(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 103222286 B
(45)授权公告日 2014.07.09

(21)申请号 201180020891.1
(22)申请日 2011.11.18
(85)PCT国际申请进入国家阶段日
2012.11.05
(86)PCT国际申请的申请数据
PCT/CN2011/082442 2011.11.18
(87)PCT国际申请的公布数据
WO2012/167566 ZH 2012.12.13
(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 刘云海 雕峻峰
(51)Int.Cl.
H04W 4/18(2006.01)

(56)对比文件
US 2005238035 A1,2005.10.27, 全文.
CN 101394349 A,2009.03.25, 说明书第4页第19行至第12页第6行,附图1,2.
CN 101222495 A,2008.07.16, 全文.
US 2010082849 A1,2010.04.01, 全文.

审查员 高菲

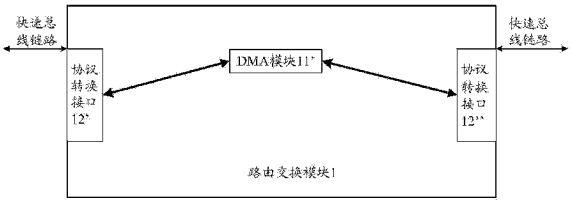
权利要求书3页 说明书8页 附图6页

(54)发明名称

路由交换装置、网络交换系统和路由交换方法

(57)摘要

本发明涉及路由交换装置、网络交换系统和路由交换方法。该路由交换装置包括：一个或多个直接内存访问模块和至少两个协议转换接口，其中所述直接内存访问模块用于产生跨网络节点的连续的访问请求并控制所述至少两个协议转换接口中的数据传输，每个所述协议转换接口用于转换所述路由交换装置内部与外部传输的数据的通信协议并连接所述路由交换模块与外部网络节点。从而，本发明可以通过引入路由交换装置替代网络交换机，使跨节点的内存访问和IO空间访问可以不通过代理而直接进行，进而减少跨节点的内存访问和IO空间访问的延时，提高系统的整体性能。



CN 103222286 B

1. 一种路由交换装置，其特征在于，包括：

一个或多个直接内存访问模块；

至少两个协议转换接口，

其中所述直接内存访问模块用于产生跨网络节点的连续的访问请求并控制所述至少两个协议转换接口中的数据传输，每个所述协议转换接口用于转换所述路由交换装置内部与外部传输的数据的通信协议并连接所述路由交换模块与外部网络节点；所述直接内存访问模块包括：

直接内存访问控制器；以及

直接内存访问通道；

其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道与所述协议转换接口的连接；所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；所述协议转换接口包括：

快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；

局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；

寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；

控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转换接口的性能或操作。

2. 根据权利要求 1 所述的路由交换装置，其特征在于，所述网络节点是服务器节点或者另一路由交换装置。

3. 一种网络交换系统，其特征在于，包括：

至少两个服务器节点；

一个或多个路由交换装置，其中每个所述路由交换装置通过快速总线链路与所述至少两个服务器节点或者另一所述路由交换装置连接，并且包括一个或多个直接内存访问模块和至少两个协议转换接口，其中所述直接内存访问模块用于产生跨网络节点的连续的访问请求并控制所述至少两个协议转换接口中的数据传输，每个所述协议转换接口用于转换所述路由交换装置内部与外部传输的数据的通信协议并连接所述路由交换模块与外部网络节点；

所述直接内存访问模块包括：

直接内存访问控制器；以及

直接内存访问通道；

其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道与所述协议转换接口的连接；

所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；

所述协议转换接口包括：

快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；

局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；

寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；

控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转换接口的性能或操作。

4. 根据权利要求 3 所述的网络交换系统，其特征在于，所述至少两个服务器节点包括第一服务器节点与第二服务器节点，所述路由交换装置至少包括：

直接内存访问模块，控制所述第一服务器节点、所述第二服务器节点分别与所述路由交换装置之间的数据传输；

第一协议转换接口，通过快速总线链路与所述第一服务器节点连接，并转换在所述第一服务器节点与所述直接内存访问模块之间传输的数据的通信协议；

第二协议转换接口，通过快速总线链路与所述第二服务器节点连接，并转换在所述第二服务器节点与所述直接内存访问模块之间传输的数据的通信协议。

5. 根据权利要求 3 或 4 所述的网络交换系统，其特征在于，每个所述服务器节点还包括一个符合快速总线链路规范的插槽，每个所述路由交换装置还包括一个符合快速总线链路规范的插槽。

6. 一种由路由交换装置执行的路由交换方法，其特征在于，所述路由交换装置至少包括直接内存访问模块、第一协议转换接口及第二协议转换接口，所述第一协议转换接口和所述第二协议转换接口分别用于转换所述路由交换装置内部与外部传输的数据的通信协议；所述方法包括：

所述直接内存访问模块通过配置接口获取访问参数；

所述直接内存访问模块根据所述访问参数，通过所述第一协议转换接口向与其连接的网络节点发出读请求，并从所述网络节点读出传输数据；

所述直接内存访问模块将获得的所述传输数据传送到第二协议转换接口，并通过所述第二协议转换接口向与其连接的网络节点发出写请求，并将所述传输数据写入与所述第二协议转换接口连接的网络节点；所述直接内存访问模块包括直接内存访问控制器和直接内存访问通道，其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道分别与所述第一协议转换接口、所述第二协议转换接口的连接；

所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；

所述第一协议转换接口和所述第二协议转换接口中的至少一个包括：

快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；

局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；

寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；

控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转换接口的性能或操作。

7. 根据权利要求 6 所述的方法，其特征在于，所述配置接口是快速总线协议接口或者系统管理总线接口。

8. 根据权利要求 6 中任一项所述的方法，其特征在于，所述访问参数包括数据读取的

网络节点的起始地址及数据大小，以及数据写入的网络节点的数据缓冲区。

9. 根据权利要求 6 至 8 中任一项所述的方法，其特征在于，所述网络节点是服务器节点。

路由交换装置、网络交换系统和路由交换方法

技术领域

[0001] 本发明实施例涉及信息技术领域，并且更具体地，涉及路由交换装置、网络交换系统、路由交换的方法。

背景技术

[0002] 在目前的服务器硬件市场上，单个处理器的计算能力和储存处理能力都是非常有限的，但在某些特殊的应用场合，对服务器的处理能力和内存容量都有很高的要求，例如银行的账务系统的处理，或者通讯交换中心的数据库处理，或者气象信息的分析处理等。为了满足这些特殊的应用，往往需要把多个处理器连接在一起，形成一个大型的计算系统，这就涉及到处理器或计算系统之间的互联。通常可以由两种方法把多个处理器或计算系统连接起来，一种是通过专门协议把多个处理器连接起来，另一种是通过通用协议（例如PCIE(Peripheral Component Interconnect Express，快速周边元件扩展接口)）把多个小计算系统连接起来形成一个大的计算系统。

