该程序代码包括用于执行上述各方面提供的方法。

第五方面，提供一种包含指令的计算机程序产品，当该计算机程序产品在计算机上运行时，使得计算机用于执行上述各方面提供的方法。

第六方面，提供一种芯片，该芯片包括处理模块与通信接口，该处理模块通过该通信接口读取存储器上存储的指令，用于执行上述各方面提供的方法。

可选地，作为一种实现方式，该芯片还可以包括存储模块，该存储模块中存储有指令，该处理模块用于执行该存储模块上存储的指令，当该指令被执行时，该处理模块用于执行上述各方面提供的方法。

第七方面，提供一种芯片，该芯片包括用于执行第一方面提供的方法的第一主机和用于执行第二方面提供的方法的第二主机。

第八方面，提供一种计算机系统，该计算机系统包括第七方面所示的芯片。

第九方面，提供一种终端设备，该终端设备包括第七方面所示的芯片。例如，终端设备包括但不限于手机、车辆等终端。

第十方面，提供一种通信系统，该系统包括用于执行第一方面提供的方法的第一主机和用于执行第二方面提供的方法的第二主机。

#### 附图说明

图1为本申请实施例提供的计算机设备的结构示意图。

图2为本申请实施例提供的数据中心示意图。

图3是本申请实施例提供的一种通用总线协议报文格式示意图。

图4是本申请实施例提供的一种通用总线协议系统中支持多路径传输的示意图。

图5是本申请实施例提供的PCIe体系示意图。

图6是本申请提供的一种通信方法的示意性流程图。

图7是本申请提供的一种资源注册的示意图。

图8是本申请提供的一种确定资源空间保序属性的示意图。

图9是本申请提供的一种条目的数据结构示意图。

图10是本申请提供的一种报文格式示意图。

图11示出了本申请实施例提供的一种通信装置1100的结构示意图。

图12示出了本申请实施例提供的一种芯片系统1200的结构示意图。

图13示意性地示出本申请实施例提供的计算机程序产品的概念性局部视图。

#### 具体实施方式

为了便于理解本申请实施例，做出以下几点说明。

第一，在本申请中示出的“至少一个”是指一个或者多个，“多个”是指两个或两个以上。另外，在本申请的实施例中，“第一”、“第二”以及各种数字编号（例如，“#1”、“#2”等）只是为了描述方便进行的区分，并不用来限制本申请实施例的范围。下文各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定，应该理解这样描述的对象在适当情况下可以互换，以便能够描述本申请的实施例以外的方案。此外，在本申请实施例中，“S610”等字样仅为了描述方便作的标识，并不是对执行步骤的次序进行限定。

第二，本申请实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其他实施例或设计方案更优选或更具优势。确切而言，使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

第三, 本申请实施例中涉及的“保存”, 可以是指的保存在一个或者多个存储器中。该一个或者多个存储器, 可以是单独的设置, 也可以是集成在编码器或者译码器, 处理器、或通信装置中。该一个或者多个存储器, 也可以是一部分单独设置, 一部分集成在译码器、处理器、或通信装置中。存储器的类型可以是任意形式的存储介质, 本申请并不对此限定。

第四, 本申请实施例中涉及的“包括”(也称“includes”、“including”、“comprises”和/或“comprising”)当在本说明书中使用指定存在所陈述的特征、整数、步骤、操作、元素、和/或部件, 但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元素、部件、和/或其分组。

第五, 本申请实施例中涉及的“如果”可被解释为意指“当...时”(“when”或“upon”)或“响应于确定”或“响应于检测到”。类似地, 根据上下文, 短语“如果确定...”或“如果检测到[所陈述的条件或事件]”可被解释为意指“在确定...时”或“响应于确定...”或“在检测到[所陈述的条件或事件]时”或“响应于检测到[所陈述的条件或事件]”。

第六, 本申请实施例中对各种所述示例的描述中所使用的术语只是为了描述特定示例, 而并非旨在进行限制。如在对各种所述示例的描述和所附权利要求书中所使用的那样, 数形式“一个(“a”, “an”)”和“该”旨在也包括复数形式, 除非上下文另外明确地指示。

第七, 本文中术语“和/或”, 仅仅是一种描述关联对象的关联关系, 表示可以存在三种关系, 例如, A 和/或 B, 可以表示: 单独存在 A, 同时存在 A 和 B, 单独存在 B 这三种情况。另外, 本文中字符“/”, 一般表示前后关联对象是一种“或”的关系。

下面将结合附图, 对本申请实施例中的技术方案进行描述。

本申请中涉及计算机系统处理器访问外部设备的配置空间。其中, 计算机系统可以是服务器, 或者是终端。终端包括但不限于台式机、笔记本电脑、智能手机等用户系统。为了便于理解, 下面对计算机系统的结构进行介绍。

参见图 1 所示的计算机系统的结构示意图, 计算机系统包括处理器 101、输入输出设备(input output device, I/O device)102、内存 103、缓存 104、内存管理单元(memory management unit, MMU)105、输入输出内存管理单元(input output management unit, IOMMU)106、外存 107 和总线 108。

处理器 101 包括至少一个内核(core)。该内核也称作计算引擎。其中, 每个内核可以独立地执行任务。当处理器 101 包括多个内核时, 可以对来自应用的任务进行划分, 使得应用能够充分利用多个内核, 在特定的时间内执行更多任务。在本实施例中, 处理器 101 可以是主处理器, 例如为中央处理机(Central Processing Unit, CPU)。

输入输出设备 102 是指具有输入数据和/或输出数据能力的硬件设备。输入输出设备 102 可以分为输入设备和输出设备。其中, 输入设备可以包括鼠标、键盘、操作杆、触控笔、麦克风等设备, 输出设备可以包括显示器、扬声器等设备。

内存 103 也称作内存存储器或主存储器, 用于暂时存放处理器 101 中的运算数据。进一步地, 内存 103 还用于暂时存放与外存 107 交换的数据。内存 103 通常可以采用动态随机存取存储器 DRAM 或者静态随机存取存储器(static random access memory, SRAM)等存储介质实现。

缓存 104(本实施例中是指处理器缓存, 如 CPU 缓存)是用于减少处理器 101 访问内存 103 所需平均时间的部件。参见图 1, 在金字塔式存储体系中, 缓存 104 位于自顶向下的第二层, 仅次于处理器 101 的寄存器(图 1 中未示出), 高于内存 103(内存 103 位于自顶向下的第三层)。通常情况下, 缓存 104 的容量远小于内存 103, 但访问速度可以接近处理器 101 的频率。

