[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 13/00 (2006.01)
H04L 12/24 (2006.01)

[12] 发明专利说明书

专利号 ZL 200410074690.1

[45] 授权公告日 2007 年 5 月 16 日

[11] 授权公告号 CN 1316385C

[22] 申请日 2004.9.13
[21] 申请号 200410074690.1
[30] 优先权
[32] 2003.12.2 [33] US [31] 10/725,778
[73] 专利权人 国际商业机器公司
地址 美国纽约
[72] 发明人 G·R·弗拉齐耶
[56] 参考文献
US2003/0031183 A1 2003.2.13
US5991797A 1999.11.23
US6594712B1 2003.7.15
US2003/0079075A1 2003.4.24
审查员 唐 嫣

[74] 专利代理机构 北京市中咨律师事务所
代理人 于 静 李 峥

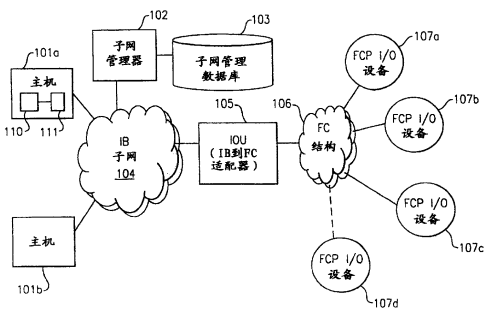
权利要求书 3 页 说明书 8 页 附图 3 页

[54] 发明名称

在 Infiniband 管理数据库上存储光纤信道信息的方法和装置

[57] 摘要

一种在 Infiniband 管理数据库上存储光纤信道信息的技术。主机计算系统，能够有效地识别出相应于光纤信道 I/O 设备的 Infiniband™ (IB) 寻址参数，所述的光纤信道 I/O 设备可以通过 IB 到光纤信道适配器访问。说明了一种有效的方法在子网管理数据库内存储与单个光纤信道 I/O 设备有关的 IB 寻址参数，并且说明了用于从该数据库中恢复 IB 寻址参数的有效机制，所述的 IB 寻址参数相应于访问所希望的光纤信道 I/O 设备可以经由的所有的物理路径。



1. 一种用于在网络中识别 I/O 设备的方法，包括：

在数据库中的服务记录内注册相应于 I/O 设备的服务名称；

在该服务名称上添加将该服务名称标识为特定 I/O 设备的名称的唯一的后缀；以及

通过在所述的数据库中查找该注册的服务名称和添加的后缀访问所述的 I/O 设备。

2. 如权利要求 1 的方法，还包括在该服务记录的其它字段内存储与所述 I/O 设备有关的其它寻址参数。

3. 如权利要求 2 的方法，还包括从主机向所述数据库的子网管理器发送一个请求，所述的请求包括能够使得向该主机返回与一个 I/O 设备有关的所有服务记录的参数，从而该主机可以访问该 I/O 设备。

4. 如权利要求 3 的方法，还包括通过检查对所述数据库的子网管理器的单个请求的响应，为网络中的主机识别到所述 I/O 设备的所有物理路径。

5. 如权利要求 4 的方法，其中所述的网络包括提供了到该 I/O 设备的访问的 IOU 适配器，并且识别该 I/O 设备包括识别该 IOU 适配器，从而对该 I/O 设备的访问不需要轮询所述的 IOU 适配器。

6. 如权利要求 5 的方法，其中寻址参数包括对该 I/O 设备不可访问的指示，从而不需要轮询 IOU 适配器就可以由对该 I/O 设备的寻址确定该 I/O 设备不可访问。