[0003] 对于采用通用协议的计算系统，如果一个网络节点的处理器要访问另一个网络节点的内存或输入输出 (Input/Output，简写为 IO) 空间，该网络节点需要将相关命令和参数通过网络协议发送到另一个网络节点，由该另一个网络节点的处理器代为执行相关的访问命令，这样跨节点的内存访问和 IO 空间访问的效率很低、速度很慢，从而严重影响系统性能。

发明内容

[0004] 本发明实施例提供一种路由交换装置、网络交换系统和路由交换的方法，能够解决跨节点内存和 IO 空间访问的延时问题。

[0005] 一方面，提供了一种路由交换装置，包括一个或多个直接内存访问 DMA(Direct Memory Access) 模块和至少两个协议转换接口，其中 DMA 模块用于产生跨网络节点的连续的访问请求并控制该至少两个协议转换接口中的数据传输，每个协议转换接口用于转换路由交换装置内部与外部传输的数据的通信协议并连接路由交换模块与外部网络节点；

[0006] 所述直接内存访问模块包括：

[0007] 直接内存访问控制器；以及直接内存访问通道；

[0008] 其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道与所述协议转换接口的连接；所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；所述协议转换接口包括：

[0009] 快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转

换接口的性能或操作。

[0010] 另一方面，提供了一种网络交换系统，包括：至少两个服务器节点；一个或多个路由交换装置，其中每个路由交换装置通过快速总线链路（Fast Bus Link）与至少两个服务器节点或者另一路由交换装置连接，并且包括一个或多个 DMA 模块和至少两个协议转换接口，其中 DMA 模块用于产生跨网络节点的连续的访问请求并控制该至少两个协议转换接口中的数据传输，每个协议转换接口用于转换路由交换装置内部与外部传输的数据的通信协议并连接路由交换模块与外部网络节点；

[0011] 所述直接内存访问模块包括：直接内存访问控制器；以及直接内存访问通道；其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道与所述协议转换接口的连接；所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；所述协议转换接口包括：快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转换接口的性能或操作。

[0012] 再一方面，提供了一种由路由交换装置执行的路由交换方法，路由交换装置至少包括 DMA 模块、第一协议转换接口及第二协议转换接口，所述第一协议转换接口和所述第二协议转换接口分别用于转换所述路由交换装置内部与外部传输的数据的通信协议；该方法包括：DMA 模块通过配置接口获取配置信息和访问参数；DMA 模块通过第一协议转换接口向与其连接的网络节点发出读请求，从该网络节点读出传输数据；DMA 模块将获得的传输数据传送到第二协议转换接口，并通过第二协议转换接口向与其连接的网络节点发出写请求，并将所述传输数据写入与第二协议转换接口连接的网络节点；所述直接内存访问模块包括直接内存访问控制器和直接内存访问通道，其中所述直接内存访问控制器根据配置信息控制所述直接内存访问通道分别与所述第一协议转换接口、所述第二协议转换接口的连接；所述直接内存访问模块还包括存储器，作为所述直接内存访问通道的缓冲区；所述第一协议转换接口和所述第二协议转换接口中的至少一个包括：快速总线协议功能块，通过快速总线链路与所述外部的网络节点进行数据交换；局部总线接口功能块，其与所述快速总线协议功能块连接，用于所述路由交换装置内部的数据传输；寄存器，与所述快速总线协议功能块和所述局部总线接口功能块连接，所述寄存器中存储有用于配置所述快速总线协议功能块和所述局部总线接口功能块的指令；控制功能块，与所述快速总线协议功能块和所述局部总线接口功能块连接，用于控制所述协议转换接口的性能或操作。

[0013] 本发明实施例可以通过引入路由交换装置替代网络交换机，使跨节点的内存访问和 IO 空间访问可以不通过代理而直接进行，进而减少跨节点的内存访问和 IO 空间访问的延时，提高系统的整体性能。

附图说明

[0014] 为了更清楚地说明本发明实施例的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实

施例，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的附图。

[0015] 图 1 是根据本发明实施例的路由交换装置的结构示意图。

[0016] 图 2 是根据本发明实施例的路由交换装置中 DMA 模块与协议转换接口的结构示意图。

[0017] 图 3 是根据本发明实施例的网络交换系统的结构示意图。

[0018] 图 4 是根据本发明实施例的由路由交换装置执行的路由交换方法的流程图。

[0019] 图 5 是根据本发明实施例的网络交换系统中路由交换方法的流程图。

[0020] 图 6 是根据本发明一个具体实施例的通过路由交换装置实现数据读取访问的流程图。

[0021] 图 7 是根据本发明另一具体实施例的通过路由交换装置实现数据写入访问的流程图。

具体实施方式

[0022] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动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。

[0023] 图 1 示出了根据本发明实施例的路由交换装置。该路由交换装置是一个硬件设备，它可以把多个网络节点连接起来。这里所述网络节点可以是服务器节点，也或者是另一路由交换装置。

[0024] 在图 1 中，路由交换装置 1 包括一个或多个 DMA 模块 11 和至少两个协议转换接口 12。在图 1 中仅示意性地示出了一个 DMA 模块和两个协议转换接口，即 DMA 模块 11’、协议转换接口 12’和协议转换接口 12’’。本领域技术人员很容易理解，可以根据应用的需要设计路由交换模块 1 具有一个以上的 DMA 模块以及两个以上的协议转换接口。一般地，DMA 模块 11’会产生跨网络节点的访问请求，并进而控制两个协议转换接口 12’、12’’之间的数据传输，协议转换接口 12’和协议转换接口 12’’转换将来自外部的网络节点的数据的通信协议转换成路由交换装置 1 内部数据传输所需的通信协议，或者将路由交换装置 1 内部数据传输所需的通信协议转换成来自外部的网络节点的数据的通信协议，并连接路由交换装置 1 与外部的网络节点。