内存管理单元 105 是一种用于处理数据访问请求的计算机硬件。内存管理单元 105 具体用于对数据访问请求中的虚拟地址(virtual address, VA)进行映射。其中, 内存管理单元 105 可以截获处理器 101 的内核发出的数据访问请求, 将数据访问请求中的虚拟地址映射(或者翻译)为物理地址(physical address, PA), 以便于根据该物理地址访问内存 103。

输入输出内存管理单元 106 实质是一种内存管理单元。类似于内存管理单元 105 将处理器 101 可见的虚拟地址映射为物理地址, 输入输出内存管理单元 106 用于将输入输出设备可见 102 的虚拟地址(也可以称作设备地址或 I/O 地址)映射为物理地址。

外存 107 也称作外部存储器、辅存, 通常用于持久化保存数据。例如, 外存 107 可以持久化存储处理器 101 中的运算数据。即使供电异常, 已经写入该外存 107 的数据仍然能够保存, 避免了数据丢失。具体实现时, 外存 107 包括至少一个非易失性存储器 1071, 当外存包括多个非易失性存储器时, 这多个

非易失性存储器可以是相同类型，也可以是不同类型。例如，在图 1 的示例中，外存 107 可以包括两种类型的非易失性存储器，例如为存储级存储器(storage class memory, SCM)和固态硬盘(solid state drive, SSD)。

总线 108 用于将计算机系统的各个功能部件连接。总线 108 是计算机系统各种功能部件之间传送信息的公共通信干线。总线 108 可以由导线形成的传输线束。根据连接对象不同，总线 108 还可以分为内部总线和外部总线。

其中，内部总线采用内部总线协议传送信息。内部总线协议包括用于访问所述计算机系统的内存空间的总线协议。外部总线采用外部总线协议传送信息。外部总线协议包括用于访问所述计算机系统的外存空间的总线协议。其中，内存空间是指内存的地址空间，外存空间是指外存的地址空间。

在一些实施例中，内部总线协议包括但不限于外设部件互连标准(peripheral component interconnect, PCI)总线、外设部件互连标准高速(PCI Express, PCIe)协议、快速通道互联(Intel™ Quick Path Interconnect, QPI)协议、通用总线(unified bus)协议。外部总线协议包括但不限于小型计算机系统专用接口(small computer system interface, SCSI)协议或者串行连接小型计算机系统专用接口(Serial Attached SCSI, SAS)协议。

需要说明的是，图 1 所示的计算机系统是以外存 107 为远端外存进行示例说明。如图 1 所示，外存 107 包括网卡 1072。该网卡 1072 例如可以是 smart NIC 网络接口卡(也即网络适配卡)。外存 107 通过该网卡 1072 接入网络，进而通过网络与计算机系统 101 的其他部件连接。网络可以是有线通信网络，如光纤通信网络，也可以是无线通信网络，例如是无线局域网(wireless local area network, WLAN)或者是第五代(the fifth generation, 5G)移动通信网络。

在一些可能的实现方式中，计算机系统的外存 107 也可以是本地外存，计算机系统的其他部件如处理器 101 可以通过总线 108 连接上述本地外存。在另一些可能的实现方式中，计算机系统可以既包括远端外存，又包括本地外存。此外，本申请实施例可以适用于集中式存储，或者分布式存储场景，本实施例对此不作限定。

示例性地，本申请实施例中提供的访问总线设备的配置空间寄存器的方式还可以应用于图 2 所示的跨网络通信的服务器集群，如图 2 所示的数据中心。其中，图 2 中所示的交换机或服务器内部结构如上文图 1 中所示。

另外，本申请中涉及的计算机系统所支持的内部总线协议包括通用总线协议，计算机系统之间可以建立传输层的连接。其中，通用总线协议还可以称为灵衢总线协议或统一总线协议，一种总线协议标准，本申请对于该通用总线协议的名称不做限定。

通用总线协议打破现有各种协议壁垒，去除中间不必要的转换开销，从而实现极致低时延。通用总线协议定义了独立的事务层(transaction, TA)和传输层(transport, TP)。传输层之间有连接，事务层之间无连接。针对主机(host)来说，两个主机之间的事务层之间无连接，传输层之间有连接，那么两个主机中的任一主机的所有事务都承载在传输层上传输。通用总线协议包含传输层和事务层，传输层负责网络丢包重传，保证可靠传输，事务层处理各自不同的事务。传输层从网络收到包，剥离传输层包头，转发给事务层。

通用总线协议报文格式如图 3 所示。具体地，通用总线协议报文格式中的字段定义如下表 1 所示：

表 1

| 名称(name)                             | 描述(description)                                                      |
|--------------------------------------|----------------------------------------------------------------------|
| 通用总线链接层(unified bus LINK)            | 通用总线协议定义的链接(LINK)层，通用总线链接层也可替换为以太 MAC。                               |
| 网络分区 ID(network partition ID, NPI)   | 用于物理网络隔离。                                                            |
| IP                                   | IP 协议头                                                               |
| 用户数据报协议(user datagram protocol, UDP) | UDP 协议头，UDP 目的端口等于 4792 表示通用总线报文，后面是传输层头(transport header, TPH)。     |
| TPH                                  | 传输层头，包含 TP 操作码(opcode)、源 TPN、目的 TPN、包序列号(packet sequence number)等内容。 |
| 通用总线分区 ID(unified bus                | 用于租户隔离。                                                              |

|                                                       |                                                                                                                                 |
|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| partition ID , UPI)                                   |                                                                                                                                 |
| 通用总线实体 ID (unified bus entity ID , UEID) 或者简称为 EID    | 包括源实体标识 (source entity ID) 和目的实体标识 (destination entity ID)。EID 位宽 128bit, 全网唯一。一个 EID 可以表示一个虚拟机, 也可以表示一个 SSD 控制器。               |
| 事务层头 (transaction header, TAH)                        | 事务层头。包含 TA 操作码 (opcode) (表示事务操作类型, 如 send、read、write、atomic 或事务层应答)、目的 JFR 号、事务层顺序段号 (segment sequence number, TASSN)、读写地址和长度等。 |
| 载荷 (payload)                                          | 通用总线协议报文的载荷部分                                                                                                                   |
| 不变的循环冗余校验码 (invariable Cyclic Redundancy Check, ICRC) | 用于对通用总线协议报文进行校验                                                                                                                 |