7. 如权利要求 2 的方法，其中该网络是 Infiniband 网络。

8. 如权利要求 7 的方法，其中所述的其它寻址参数包括与该 I/O 设备有关的 IOCGUID。

9. 如权利要求 8 的方法，其中该 I/O 设备是 FCP I/O 设备，

并且所述其它的寻址参数包括对提供到所述 FCP I/O 设备的访问的相应 SRP I/O 设备的标识。

10. 一种用于识别网络中的 I/O 设备的装置，包括：

数据库，具有用于存储相应于 I/O 设备的服务名称的服务记录；

用于将唯一的后缀添加在服务名称上，将服务名称标识为特定 I/O 设备的名称的设备；以及

主机，其通过在数据库中查找注册的服务名称和附加的后缀访问所述的 I/O 设备。

11. 如权利要求 10 的装置，还包括存储在服务记录的其它字段内的与所述 I/O 设备有关的其它寻址参数。

12. 如权利要求 11 的装置，还包括所述主机内的请求产生器，它向管理所述数据库的子网管理器发送请求，所述的请求包括有能够使得向所述主机返回与一个 I/O 设备有关的所有服务记录、从而主机可以访问该 I/O 设备的参数。

13. 如权利要求 12 的装置，其中所述的主机通过检查对由该主机向所述数据库的子网管理器发送的单个请求的响应，识别出从所述的主机到所述的 I/O 设备的所有物理路径。

14. 如权利要求 13 的装置，其中所述的网络包括 IOU 适配器，它提供了到该 I/O 设备的访问，并且其中所述寻址参数包括对该 IOU 适配器的标识，从而对该 I/O 设备的访问不需要轮询所述的 IOU 适配器。

15. 如权利要求 14 的装置，其中所述寻址参数包括该 I/O 设备不可访问的指示，从而不需要轮询该 IOU 适配器就可以由对该 I/O 设备的寻址确定该 I/O 设备不可访问。

16. 如权利要求 11 的装置，其中该网络是 Infiniband 网络。

17. 如权利要求 16 的装置，其中所述的其它寻址参数包括与该 I/O 设备有关的 IOCGUID。

18. 如权利要求 17 的装置，其中 I/O 设备是 FCP I/O 设备，并且所述的其它寻址参数包括对提供了到所述 FCP I/O 设备的访问的相应的 SRP I/O 设备的标识。

在 Infiniband 管理数据库上存储光纤信道信息的方法和装置

技术领域

本发明涉及 Infiniband™（IB）输入/输出单元（IOU），它能够使主机计算系统访问符合小型计算机系统接口（SCSI）到光纤信道（FC）的映射的输入/输出（I/O）设备。这种 I/O 设备被称为 FCP I/O 设备。

背景技术

下面的参考文献与本发明有关:

1. 在 Infiniband Architecture Specification, Volume1, release 1.1, Infiniband Trade Association 的第 14、15、16 章中可以找到关于 IB 子网管理的信息。Infiniband 是作为 Portland Oregon 的 Infiniband Trade Association 从事商业活动的 SYSTEM I/O 的商标。
2. 在 Fibre Channel-Framing and Signalling(FC-FS) rev 1.9, American National Standards Institute, Inc.中可以找到关于光纤信道的信息。
3. 在 Fibre Channel Protocol for SCSI, second version, (FCP-2), American National Standards Institute, Inc.中可以找到关于 FCP I/O 的信息。
4. 在 SCSI Remote Direct Memory Access(RDMA) Protocol(SRP), rev 16a, American National Standards Institute, Inc.中可以找到关于 SCSI 映射到 IB 的信息。
5. 在 Fibre Channel HBA API(FC-HBA), rev 8, American

National Standards Institute, Inc 中可以找到关于 SRP 目的端口标识符的信息。

在包含有能够使连接到 Infiniband 网络上的主机计算系统可以访问 I/O 设备的 I/O 单元（IOU）的 Infiniband 配置中，对于主机而言，FCP I/O 设备是以 SRP I/O 设备出现的。由于所有的 SRP I/O 设备都由 SRP 目的端口标识符（ID）唯一地标识，因此在每个 FCP I/O 设备的全球范围唯一端口名称（WWPN）和相应的 SRP 目的端口标识符之间存在着一一对应的关系。SRP 目的端口标识符由两个 64 位组成。第一个是“IOCGUID”，它标识着控制着该 SRP 目的端口的 IOC。第二个是一个 64 位的扩展。对于 IB 到 FC 适配器的情况，所述的扩展被设置为与该 SRP 目的端口标识符相应的 FCP I/O 设备的 WWPN。例如，SRP 目的端口标识符“IOCGUID1.WWPNA”相应于 WWPN 为 WWPNA 的 FCP I/O 设备，该设备可以通过 IOCGUID 为 IOCGUID1 的 IOC 访问；SRP 目的端口标识符“IOCGUID1.WWPNB”相应于 WWPN 为 WWPNB 的 FCP I/O 设备，该设备也可以通过 IOCGUID 为 IOCGUID1 的 IOC 访问；SRP 目的端口标识符“IOCGUID2.WWPNC”相应于 WWPN 为 WWPNC 的 FCP I/O 设备，该设备可以通过 IOCGUID 为 IOCGUID2 的 IOC 访问；并且依此类推。