[0025] 图 2 示例性地示出了 DMA 模块 11 与协议转换接口 12 的内部结构图。

[0026] 例如，在图 2 中，DMA 模块 11 包括 DMA 控制器 111 和多个 DMA 通道 112。其中，DMA 控制器 111 根据配置信息控制 DMA 通道 112 与协议转换接口 12 的连接。此外，DMA 模块 11 还可以包括多个存储器 113，通常存储器 113 作为 DMA 通道 112 的缓冲区。例如，为了便于理解，在图 2 中具体示出了存储器 113 与 DMA 通道 112 一一对应的关系，DMA 通道 112_1 至 DMA 通道 112_n，存储器 113_1 至存储器 113_n。当然，存储器 113 也可以与 DMA 通道 112 不一一对应。所有的 DMA 通道 112 均具有相同的功能与配置，所有的存储器 113 也可具有相同的功能与配置。存储器 113 例如可以是寄存器或 RAM(random access memory，随机存取存储器)，其数据容量可以是 2 个字节、4 个字节、8 个字节、32 个字节等，本发明不限制其容

量大小。当通过一个 DMA 通道 112 传输数据的 2 个网络节点的速度相差明显时，就可以用存储器 113 进行数据缓冲，避免传输的拥塞。

[0027] 协议转换接口 12 包括快速总线协议功能块 121、局部总线接口功能块 122、寄存器 123 和控制功能块 124。具体而言，快速总线协议功能块 121 通过快速总线链路与外部的网络节点进行数据交换。该快速总线协议功能块 121 是实现、支持快速总线链路协议的功能模块，保证各网络节点之间可以通过快速总线链路进行数据传输。所谓快速总线协议是把路由交换装置 1 与各网络节点连接起来进行数据传输的通用协议，例如，PCIE 协议及其升级、或者其他通用协议。快速总线协议功能块 121 是网络节点和路由交换装置 1 之间的命令、数据的传输通道，也可用于配置和管理路由交换装置。而局部总线接口功能块 122 与快速总线协议功能块 121 连接，并用于路由交换装置 1 内部的数据传输。以保证数据从一个网络节点正确地传送到另一个网络节点。寄存器 123 与快速总线协议功能块 121 和局部总线接口功能块 122 连接，其中存储有用于配置快速总线协议功能块 121 和局部总线接口功能块 122 的指令，以便对快速总线协议功能块 121 和局部总线接口功能块 122 以及协议转换接口 12 内部的其他功能块进行配置。控制功能块 124 与快速总线协议功能块 121 和局部总线接口功能块 122 连接，并用于控制协议转换接口 12 的性能或操作，例如能耗管理、热插拔管理等。

[0028] 综上所述，协议转换接口 12 用于在 2 种协议之间进行转换，以实现在不同的实体之间进行数据传输。从一个协议转换接口 12 流向各网络节点、或者从各网络节点流向另一协议转接口 12 的信息格式以及传递信息的方法一定是符合协议要求的。在路由交换装置 1 内部进行数据的传输时使用的是内部数据传输协议，在不同的实施例中可以采用不同的数据传输协议。协议转换接口 12 有 2 个主要的目的 :第一，在网络节点和路由交换装置 1 之间进行通信协议转换 ;第二，把路由交换装置 1 的内部资源空间与各网络节点的空间隔离开来，使各网络节点和路由交换装置 1 成为相互独立的实体，即便路由交换装置 1 内部也是支持快速总线链路协议。

[0029] 通过协议转换接口 12 可以把路由交换装置 1 与多个网络节点连接在一起。同样也可以通过协议转换接口 12 把多个路由交换装置 1 级联起来，从而形成一个更大的路由交换装置网络。

[0030] 需要说明的是，图 2 所示的路由交换装置 1 内部的功能模块划分只是为了描述方便，并非表示路由交换装置 1 只能由这些功能模块组成，路由交换装置 1 还可以包含其它的功能模块，图中所示的功能块也可以组合、分解从而形成新的功能模块。这些变化均在本发明的覆盖范围内。

[0031] 下面将结合图 3 描述根据本发明实施例的网络交换系统。

[0032] 在图 3 中，网络交换系统 3 包括至少两个服务器节点 2 以及一个或多个路由交换装置 1，其中每个路由交换装置 1 通过快速总线链路与服务器节点 2 或者另一个路由交换装置 1 连接。例如，如图 3 所示，示出了第一服务器节点 2’与第二服务器节点 2’’。路由交换装置 1 包括至少一个 DMA 模块 11，例如 DMA 模块 11’，该 DMA 模块 11’控制第一服务器节点 2’、第二服务器节点 2’’ 分别与路由交换装置 1 之间的数据传输。第一协议转换接口 12’通过快速总线链路与第一服务器节点 2’连接，并转换在第一服务器节点 2’与 DMA 模块 11’之间传输的数据的通信协议。第二协议转换接口 12’’ 通过快速总线链路与第二服务器节

点 2’’ 连接，并转换在第二服务器节点 2’’ 与 DMA 模块 11’ 之间传输的数据的通信协议。

[0033] 可考虑在每个服务器节点 2 上设计一个符合快速总线规范的插槽，或者在每个路由交换装置 1 上也设计一个符合快速总线规范的插槽。这样就可以根据需要采用一头或两头带插头的链路连线将网络节点与路由交换装置 1 进行连接。

[0034] 图 4 示出了根据本发明实施例的由路由交换装置 1 执行的路由交换方法，其中路由交换装置 1 至少包括 DMA 模块 11、第一协议转换接口 12’及第二协议转换接口 12’’。由路由交换装置 1 执行的路由交换方法包括下列步骤。

[0035] 41，DMA 模块 11 通过配置接口获取访问参数。

[0036] 配置接口可以是快速总线协议接口或者其他类型的接口，例如系统管理总线 SMBus(System Management Bus) 接口，也可通过该配置接口获取配置信息。

[0037] 一般而言，配置信息包括用于配置路由交换装置 1 所需的信息，例如工作频率、工作方式、数据传输类型、通道选择、DMA 模块选择等等。而访问参数可以包括被读取数据的网络节点（例如，服务器节点）的起始地址及数据大小，以及被写入数据的网络节点的数据缓冲区。

[0038] 当路由交换装置 1 通过配置接口获取了配置信息以及访问信息之后，就可以根据配置信息对自身进行配置，同时也明确了数据访问将涉及的外部网络节点信息等内容。