具体地, 通用总线协议的事务层和应用的交互接口称为 Jetty, 应用的消息可以通过一个 Jetty 发到任何目的地, 也可以通过一个 Jetty 接收来自任何源的消息。只能发送的 Jetty, 定义为 (Jetty for send, JFS); 只能接收的 Jetty, 定义为 (Jetty for receive, JFR)。

图 4 是本申请实施例提供的一种通用总线协议系统中, 支持多路径传输的示意图。

如图 4 所示, 本申请涉及的通用总线协议系统中, 支持多路径, 包括源多端口 (如图 4 中所示的主机 A 中的端口#0 (port#0)、端口#1、端口#2 和端口#3), 目的多端口 (如图 4 中所示的主机 B 中的端口#0、端口#1、端口#2 和端口#3) 以及网络多路径 (如图 4 中所示的路径#1、路径#2、...、路径#n)。其中, 主机 A 可以理解为源设备, 主机 B 可以理解为目的设备, 主机 A 可以通过多路径向主机 B 发送报文。

可选地, 主机 A 和/或主机 B 中还可以包括处理器 (processing unit, PU) (如, 中央处理器 (central processing unit, CPU))、控制器 (controller) 和片上系统 (system on chip, NOC) 等, 本申请中对于主机中包括的硬件或软件系统不做任何的限定, 可以参考目前相关技术中关于主机的介绍, 本申请中对此不进行赘述。

示例性地, 在图 4 所示的通用总线协议系统中, 两个设备之间可以通过源多端口或目的多端口, 以及网络多路径传输报文, 以期提升通信带宽。

上文中结合图 1 至图 4 简单介绍本申请能够应用的场景以及涉及的主机内部逻辑单元, 为了便于理解本申请实施例, 对本申请涉及的一些基本概念做简要说明。

1、报文保序: 按照一定的顺序发送和/或接收报文, 实现报文保序。例如, 在当前 PCIe 体系中, CPU 通过用于访问的报文 (如, 存储 (store) 指令、载入 (load) 指令等) 访问外设的不同属性的内存时, 可以采用的不同的保序方式实现报文保序。

示例性地, 内存属性分为两类:

设备 (device) 型内存: 包括设备寄存器的输入输出 (input output, IO) 内存空间。

普通 (normal) 型内存: 包括静态随机存取存储器 (static random access memory, SRAM) 或动态随机存取存储器 (dynamic random access memory, DRAM) 等内存空间。

当访问设备 (device) 型内存空间时, 通常按照端点保序 (endpoint order) 方式实现报文保序; 当访问普通 (normal) 型内存时, 通常按照请求保序 (request order) 方式实现报文保序。其中, 端点保序可以实现到同一端点的多个报文的顺序; 请求保序可以实现到同一地址的多个报文的顺序。

为便于理解, 结合图 5 简单说明在当前 PCIe 体系中, 2 个设备之间通信如何实现报文保序。

如图 5 所示, PCIe 组网为树形结构, 包括如图 5 中所示的 PCIe 根复合体 (root complex, RC)、交换 (switch) 设备和 PCIe 终端设备 (endpoint, EP)。其中, PCIe RC 负责管理 CPU 和其他设备之间的通信, switch 设备用于将一个 PCIe 接口拓展为多个 PCIe 接口, PCIe EP 为没有转发数据的功能的设备。

应理解, 图 5 仅为示例, 对本申请的保护范围不构成任何的限定, 只是举例说明 PCIe 组网中不同设备之间进行通信的场景下如何实现报文保序。例如, 在图 5 所示的场景下, CPU 下发了 3 个存储 (store) 请求, 且这三个存储请求之间没有保序关系, 但是 PCIe RC 接收到第一个存储请求之后, 会回复 Comp 响应, 表示当前的请求已经到达且需要保证命令的执行顺序和接收顺序是一致的, 因此需要将这三个报文标记为 SO, 即要求强序执行。

在当前 PCIe 体系中报文保序存在以下特点:

1) 仅支持单路径 (single path)，在当前 PCIe 体系保序 (ordering) 机制中，通过点对点的保序机制，以期保证整个报文传输路径上的序，进而达到端到端的保序；

2) 即使 PCIe EP 侧有多个 Bar 空间给 PCIe RC 使用时，输入输出内存映射 (memory map I/O, MMIO) 空间访问依旧是走同一个路径，遵守同一套保序机制；

3) 在当前 PCIe 系统设计中，逻辑实现都是按照加严的，xPU 发出的 load/store 指令在接入到 PCIe 链路时，默认 order 属性全部采用“SO”。

2、保序报文：由上述可知报文保序针对的是有保序需求的报文，本申请中有保序需求的报文称为保序报文，没有保序需求的报文称为非保序报文。示例性地，有保序需求的报文包括但不限于：依赖于其他报文的接收情况而发送的报文，例如，报文#1 需要在报文#2 被成功接收的情况下才可以发送，则报文#1 为保序报文；同理，没有保序需求的报文包括但不限于：不依赖于其他报文的接收情况而发送的报文。

3、传输链路组 (transport group, TPG)：源 (initiator) 设备和目标 (target) 设备之间可以在传输层建立至少一个 TPG，每个 TPG 包含多个传输链路 (transport port, TP)，源设备向目标设备发送报文时，可以在 TPG 内的多个 TP 间负载均衡，不同报文可以通过不同的网络路径传输。

4、资源注册：在通用总线系统中，目标设备上的内存片段 (memory segment) 和/或功能实体 (function entity, FE) 可以通过注册的方式，提供给源设备使用。其中，内存片段和/或功能实体是目标设备对自身资源划分的单元，表征具备一定隔离性的设备资源。例如，内存片段是一段连续的虚拟地址 (virtual address, VA) 空间，每个内存片段对应一段物理内存，目标设备创建内存片段并注册，源设备申请使用目标设备的内存片段。

示例性地，源设备获取到内存片段信息后，将内存片段的通用总线地址 (unified bus address, UBA) 映射到本地进程 VA 空间中，得到映射地址 (mapped VA, mVA)。

应理解，本申请中对于目标设备的资源注册流程不做限定，例如，源设备和目标设备之间可以通过带内交换机制，完成资源的申请注册，也可以通过带外交换的机制。

上文结合图 2 简单介绍了本申请提供的通信方法适用的场景，以及介绍了本申请涉及的基本概念。并在基本概念中介绍了当前 PCIe 系统设计中报文保序的方式，由前文所述，当前 PCIe 系统设计中报文保序仅支持单路径，不具备多路径能力，无法提供更大的交互带宽。也就是说该报文保序的方式不适用于图 4 所示的通用总线协议系统中，因为图 4 所示的通用总线协议系统支持多路径传输。