除了目的端口标识符之外，还以 IB“服务名称”标识 SRP I/O 设备。IB 服务名称唯一地标识“服务提供者”，诸如信息数据库、通信功能、或（如本例中的）SRP I/O 设备功能。SRP I/O 设备的服务名称是 40 个字节的 UTF－8 字符串。服务名称的前 24 个字节包括如下的字符串：

‘SRP.T10:xxxxxxxxxxxxxxxx’

服务名称中的‘SRP.T10’部分将该服务标识为一个 SRP 目的端口；服务名称中的‘xxxxxxxxxxxxxxxx’部分是 SRP 目的端口标识符的 64 位的“扩展”部分的 16 个字节的十六进制编码。注意在 IB 到 FC 适配器的情况中，SRP 目的端口标识符的 64 位扩展被设置为与该 SRP 目的端口相

应的 FCP I/O 设备的 WWPN；因此，可以通过检查服务名称的‘xxxxxxxxxxxxxxxx’部分推导出与特定服务名称相应的 FCP I/O 设备。SRP 服务名称的其它 16 字节包含空字符。与具有 WWPN x ‘5347 9899 5348 8888’ 的 FCP I/O 设备相应的 SRP 服务名称的一个例子为 ‘SRP.T10:5347989953488888’，其后为空字符。

因为上述的服务名称和 SRP 目的端口标识符格式中包括相应的 FCP I/O 设备的 FC WWPN，主机能够为与 FCP I/O 设备相应的 SRP I/O 设备产生 SRP 服务名称。因为主机通常通过 FCP I/O 设备的 WWPN 识别 FCP I/O 设备，这使得每当主机需要访问具有给定 WWPN 的特定 FCP I/O 设备时，可以容易地确定相应的 SRP I/O 设备的 SRP 服务名称。

然而为了访问 SRP I/O 设备，主机还需要支持该服务的 IB 到 FC 适配器的 IB 地址。在本发明之前，不存在有效的方法可以使得主机能够确定支持与该 FCP I/O 设备相应的服务名称的 IB 到 FC 适配器的 IB 地址。因此，如果不使用本发明，试图访问具有特定 WWPN 的 FCP I/O 设备的主机将需要轮询 IB 子网内的所有 I/O 单元（IOU），以便确定那些是 IB 到 FC 适配器的 IOU。然后主机将轮询每个 IB 到 FC 适配器以确定特定的一个或多个 IB 到 FC 适配器，它（它们）提供了对具有所希望的 WWPN 的 FCP I/O 设备的访问。当 IB 子网为任何显著大小时，这种涉及到向子网中的每个 IOU 发送多个查询的轮询操作是不实际的，并且产生不可接受的性能下降。

发明内容

本发明定义了用于有效地在 IB 子网管理数据库内存储并检索关于 FCP I/O 设备的信息的方法。所述存储信息的方法能够使主机快速地确定 IB 寻址参数，通过所述的参数可以通过 IB 到 FC 适配器访问 FCP I/O 设备。

通过如本发明所述在 IB 子网管理数据库上注册 SRP 服务名称并且通过查询该数据库，克服了现有技术的缺点，即，缺少有效的方法发现 IB

到 FC 适配器的 IB 地址，所述的 IB 到 FC 适配器提供了对由 WWPN 标识的特定 FCP I/O 设备的访问。使用本发明降低了主机系统的复杂度，提高了主机系统的性能，并且提供了下面总结出的其它优点。