[0039] 42，于是，DMA 模块 11 通过第一协议转换接口 12’ 向与其连接的网络节点发出读请求，并从该网络节点读出传输数据。

[0040] 43，之后，DMA 模块 11 将获得的传输数据传送到第二协议转换接口 12’’，并通过第二协议转换接口 12’’ 向与其连接的网络节点发出写请求，最后将传输数据写入与第二协议转换接口 12’’ 连接的网络节点。

[0041] 在图 4 中仅仅示意性地描述了 DMA 模块 11 与第一协议转换接口 12’、第二协议转换接口 12’’ 实现路由交换方法的流程。本领域技术人员很容易理解，路由交换装置 1 中可以包括多于 2 个的协议转换接口以及多个 DMA 模块 11，不同的 DMA 模块 11 可以与不同的协议转换接口组合，这样可以实现并行处理，以进一步提高路由交换装置 1 的处理速度。

[0042] 为了简化说明，下面的描述仍以一个 DMA 模块 11 与两个协议转换接口 12’、12’’ 为例。应理解，路由交换装置 1 中任意一个 DMA 模块以及与该 DMA 模块关联的两个协议转换接口之间的数据访问操作与这里描述的情况基本相同。

[0043] 这里，假设第一协议转换接口 12’ 与一个网络节点（例如，第一服务器节点）电连接，第二协议转换接口 12’’ 与另一个网络节点（例如，第二服务器节点）电连接。在第二服务器节点需要读取第一服务器节点中的数据的情况下，或者在第一服务器节点需要向第二服务器节点写入数据的情况下，路由交换装置 1 中的 DMA 模块 11 就通过第一协议转换接口 12’ 向与其连接的第一服务器节点发出读请求，进而从该第一服务器节点读出传输数据。当传输数据通过第一协议转换接口 12’ 传送回路由交换装置 1 中，并经由 DMA 模块 11 传送到第二协议转换接口 12’’ 后，DMA 模块 11 又通过第二协议转换接口 12’’ 向与其连接的第二服务器节点发出写请求，最后将传输数据写入与第二协议转换接口 12’’ 连接的第二服务器节点。

[0044] 在上述过程中，本领域技术人员可理解，其中的“第一”、“第二”的表述并非用于指定具体的部件，而仅仅为了方便说明。在第一服务器节点需要读取第二服务器节点中的数

据的情况下，或者在第二服务器节点需要向第一服务器节点写入数据的情况下，路由交换装置 1 可以由 DMA 模块 11 通过第二协议转换接口 12’’向第二服务器节点发出读请求并读出传输数据，再通过第一协议转换接口 12’向第一服务器节点发出写请求并写入传输数据来实现数据访问的过程。

[0045] 进一步地，对于 DMA 模块 11 而言，该 DMA 模块 11 包括 DMA 控制器 111 和 DMA 通道 112，其中 DMA 控制器 111 根据配置信息控制 DMA 通道 112 分别与第一协议转换接口 12’、第二协议转换接口 12’’ 的连接，如图 2 和图 3 所示。本领域技术人员可以理解，当路由交换装置 1 的 DMA 模块 11 依据接收到的访问参数确定需要进行数据访问的网络节点后，DMA 控制器 111 就会为该次数据访问配置一个 DMA 通道 112，并使得该 DMA 通道 112 通过内部总线与两个协议转换接口连接，其中这两个协议转换接口分别与需要进行数据访问的网络节点连接。

[0046] 综上所述，路由交换装置 1 实现数据访问的路由交换方法就是先获取必要的配置信息和访问参数，这里的配置信息和访问参数不一定要由上述需要数据传输的第一服务器节点与第二服务器节点提供，也可以由其他的网络节点或设备提供。之后，由路由交换装置 1 通过 DMA 模块 11 经由一个协议转换接口向能够提供传输数据的网络节点发出读请求并读出相关数据，再经由另一个协议转换接口向需要传输数据的网络节点发出写请求并写入读出的数据来实现数据访问的路由交换方法。从而，本发明实施例的由路由交换装置执行的路由交换方法由于将路由交换装置替代了现有技术的网络交换机，使跨节点的内存访问和 IO 空间访问可以不通过代理而直接进行，进而减少跨节点的内存访问和 IO 空间访问的延时，提高系统的整体性能。

[0047] 以下将结合图 5 至图 7 分别描述具有网络节点和路由交换装置等实体的网络交换系统中的路由交换方法。其中，该网络交换系统包括至少两个服务器节点以及一个或多个路由交换装置，其中每个路由交换装置通过快速总线链路分别与该至少两个服务器节点或者另一所述路由交换装置连接。下面以仅涉及一个路由交换装置以及两个网络节点（即第一服务器节点和第二服务器节点）的最简单的网络交换系统为例，例如如图 3 所示的网络交换系统，说明在该网络交换系统中各实体如何实现路由交换方法的过程。具体包括下列步骤：

[0048] 51，通过配置接口将访问参数传送给路由交换装置 1，并启动路由交换装置 1 开始工作。

[0049] 52，然后，该路由交换装置 1 控制第一服务器节点 2’对第二服务器节点 2’’ 进行读写操作。

[0050] 53，最后，当第一服务器节点 2’对第二服务器节点 2’’ 的读写操作结束后，该路由交换装置 1 停止第一服务器节点 2’ 与第二服务器节点 2’’ 之间的数据交换。

[0051] 在路由交换装置 1 停止第一服务器节点 2’ 与第二服务器节点 2’’ 之间的数据交换之后，该路由交换装置 1 将通知第一服务器节点 2’ 数据的读写操作完毕，以便第一服务器节点 2’ 结束本次数据访问操作。

[0052] 如图 3 所示，路由交换装置 1 至少包括 DMA 模块 11、第一协议转换接口 12’和第二协议转换接口 12’’。第一协议转换接口 12’通过快速总线链路与第一服务器节点 2’连接，第二协议转换接口 12’’ 通过快速总线链路与第二服务器节点 2’’ 连接。进一步地，DMA 模

块 11 包括 DMA 控制器 111 和 DMA 通道 112，其中 DMA 控制器 111 根据配置信息控制 DMA 通道 112 分别与第一协议转换接口 12’、第二协议转换接口 12’’ 的连接。