另外，由上文所示的 PCIe 系统设计中报文保序的特点可知，Load/store 指令接入网络固定按照“SO”进行保序保序，在接收侧执行效率低，以及对于不存在序关系的流，因为使用同一个物理链路，引入序关联，交互效率下降。

本申请提供一种通信方法，以期在多路径传输的场景下，实现报文保序。其中，多路径传输的场景包括但不限于图 4 所示的通用总线协议系统支持的多路径传输，或者还可以是其他支持的多路径传输系统（如，能够支持多路径传输的 PCIe 系统）。

应理解，本申请实施例提供的通信方法可以应用于计算机系统中，例如，图 2 所示的跨网络通信系统中。

还应理解，下文示出的实施例并未对本申请实施例提供的方法的执行主体的具体结构特别限定，只要能够通过运行记录有本申请实施例的提供的方法的代码的程序，以实现本申请实施例提供的方法即可。例如，本申请实施例提供的方法的执行主体可以是设备，或者，是设备中能够调用程序并执行程序的功能模块。

图 6 是本申请提供的一种通信方法的示意性流程图。应用于第一主机和第二主机之间进行通信的场景下，如图 2 所示的场景中。

在图 6 所示的实施例中，第一主机可以作为发送端（或者称为源设备 (initiator)），第二主机可以作为接收端（或者称为目标设备 (target)），对于接收端来说，接收端为本端设备，发送端为远端设备，对于发送端来说，发送端为本端设备，接收端为远端设备。

示例性地，上述的第一主机可以是物理机或者虚拟机，若第一主机为虚拟机，第一主机执行的步骤可以由虚拟机执行；同理，第二主机可以是物理机或者虚拟机，若第二主机为虚拟机，第二主机执行的步骤可以由虚拟机执行。

具体地，图 6 所示的方法包括以下步骤：

S610，第一主机获取第二主机中第一资源空间的信息。

该实施例中，第一主机为发送报文的发送端，如，第一主机为发送 load 和/或 store 指令的设备。第二主机为接收报文的接收端，如，第二主机为接收 load 和/或 store 指令的设备。

示例性地，该实施例中第一主机可以理解为使用资源空间的设备，第二主机可以为提供资源空间给第一主机使用的设备。例如，第一主机可以为通用总线系统中的用户主机（user host），第二主机可以为通用总线系统中的基设备（home device）；还例如，第一主机和第二主机可以为通用总线系统中的基设备，或者第一主机和第二主机可以为通用总线系统中的用户主机；又例如，第一主机和第二主机可以为通用总线系统中的其他设备，该实施例中对此不做限定。其中，第二主机可以将自身的资源提供给第一主机使用，例如第二主机将 FE 或内存片段级别的资源提供给第一主机使用，以使得第一主机能使用通用总线系统中的其他设备资源。

为了便于理解，下面结合图 7 详细介绍该实施例中第二主机如何将自身的资源提供给第一主机使用，以及第一主机如何使能使用第二主机的资源。

从图 7 中可以看出，通用总线系统中包括第一主机（即用户主机）、第二主机（即基设备）和实体管理（function management, FM）设备。其中，第一主机为使用第二主机的资源的设备，第二主机是为第一主机提供可用资源的设备，FM 设备为系统中实现管理功能的设备，如，FM 设备可以通过扫描获取系统中的至少一个基设备。

示例性地，FM 设备通过扫描发现第二主机之后，将第二主机纳入管理，并且 FM 设备可以将第二主机的资源（如，FE 资源或内存片段等）注册到第一主机中，第一主机可以创建驱动第二主机的驱动（如图 7 中所示的第一主机的设备驱动），以使得第一主机可以使用第二主机的资源。

应理解，上述图 7 所示的第二主机将自身的资源提供给第一主机使用的方式仅为示例，对本申请的保护范围不构成任何的限定，第二主机也可以通过其他方式将自身的资源注册到第一主机，以使得第一主机可以使用第二主机的资源，该实施例中对此不做限定。

示例性地，该实施例中第一主机可以根据 FM 设备指定的第二主机的通信地址和第二主机的描述信息，获取第二主机中资源空间的信息。

具体地，第一资源空间的信息包括第一资源空间的起始地址、第一资源空间的长度信息和第一资源空间对应的保序属性信息，第一资源空间对应的保序属性信息用于指示对第一资源空间进行访问的第一访问报文对应的保序要求。

应理解，第一主机可以获取第二主机中至少一个资源空间的信息，上述的第一资源空间为该至少一个资源空间中的任意一个，第一资源空间的信息包括该第一资源空间的起始地址和该第一资源空间的长度。第一资源空间相对应的保序要求由该第一资源空间对应的保序属性信息指示，该第一资源空间对应的保序要求可以理解为对该第一资源空间进行访问的访问报文的保序要求。例如，第一主机还可以获取第二主机中的第二资源空间的信息，第二资源空间的信息包括第二资源空间的起始地址、第二资源空间的长度信息和第二资源空间对应的保序属性信息，第二资源空间对应的保序属性信息用于指示对第二资源空间进行访问的第二访问报文对应的保序要求。

作为一种可能的实现方式，不同资源空间对应的保序属性信息由第二主机提供。例如，第一主机所获取的至少一个资源空间的信息中包括的资源空间对应的保序属性信息。

在该实现方式下，不同资源空间对应的保序属性信息由第二主机确定。例如，第二主机可以基于资源空间的功能确定该资源空间对应的保序属性信息，并将相应的保序属性信息在资源空间注册流程中提供给第一主机。

为了便于理解，结合图 8 说明第二主机如何确定不同资源空间的保序属性信息。

如图 8 所示，第二主机上具有多个 FE 资源供使用，一个或多个 FE 对应的资源空间在功能设计时确定了不同用途，如：某些 FE 对应的资源空间用作配置（configure）；某些 FE 对应的资源空间用作中断信息；某些 FE 对应的资源空间用作下发命令的命令队列（command queue）；某些 FE 对应的资源空间用作 command queue 的门铃（doorbell）；某些 FE 对应的资源空间用于数据存储等等。

不同的 FE 对应的资源空间可以有不同的 order 属性。例如，用作下发命令的命令队列、或用于数据存储的 FE 对应的资源空间可以是 RO/NO；还例如，用作 doorbell 的 FE 对应的资源空间可以是 SO。不同的 FE 对应的资源空间之间可以有或没有保序要求，例如，不同 FE 资源空间之间可以没有序要求。

作为另一种可能的实现方式，不同资源空间对应的保序属性信息由第一主机确定。

在该实现方式下，第二主机为第一主机提供不同资源空间供第一主机使用时，未提供不同资源空间

所对应的保存属性信息，第一主机可以根据不同资源空间的实际用途确定不同资源空间的保序属性。

例如，第二主机为第一主机提供了资源空间#1 和资源空间#2，第一主机决定资源空间#1 用于存储数据，资源空间#2 用于接收中断指令，则第一主机确定资源空间#1 对应的保存属性信息为指示不需要保存的信息，资源空间#2 对应的保存属性信息为指示需要强保存的信息。

进一步地，该实施例中第一主机获取第二主机中资源空间的信息，确定不同资源空间对应的保序属性信息之后，可以保存资源空间对应的保序属性信息以及资源空间对应的路径信息，例如，将不同资源空间的信息配置到第一主机的通用总线译码器（UB decoder）表项中，该表项用于确定报文的路由路径；还例如，可以将保存资源空间对应的保序属性信息以及资源空间对应的路径信息保存在存储空间中，或者通过其他方式记录，该实施例中对此不做任何限定。

为了便于理解，该实施例中以将保存资源空间对应的保序属性信息以及资源空间对应的路径信息配置到通用总线译码器（UB decoder）表项中为例进行说明，则图 6 所示的方法流程还包括：

S620，第一主机配置译码器表项。

为了便于描述，下文中以第一主机将第一资源空间的信息配置到译码器表项中为例进行说明。其中，第一资源空间可以为第二主机提供的至少一个资源空间中的任意资源空间。第一主机将第一资源空间的信息配置到译码器表项中，译码器表项中包括第一条目，第一条目包括用于指示访问第一资源空间的第一访问报文对应的路径信息（可以简称为第一路径信息）和第一资源空间对应的保序属性信息。

例如，第一主机获得第二主机中多个资源空间的信息，其中，多个资源空间分别对应的保序属性可能不同。

在该实施例中，第一资源空间对应表项中的第一条目，其中，表项中一个条目（entry）的数据结构如图 9 所示，包括 UBA、传输组号（transport group number, TPG Number）、路径信息或者称为保序标识（order ID）、保序属性信息或者称为保序类型（order type）、目的对象实体标识（destination entity ID, DstEID）、令牌标识（TokenID）和令牌值（TokenValue）。

示例性地，条目中各个字段的含义如下表 2 所示：

表 2

| 名称（Name）   | 位宽（比特）Width(bit) | 描述（Description）                                                                                                                      |
|------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| UBA        | xx bit           | 经转换后输出到 UB 线路上 UBA [63:12]。低 12bit 与输入地址低 12bit 相同。                                                                                  |
| TPG Number | xx bit           | 该域段用于指示该段地址所对应的 TPG 号。                                                                                                               |
| Order ID   | xx bit           | 用于指示当前访问地址空间的流 Order ID。<br>建议作为与目的设备路径可达的多个源出端口中选择出一条唯一的出端口，并将此 order id 作为 LB 填充在报文中，并使能 HASH 路由机制，由此保证 load/store 全网走唯一一条路径到达接收侧； |
| Order Type | xx bit           | 用于指示当前访问地址空间的流的 Order 类型，比如：NO/RO/SO。该域段直接用作控制器填充报文中。                                                                                |
| DstEID     | xx bit           | 用于指示通信的目的对象 EID。                                                                                                                     |
| TokenID    | xx bit           | 用于指示通信 memory 的索引。                                                                                                                   |
| TokenValue | xx bit           | TokenValue，权限校验值。                                                                                                                    |

表 2 中所涉及的当前访问地址空间的流包括当前访问地址空间的至少一个报文，即访问相同的地址空间的一个或多个访问报文称为“访问地址空间的流”。

应理解，上述的译码器主要功能是：支持第一主机根据报文的物理地址（hard physical address, HPA）查询报文的 UBA 以及报文的部分域段转换信息功能。

示例性地，第一主机查询译码器获得以下信息：

TPG：指示传输当前报文的 TPG，该 TPG 中维护源端和目的端之间的多个路径可达的发送端口/路径。

Order ID：用于指示当前访问地址空间的流 Order ID，Order ID 作为 TPG 内部多个路径可达的发送端口/路径中选择出一条唯一的出端口/路径，并将此 Order ID 作为负载均衡（load balancing, LB）域填充在报文中，可以使能 HASH 路由机制，由此保证报文经过唯一一条路径到达接收侧。

Order Type: 用于指示当前访问地址空间的流的 Order 类型, 比如: NO、RO 或 SO。该域段直接用作控制器填充报文中“ODR”域段。

应理解, 第一主机可以获取第二主机中至少一个资源空间的信息, 并将该至少一个资源空间的信息配置到第一主机的通用总线译码器, 上述的第一资源空间为该至少一个资源空间中的任意一个。例如, 针对第一主机获取的第二主机中的第二资源空间的信息, 第一主机将第二资源空间的信息配置到译码器表项中, 译码器表项中包括第二条目, 第二条目包括用于指示访问第二资源空间的第二访问报文对应的路径信息和第二资源空间对应的保序属性信息。

作为示例而非限定, 上述的第二访问报文对应的路径信息(可以简称为第二路径信息)和第一访问报文对应的路径信息不同, 第二资源空间对应的保序属性信息和第一资源空间对应的保序属性信息不同, 即该实施例保序要求不同的访问报文可以通过不同的路径传输, 从而保证在多路径传输的场景下, 实现报文的保序传输。

上述的表项配置完成之后, 若第一主机向第二主机发起访问, 以期访问第二主机给第一主机提供的第一资源空间, 第一主机可以通过查询译码器表项获取第一访问报文需要满足的保序要求, 并基于查询译码器表项获取到的信息(如, 保序属性信息、传输路径信息等)生成第一访问报文, 则图 6 所示的方法流程还包括:

S630, 第一主机生成第一访问报文。

具体地, 第一主机组发第一访问报文时, 根据所查序属性信息, 填充发送数据包相关域段, 以及发送路径的选择。

示例性地, 第一访问报文的格式如图 10 所示, 包括但不限于路由模式(routing mode)域和 LB 域, 其中, RM 和 LB 含义如下表 3 所示:

表 3

|    |                                |
|----|--------------------------------|
| RM | 报文路由模式指示域段                     |
| LB | 报文中的负载分担因子, 可参与路由多路径负载分担时的哈希计算 |