[0053] 因此，路由交换装置 1 控制第一服务器节点 2’对第二服务器节点 2’’进行读写操作包括：DMA 模块 11 通过第二协议转换接口 12’’不断从第二服务器节点 2’’读出数据，并通过第一协议转换接口 12’将所读出的数据写入第一服务器节点 2’，直到从第二服务器节点 2’’读出全部数据；或者 DMA 模块 11 通过第一协议转换接口 12’不断从第一服务器节点 2’读出数据，并通过第二协议转换接口 12’将所读出的数据写入第二服务器节点 2’’，直到向第二服务器节点 2’’ 写入全部数据。

[0054] 具体而言，第一服务器节点 2’通过路由交换装置 1 从第二服务器节点 2’’读出数据的过程包括以下步骤，如图 6 所示。61，DMA 模块 11 通过第二协议转换接口 12’’ 向第二服务器节点 2’’ 发出读请求。62，第二服务器节点 2’’ 依据该读请求通过第二服务器节点 2’’ 的内部总线将所读出的数据传送到第二协议转换接口 12’’。63，DMA 控制器 111 控制 DMA 通道 112 将所述数据传送到第一协议转换接口 12’。然后，64，DMA 模块 11 通过第一协议转换接口 12’ 向第一服务器节点 2’ 发出写请求。65，第一服务器节点 2’ 依据该写请求写入所述数据。最后，66，判断是否全部所需数据均读取完毕，如果不是，则获取下一数据。

[0055] 或者，第一服务器节点 2’通过路由交换装置 1 向第二服务器节点 2’’写入数据的过程包括以下步骤，如图 7 所示。71，DMA 模块 11 通过第一协议转换接口 12’ 向第一服务器节点 2’ 发出读请求。72，第一服务器节点 2’ 依据该读请求通过第一服务器节点 2’ 的内部总线将写入数据传送到第一协议转换接口 12’。73，DMA 控制器 111 控制 DMA 通道 112 将所述数据传送到第二协议转换接口 12’’。然后，74，DMA 模块 11 通过第二协议转换接口 12’’ 向第二服务器节点 2’ 发出写请求。75，第二服务器节点 2’ 依据该写请求写入所述数据。最后，76，判断是否全部所需数据均写入完毕，如果不是，则获取下一数据。

[0056] 由上可知，第二服务器节点 2’’通过路由交换装置 1 从第一服务器节点 2’读出数据的过程或者第二服务器节点 2’’ 通过路由交换装置 1 向第一服务器节点 2’ 写入数据的过程与上述描述的具体过程相似。类似地，当网络交换系统中有更多的路由交换装置以及网络节点时，它们之间的数据访问过程与上述过程也是基本相同的。从而，本发明实施例的在网络交换系统中的路由交换方法由于将路由交换装置替代了现有技术的网络交换机，使跨节点的内存访问和 IO 空间访问可以不通过代理而直接进行，进而减少跨节点的内存访问和 IO 空间访问的延时，提高系统的整体性能。

[0057] 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发明的范围。

[0058] 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、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0059] 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。

[0060] 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。

[0061] 另外，在本发明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。

[0062] 以上所述，仅为本发明的具体实施方式，但本发明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发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应所述以权利要求的保护范围为准。