需要说明的是第一访问报文中还包括其他信息, 该实施例中主要涉及报文中 RM 域和 LB 域。其他域可以参考目前通用总线协议中报文格式相关描述, 这里不再赘述。

具体地, 报文头中 RM 域定义可以如下表 4 所示:

表 4

| 报文头字段     | 解释                                                                                                 |
|-----------|----------------------------------------------------------------------------------------------------|
| RM (2bit) | Bit0: 为 0 表示基于流的 Hash, 为 1 表示自适应路由;<br>Bit1: 为 0 表示在路由表配置的全部路径中进行选路, 为 1 表示在路由表配置的路径中选择最短路径子集进行选路。 |

可选地, 若 RM 域指示当前第一访问报文的传输路径选择方式为基于流的 Hash 方式, 则第一主机和中间交换机(switch)可以基于元组进行 Hash 选路。例如, 基于流的 Hash 中, 第一主机和中间交换机可以选择报文头的一个或多个域段进行 Hash 选路, 如 IP 报文格式的五元组{srcIP、dstIP、protocol、传输层源端口、传输层目的端口}; 对于压缩报文格式, 可以使用{ srcCNA, dstCNA, LB}进行 Hash。Hash 选路也可以选择报文在网络设备的参数信息等参与 hash 计算(如交换机的入端口号等)。

进一步地, 第一访问报文的 TAH 中还包括 ODR 域, 其中, ODR 域定义可以如下表 5 所示:

表 5

| 名称<br>(Name) | 位宽 (比特)<br>Width(bit) | 描述 (Description)                                                                                                                                                            |
|--------------|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ODR          | 3                     | 请求报文 Order 属性标识。<br>ODR[1:0]:指示报文的保序属性, 其中:<br>2'b00:NO, 标识当前报文和其他报文没有任何保序要求;<br>2'b01:RO, 标识与后续的 SO 报文有保序要求的报文, SO 报文不能超越 RO 报文;<br>2'b10: SO, 标识有保序要求的报文, 和该报文与之前的 RO 报文有 |

|  |  |                                                                                                                                            |
|--|--|--------------------------------------------------------------------------------------------------------------------------------------------|
|  |  | 保序要求；<br>2'b11：保留。<br>ODR[2]：完成保序（completion order），表示当前报文和前面的报文存在 completion 序，给接收侧使用；<br>1'b0：NO，表示当前报文和前面报文没有保序要求；<br>1'b1：当前报文有完成序的要求。 |
|--|--|--------------------------------------------------------------------------------------------------------------------------------------------|

示例性地，第一主机组成第一访问报文之后，可以通过多路径网络（如图 4 中所示的网络）向第二主机发送该第一访问报文，则图 6 所示的方法流程还包括：

S640，第一主机向第二主机发送第一访问报文。

具体地，交换机基于第一访问报文中 RM 域确定多条路径中用于传输第一访问报文的路径。进一步地，第二主机接收到第一访问报文之后，可以根据第一访问报文中的“ODR”域确定第一访问报文的保序要求。

图 6 的通信方法中，通过资源注册阶段，识别出访问某些资源空间的报文之间没有序要求，可以乱序执行，或者识别出访问某些资源空间的报文有保序要求，也就是说可以提前获知访问不同资源空间的报文之间的保序要求。进一步地，将相关信息配置在译码器表项中，访问请求发送时，可以通过查询译码器表项中相关信息，得到对应的保序属性，并根据保序属性来填充报文相关域段，完成保序场景下，依旧能够将多路径利用起来。

为了便于理解，下面结合具体的示例说明图 6 所示的通信方法中，如何实现在多路径场景下报文的保序传输。

示例一：

步骤一：第一主机的 xPU 下发的 Load/store 操作，经内部总线到达第一主机中的通用总线控制器（unified bus controller）；

步骤二：通用总线控制器判断此 Load/store 操作需要通过通用总线链路发送至第二主机；

步骤三：通用总线控制器向译码器表项发起查表请求；

步骤四：根据 HPA 进行内部查表，获取到对应的信息，信息中包含了第二主机注册资源空间时提供的 UBA，Order ID，Order Type 等信息；

步骤五：通用总线控制器根据查译码器表项所得目的通信对象信息后，查询内部路由表，得到多个路径可达的“出端口/路径”，通用总线控制器使用“Order ID”作为多个路径可达的“出端口/路径”的选路输入，将相同“Order ID”的报文选路到同一个“出端口/路径”；

步骤六：通用总线控制器进行组包，将查询译码器表项得到的“order type”填充到报文的“ODR”，将“Order ID”填充到报文的“LB”，将报文中的“RM 的 bit0 设置为 0”，将其他信息（例如，UBA，DstEID，TokenID，.....）填充到报文对应位置。

步骤七：报文传输至交换机时，交换机检测到报文中“RM 的 bit0 设置为 0”，采用 Hash 路由机制，对于压缩报文格式，可以使用{ srcCNA, dstCNA, LB}进行 Hash，因此相同{ srcCNA, dstCNA, LB}可以经过确定的一条路径到达第二主机。

步骤八：对于第二主机，同一个源设备发过来的相同 Order ID 的 load/store 对应的报文从同一个物理端口接收，第二主机根据报文中的“ODR”进行序处理。

应理解，本申请实施例中的图 5 至图 10 所示的具体的例子只是为了帮助本领域技术人员更好地理解本申请实施例，而非限制本申请实施例的范围。还应理解，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

还应理解，在本申请的各个实施例中，如果没有特殊说明以及逻辑冲突，不同的实施例之间的术语和/或描述具有一致性、且可以相互引用，不同的实施例中的技术特征根据其内在的逻辑关系可以组合形成新的实施例。

上述主要从方法的角度对本申请实施例提供的方案进行了介绍。为了实现上述功能，其包含了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，本申请能够以硬件或硬件和计算机软件的结合形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行，取决于技术方案的特定应用和设计约束条件。专

业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

以下，结合图 11 至图 13 详细说明本申请实施例提供的通信装置。应理解，装置实施例的描述与方法实施例的描述相互对应，因此，未详细描述的内容可以参见上文方法实施例，为了简洁，部分内容不再赘述。

本申请实施例可以根据上述方法示例对第一主机或者第二主机进行功能模块的划分，例如，可以对各个功能划分各个功能模块，也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。需要说明的是，本申请实施例中模块的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。下面以采用对应各个功能划分各个功能模块为例进行说明。

图 11 示出了本申请实施例提供的一种通信装置 1000 的结构示意图。

一种示例，通信装置 1100 可以应用于第一主机，通信装置 1000 可以用于执行上述的通信方法，例如用于执行图 6 所示的方法。具体地，通信装置 1100 可以包括收发单元 1110 和处理单元 1120。

收发单元 1110，用于获取第二主机中第一资源空间的信息，所述第一资源空间的信息包括所述第一资源空间的起始地址、所述第一资源空间的长度信息和所述第一资源空间对应的保序属性信息，所述第一资源空间对应的保序属性信息用于指示对所述第一资源空间进行访问的第一访问报文对应的保序要求。处理单元 1120，用于将所述第一资源空间的信息配置到译码器表项中，所述译码器表项中包括第一条目，所述第一条目包括用于指示访问所述第一资源空间的第一访问报文对应的路径信息和所述第一资源空间对应的保序属性信息，其中，所述第一主机和所述第二主机之间包括多条传输路径。

作为示例，结合图 6，收发单元 1010 可以用于执行 S610 和 S640，处理单元 1020 可以用于执行 S620 和 S630。

需要说明的是，图 11 所述的装置也可以用于执行前述提及的附图所示的实施例变形所涉及的方法步骤，在此不再赘述。

另一种示例，通信装置 1100 可以应用于第二主机，通信装置 1100 可以用于执行上述的通信方法，例如用于执行图 5 所示的方法。

收发单元 1010，用于第二主机接收来自第一主机的第一访问报文，所述第一访问报文用于访问所述第二主机中的第一资源空间，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述访问报文的负载分担因子字段承载指示第一传输路径的信息。处理单元 1120，用于根据第一访问报文的保序域确定所述第一访问报文的保序要求；其中，接收第一传输路径上传输的至少一个访问报文的物理端口相同，所述第一访问报文为所述至少一个访问报文中的一个。

作为示例，结合图 6，收发单元 1010 可以用于执行 S610 和 S640。

需要说明的是，图 11 所述的装置也可以用于执行前述提及的附图所示的实施例变形所涉及的方法步骤，在此不再赘述。

本申请实施例还提供一种芯片系统 1200，如图 12 所示，该芯片系统 1200 包括至少一个处理器和至少一个接口电路。作为示例，当该芯片系统 1200 包括一个处理器和一个接口电路时，则该一个处理器可以是图 12 中实线框所示的处理器 1210(或者是虚线框所示的处理器 1210)，该一个接口电路可以是图 12 中实线框所示的接口电路 1220(或者是虚线框所示的接口电路 1220)。

当该芯片系统 1200 包括两个处理器和两个接口电路时，则该两个处理器包括图 12 中实线框所示的处理器 1210 和虚线框所示的处理器 1210，该两个接口电路包括图 12 中实线框所示的接口电路 1220 和虚线框所示的接口电路 1220。对此不作限定。处理器 1210 和接口电路 1220 可通过线路互联。例如，接口电路 1220 可用于接收信号(例如存储器中存储的指令等)。又例如，接口电路 1220 可用于向其它装置(例如处理器 1210)发送信号。

示例性的，接口电路 1220 可读取存储器中存储的指令，并将该指令发送给处理器 1210。当该指令被处理器 1210 执行时，可使得通信装置执行上述实施例中的各个步骤。当然，该芯片系统 1200 还可以包含其他分立器件，本申请实施例对此不作具体限定。

本申请另一实施例还提供一种计算机可读存储介质，该计算机可读存储介质中存储有指令，当指令在通信装置上运行时，该通信装置执行上述方法实施例所示的方法流程中该通信装置执行的各个步骤。在一些实施例中，所公开的方法可以实施为以机器可读格式被编码在计算机可读存储介质上的或者被编

码在其它非瞬时性介质或者制品上的计算机程序指令。

图 13 示意性地示出本申请实施例提供的计算机程序产品的概念性局部视图，该计算机程序产品包括用于在计算机设备上执行计算机进程的计算机程序。

在一个实施例中，计算机程序产品是使用信号承载介质 1300 来提供的。该信号承载介质 1300 可以包括一个或多个程序指令，其当被一个或多个处理器运行时可以提供以上针对图 6 描述的功能或者部分功能。因此，例如，参考图 6 中 S610~S640 的一个或多个特征可以由与信号承载介质 1300 相关联的一个或多个指令来承担。此外，图 13 中的程序指令也描述示例指令。

在一些示例中，信号承载介质 1300 可以包含计算机可读介质 1301，诸如但不限于，硬盘驱动器、紧密封盘(CD)、数字视频光盘(DVD)、数字磁带、存储器、只读存储记忆体 readonly memory, ROM)或随机存储记忆体(random access memory, RAM)等等。

在一些实施方式中，信号承载介质 1300 可以包含计算机可记录介质 1302，诸如但不限于，存储器、读/写(R/W)CD、R/W DVD、等等。

在一些实施方式中，信号承载介质 1300 可以包含通信介质 1303，诸如但不限于，数字和/或模拟通信介质(例如，光纤电缆、波导、有线通信链路、无线通信链路、等等)。信号承载介质 1300 可以由无线形式的通信介质 1303 来传达。一个或多个程序指令可以是，例如，计算机可执行指令或者逻辑实施指令。

在一些示例中，响应于通过计算机可读介质 1301、计算机可记录介质 1302、和/或通信介质 1303 中的一个或多个程序指令，提供各种操作、功能、或者动作。

应该理解，这里描述的布置仅仅是用于示例的目的。因而，本领域技术人员将理解，其它布置和其它元素(例如，机器、接口、功能、顺序、和功能组等等)能够被取而代之地使用，并且一些元素可以根据所期望的结果而一并省略。另外，所描述的元素中的许多是可以被实现为离散的或者分布式的组件的、或者以任何适当的组合和位置来结合其它组件实施的功能实体。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具體工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用，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，本申请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)执行本申请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(Read-Only Memory, ROM)、随机存取存储器(Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

## 权利要求书

1. 一种通信方法，其特征在于，所述方法包括：

第一主机获取第二主机中第一资源空间的信息，所述第一资源空间的信息包括所述第一资源空间的起始地址、所述第一资源空间的长度信息和所述第一资源空间对应的保序属性信息，所述第一资源空间对应的保序属性信息用于指示对所述第一资源空间进行访问的第一访问报文对应的保序要求；