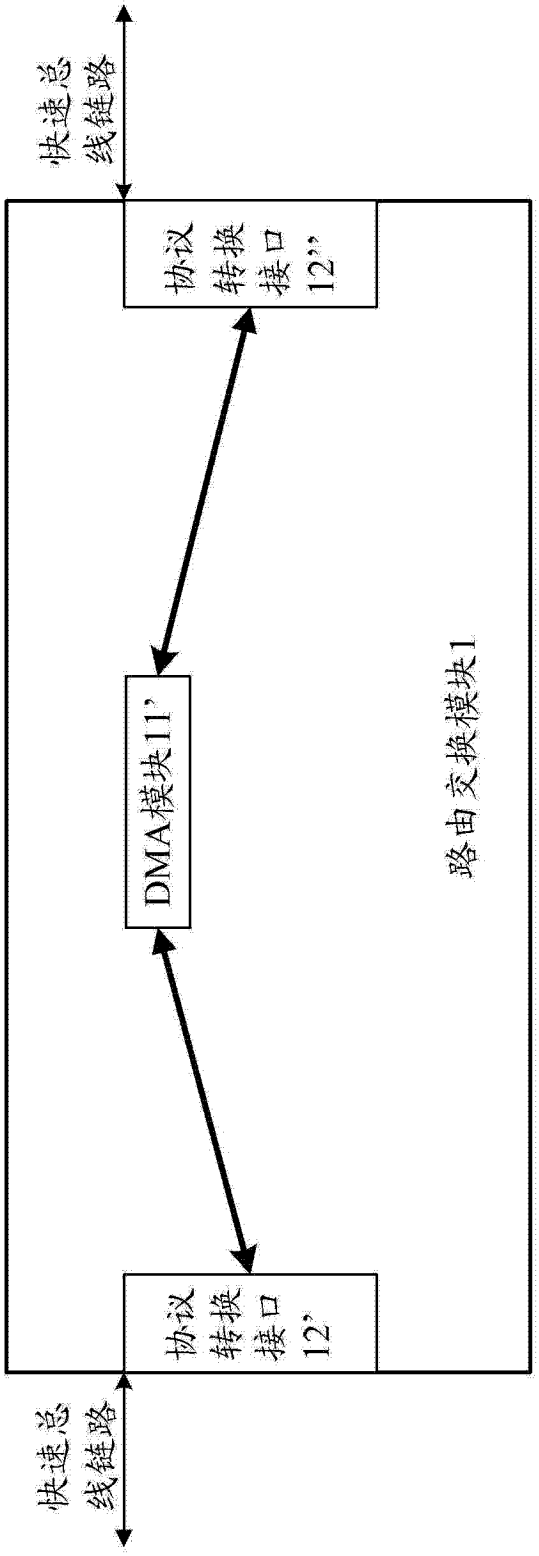


图 1

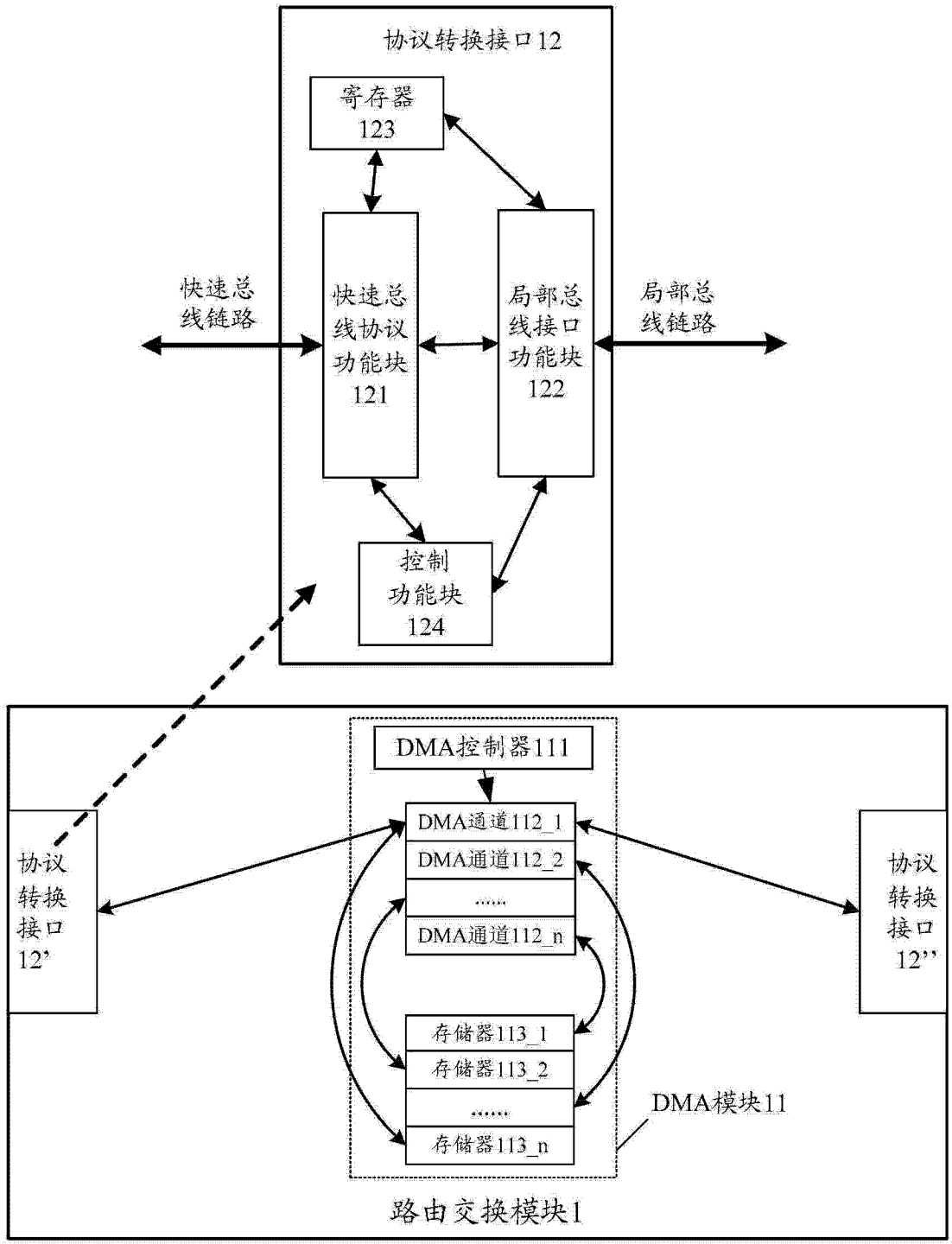

协议转换接口12
寄存器
123
快速总线链路
快速总线协议功能块
121
局部总线接口功能块
122
局部总线链路
控制功能块
124
DMA控制器111
DMA通道112_1
DMA通道112_2
.....
DMA通道112_n
协议转换接口12'
协议转换接口12''
存储器113_1
存储器113_2
.....
存储器113_n
DMA模块11
路由交换模块1

图 2

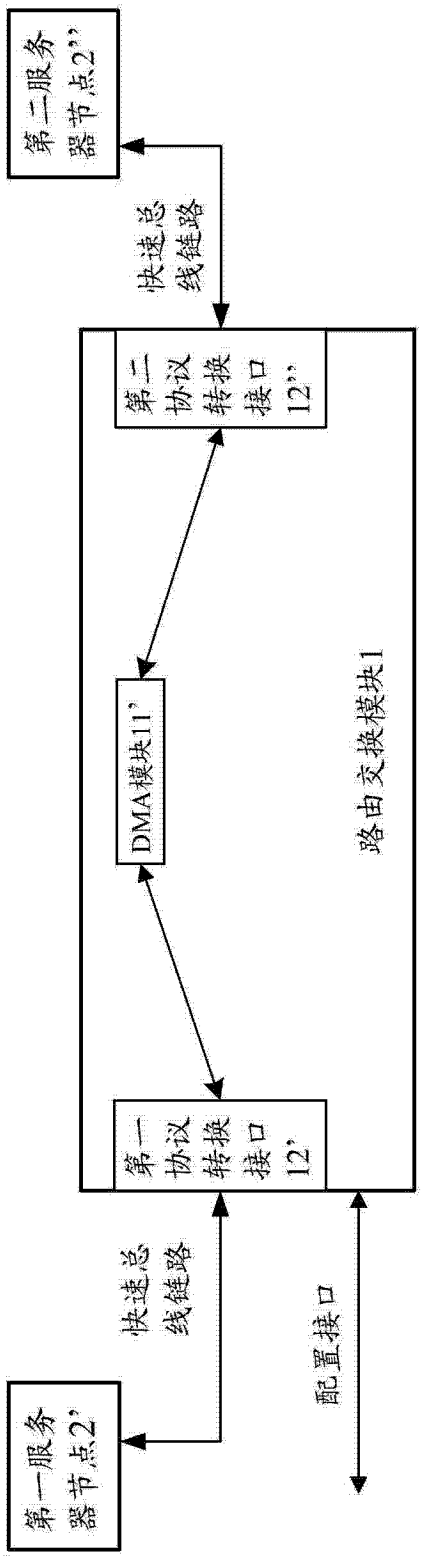

第二服务器节点2''
快速总线链路
第二协议转换接口12''
DMA模块11'
路由交换模块1
第一协议转换接口12'
快速总线链路
配置接口
第一服务器节点2'