所述第一主机保存所述第一资源空间对应的保序属性信息和第一路径信息，所述第一路径信息用于指示访问所述第一资源空间的第一访问报文对应的传输路径，

其中，所述第一路径信息所指示的传输路径为所述第一主机和所述第二主机之间包括多条传输路径中的一条。

2. 根据权利要求1所述的方法，其特征在于，所述第一主机保存所述第一资源空间对应的保序属性信息和第一路径信息，包括：

所述第一主机将所述第一资源空间的信息配置到译码器表项中，所述译码器表项中包括第一条目，所述第一条目包括所述第一路径信息和所述第一资源空间对应的保序属性信息。

3. 根据权利要求2所述的方法，其特征在于，在所述第一主机确定向所述第二主机发送第一访问报文访问所述第一资源空间时，所述方法包括：

所述第一主机查询所述译码器表项确定所述第一路径信息和所述第一访问报文对应的保序要求；

所述第一主机根据所述第一路径信息确定所述第一访问报文对应的输出端口和第一传输路径；

其中，所述第一传输路径为所述多条传输路径中用于传输所述第一访问报文的传输路径，所述第一传输路径用于传输相同路径信息对应的访问报文。

4. 根据权利要求3所述的方法，其特征在于，所述方法还包括：

所述第一主机根据所述第一传输路径和所述第一访问报文对应的保序要求生成所述第一访问报文，

其中，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述第一访问报文的负载分担因子字段承载指示所述第一传输路径的信息。

5. 根据权利要求4所述的方法，其特征在于，所述第一访问报文的路由指示字段设置为0，

其中，所述路由指示字段设置为0用于指示采用哈希路由机制确定所述第一访问报文的传输路径。

6. 根据权利要求1至5中任一项所述的方法，其特征在于，所述第一主机获取所述第一资源空间对应的保序属性信息，包括：

所述第一主机接收来自所述第二主机的所述第一资源空间对应的保序属性信息；或者，

所述第一主机根据所述第一资源空间的用途确定所述第一资源空间对应的保序属性信息。

7. 根据权利要求1至6中任一项所述的方法，其特征在于，所述方法还包括：

所述第一主机获取所述第二主机中第二资源空间的信息，所述第二资源空间的信息包括所述第二资源空间的起始地址、所述第二资源空间的长度信息和所述第二资源空间对应的保序属性信息，所述第二资源空间对应的保序属性信息用于指示对所述第二资源空间进行访问的第二访问报文对应的保序要求；

所述第一主机保存所述第二资源空间对应的保序属性信息和第二路径信息，所述第二路径信息用于指示访问所述第二资源空间的第二访问报文对应的传输路径，

其中，所述第二路径信息和所述第一路径信息不同，所述第二资源空间对应的保序属性信息和所述第一资源空间对应的保序属性信息不同。

8. 根据权利要求7所述的方法，其特征在于，所述第一主机保存所述第二资源空间对应的保序属性信息和第二路径信息，包括：

所述第一主机将所述第二资源空间的信息配置到译码器表项中，所述译码器表项中包括第二条目，所述第二条目包括所述第二路径信息和所述第二资源空间对应的保序属性信息。

9. 根据权利要求1至8中任一项所述的方法，其特征在于，所述第一访问报文对应的保序要求包括以下任意一种：

强保序SO、无需保序NO、或灵活保序RO。

10. 一种通信方法，其特征在于，所述方法包括：

第二主机接收来自第一主机的第一访问报文，所述第一访问报文用于访问所述第二主机中的第一资源

源空间，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述访问报文的负载分担因子字段承载指示第一传输路径的信息；

所述第二主机根据所述第一访问报文的保序域确定所述第一访问报文的保序要求；

其中，所述第二主机接收所述第一传输路径上传输的至少一个访问报文的物理端口相同，所述第一访问报文为所述至少一个访问报文中的一个。

11. 一种通信方法，其特征在于，包括：

第一主机获取第二主机中第一资源空间的信息，所述第一资源空间的信息包括所述第一资源空间的起始地址、所述第一资源空间的长度信息和所述第一资源空间对应的保序属性信息，所述第一资源空间对应的保序属性信息用于指示对所述第一资源空间进行访问的第一访问报文对应的保序要求；

所述第一主机保存所述第一资源空间对应的保序属性信息和第一路径信息，所述第一路径信息用于指示访问所述第一资源空间的第一访问报文对应的传输路径，

所述第二主机接收来自所述第一主机的第一访问报文，所述第一访问报文用于访问所述第二主机中的第一资源空间，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述访问报文的负载分担因子字段承载指示第一传输路径的信息；

所述第二主机根据所述第一访问报文的保序域确定所述第一访问报文的保序要求；

其中，所述第二主机接收所述第一传输路径上传输的至少一个访问报文的物理端口相同，所述第一访问报文为所述至少一个访问报文中的一个。

12. 一种通信系统，其特征在于，包括第一主机和至少一个第二主机，

所述第一主机，用于获取第二主机中第一资源空间的信息，所述第一资源空间的信息包括所述第一资源空间的起始地址、所述第一资源空间的长度信息和所述第一资源空间对应的保序属性信息，所述第一资源空间对应的保序属性信息用于指示对所述第一资源空间进行访问的第一访问报文对应的保序要求；

所述第一主机，用于保存所述第一资源空间对应的保序属性信息和第一路径信息，所述第一路径信息用于指示访问所述第一资源空间的第一访问报文对应的传输路径，

所述第二主机，用于接收来自所述第一主机的第一访问报文，所述第一访问报文用于访问所述第二主机中的第一资源空间，所述第一访问报文的保序域中承载指示所述第一访问报文对应的保序要求的信息，所述访问报文的负载分担因子字段承载指示第一传输路径的信息；

所述第二主机根据所述第一访问报文的保序域确定所述第一访问报文的保序要求；

其中，所述第二主机接收所述第一传输路径上传输的至少一个访问报文的物理端口相同，所述第一访问报文为所述至少一个访问报文中的一个。

13. 一种通信装置，其特征在于，包括：处理器，用于读取存储器中存储的指令，当所述处理器执行所述指令时，使得所述通信装置实现权利要求 1 至 10 中任一项所述的方法；或者，使得所述通信装置实现权利要求 11 所述的方法。

14. 一种计算机程序产品，其特征在于，所述计算机程序产品包括计算机程序代码，当所述计算机程序代码在计算机上运行时，权利要求 1 至 10 中任一项所述的方法被执行；或者，当所述计算机程序代码在计算机上运行时，权利要求 11 所述的方法被执行。

15. 一种计算机可读存储介质，其特征在于，包括计算机程序，当其在计算机系统上运行时，使得所述计算机系统在处理模块执行如权利要求 1 至 10 中任一项所述的方法；或者，当其在计算机系统上运行时，使得所述计算机系统在处理模块执行如权利要求 11 所述的方法。