图 3

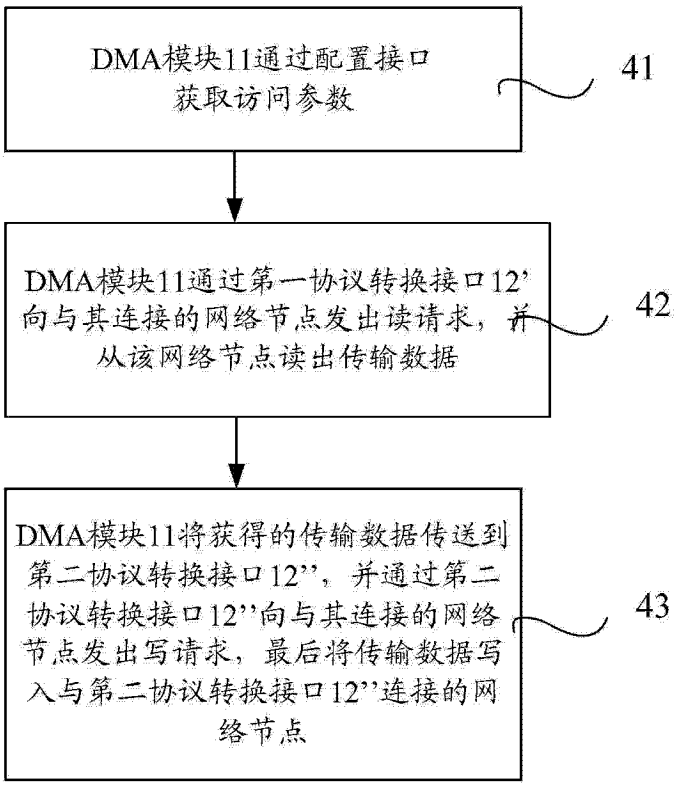

DMA模块11通过配置接口获取访问参数
41
DMA模块11通过第一协议转换接口12'向与其连接的网络节点发出读请求，并从该网络节点读出传输数据
42
DMA模块11将获得的传输数据传送到第二协议转换接口12''，并通过第二协议转换接口12''向与其连接的网络节点发出写请求，最后将传输数据写入与第二协议转换接口12''连接的网络节点
43

图 4

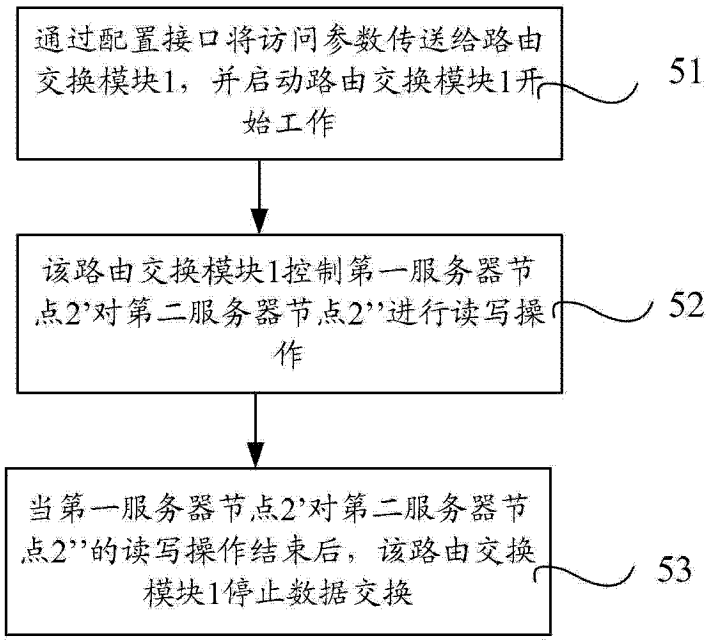

通过配置接口将访问参数传送给路由交换模块1，并启动路由交换模块1开始工作
51
该路由交换模块1控制第一服务器节点2'对第二服务器节点2''进行读写操作
52
当第一服务器节点2'对第二服务器节点2''的读写操作结束后，该路由交换模块1停止数据交换
53

图 5

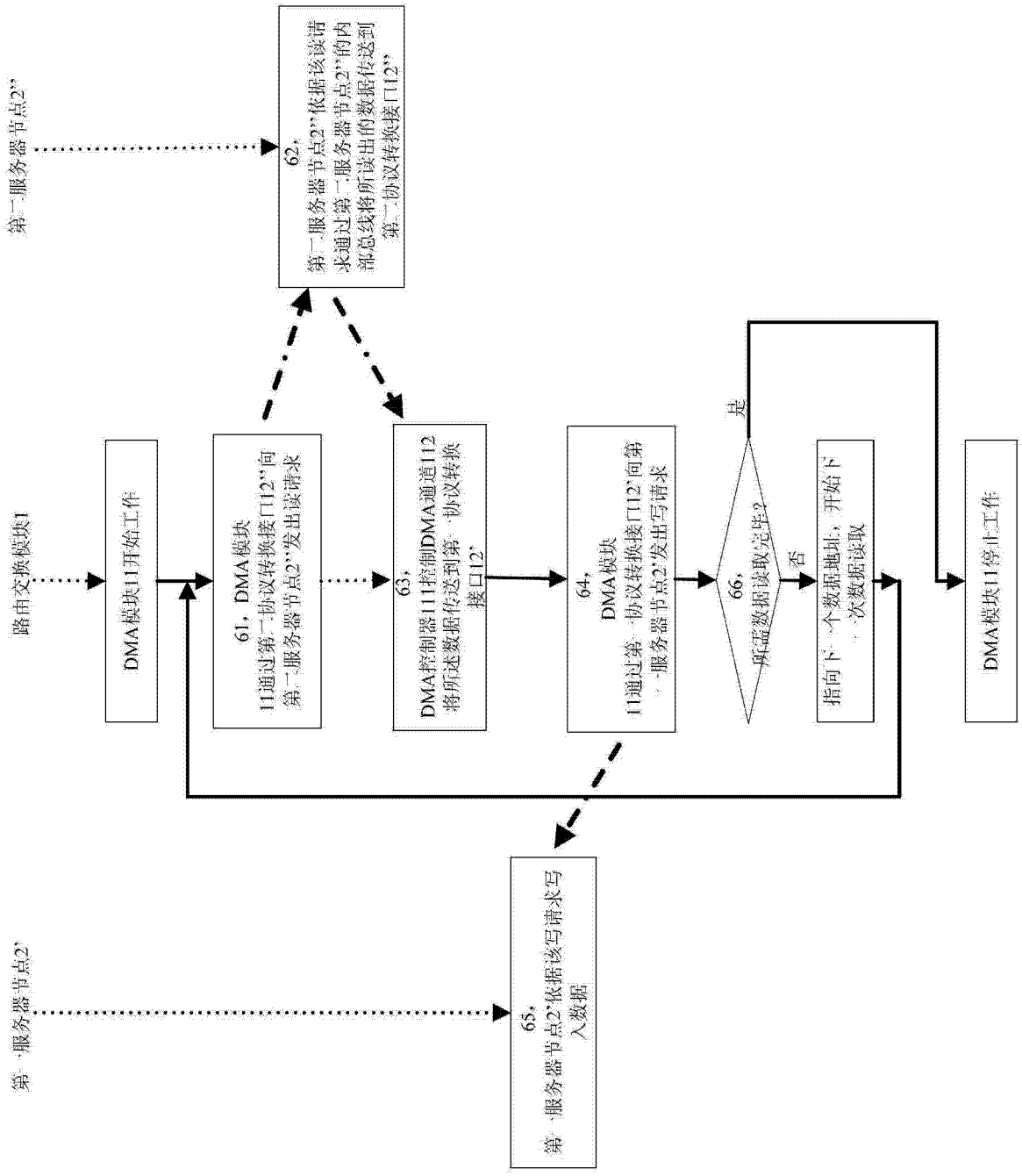

第二服务器节点2''
62，
第二服务器节点2''依据该读请求通过第二服务器节点2''的内部总线将所读出的数据传送到第二协议转换接口12''
路由交换模块1
DMA模块11开始工作
61，DMA模块11通过第二协议转换接口12''向第二服务器节点2''发出读请求
63，
DMA控制器111控制DMA通道112将所述数据传送到第一协议转换接口12'
64，
DMA模块11通过第一协议转换接口12'向第一服务器节点2'发出写请求
66，
所需数据读取完毕?
是
否
指向下一个数据地址，开始下一次数据读取
DMA模块11停止工作
第一服务器节点2'
65，
第一服务器节点2'依据该写请求写入数据

图 6

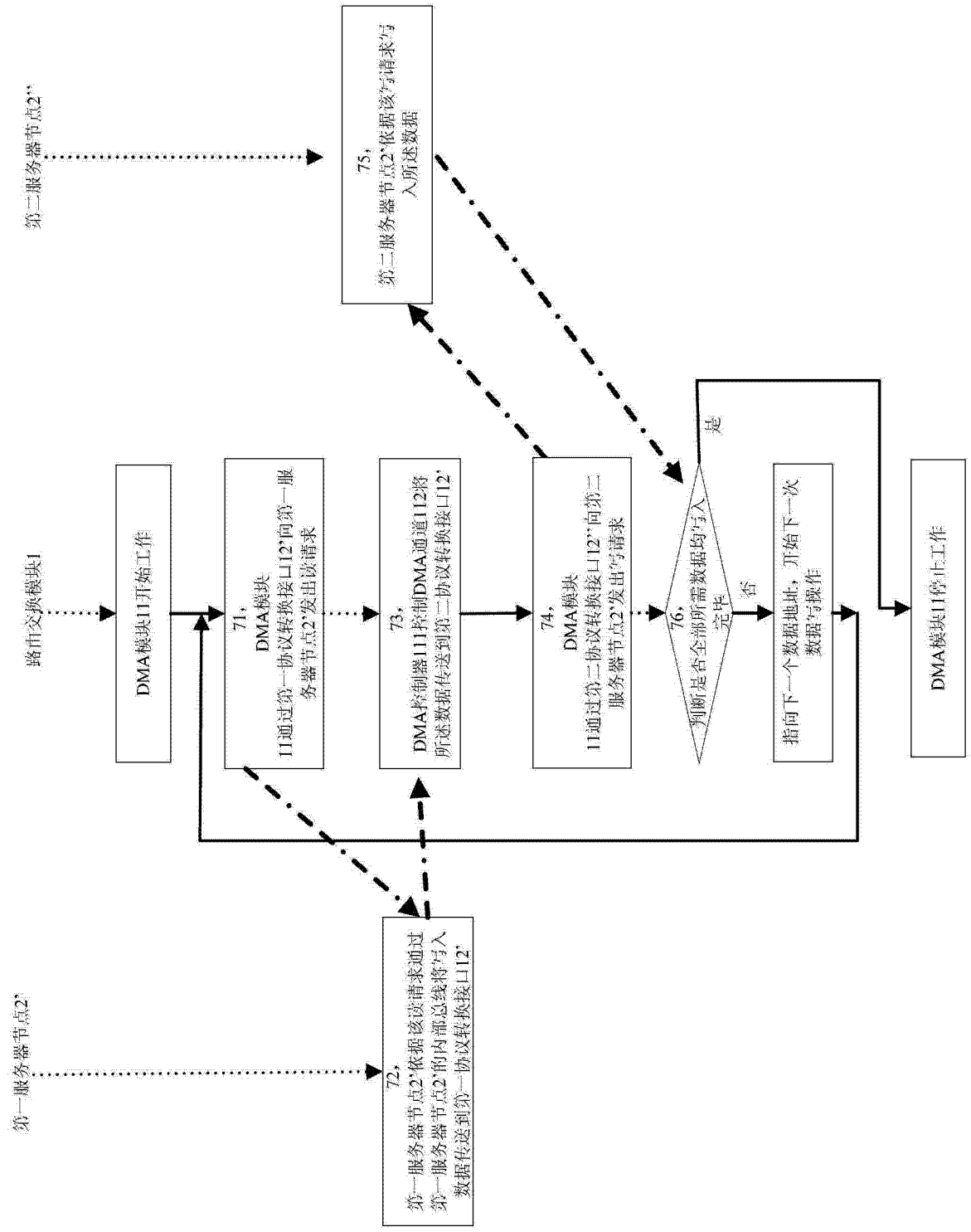

第二服务器节点2''
75,
第二服务器节点2''依据该写请求写入所述数据
路由交换模块1
DMA模块11开始工作
71,
DMA模块
11通过第一协议转换接口12'向第一服务器节点2'发出读请求
73,
DMA控制器111控制DMA通道112将所述数据传送到第二协议转换接口12'
74,
DMA模块
11通过第二协议转换接口12''向第二服务器节点2''发出写请求
76,
判断是否全部所需数据均写入完毕
是
否
指向下一个数据地址，开始下一次数据写操作
DMA模块11停止工作
第一服务器节点2'
72,
第一服务器节点2'依据该读请求通过第一服务器节点2'的内部总线将写入数据传送到第一协议转换接口12'

图 7