通过本发明的技术实现了其它的特征和优点。在此详细地说明了本发明的其它实施例和各个方面，并且认为它们是所要求保护的发明的一部分。为了更好地理解本发明以其优点和特征，请参考该说明和附图。

附图说明

通过下面结合附图进行的详细说明可以清楚地了解本发明的上述以及其它目的、特征和优点，其中：

图 1 示出了一个包含本发明的网络配置的例子，它包括 IB 子网、FC 结构（fabric）以及 IB 到 FC 适配器；

图 2 示出了包含本发明的 IB 到 FC 适配器以及相关的组件的一个例子；

图 3 示出了 IB 服务记录的一个例子；并且

图 4 示出了确定提供对特定 FCP I/O 设备的访问的一个或多个 IB 到 FC 适配器的 IB 地址的过程的流程图。

下面的详细说明以示例并参考附图的方式解释了本发明的优选实施例以及本发明的优点和特征。

具体实施方式

本发明为主机计算系统提供了有效地识别 Infiniband（IB）寻址参数的能力，所述的寻址参数相应于可以通过 IB 到光纤信道适配器访问的光纤信道 I/O 设备。本发明可以分为配置步骤和查询步骤。在配置步骤中，在一个数据库中注册 FCP I/O 设备。在查询步骤中，主机访问该数据库以便确定提供到该 FCP I/O 设备的访问的 IB 到 FC 适配器的 IB 寻址参数。下面将详细地说明这两个步骤。

参考图 1，主机系统 101a 和 101b 通过 IB 子网 104 与 IB 到 FC 适配

器 105 通信。IB 到 FC 适配器 105 通过 FC 结构 106 与 FCP I/O 设备 107a－107d 通信。每个主机系统，例如主机系统 101a 包括一个用于执行数据处理指令的处理单元 110，以及用于存储将要处理的数据和形成命令、请求以及例程的布置在计算机程序中的编码的处理指令的存储器 111。主机系统 101 和 IB 到 FC 适配器 105 间的接口协议符合参考文献 4。在主机 101 看来，每个 FCP I/O 设备 107a－107d 都是一个符合 SRP（参考文献 4）的 IB I/O 设备；因此，对主机 101 为可见的 I/O 设备 107 被称为 SRP I/O 设备，即使它们代表实际的 FCP I/O 设备。IB 到 FC 适配器 105 和 FCP I/O 设备 107 之间的接口协议符合 SCSI 到 FC 的映射（参考文献 3）。

图 2 给出了 IB 到 FC 适配器 105 的一个展开的表示，它包括两个 I/O 控制器（IOC）201a 和 201b。IOC 201 附加在 FC 结构 106 上，FC 结构 106 附加到 FCP I/O 设备 107a－107d。IOC 201a 提供到 FCP I/O 设备 107a 和 107b 的访问，并且 IOC 201b 提供到 FCP I/O 设备 107c 和 107d 的访问。所有的 FCP I/O 设备 107 由 64 位的“全球唯一的”端口名称（WWPN）唯一地标识。因此，FCP I/O 设备 107a－107d 由 WWPN A 到 WWPN D 唯一地标识。

回到图 1，在配置步骤中，在子网管理（SA）数据库 103 中注册每个 FCP I/O 设备 107a－107d。这种注册可以由 IB 到 FC 适配器 105 执行，或可以由第三方，例如执行配置例程的另一主机执行。

为了在 SA 数据库 103 中注册 FCP I/O 设备 107，需要在 SA 数据库 103 内存储标识着 FCP I/O 设备 107 的 IB 服务记录。这是通过向子网管理器 102 发送包含 SubnetAdminSet（ServiceRecord）命令的 IB 数据报而完成的。这个命令包括一个如图 3 中所示的 IB 服务记录。该服务记录存储在子网管理数据库 103 内。参考图 3，IB 服务记录中与本发明有关的字段为 SeviceName 301 和 ServiceData 302。（如 IB 规范中所说明的那样使用 ServiceID、ServiceGID、ServiceP_Key、ServiceLease 和 ServiceKey 字段，参考文献 1）

512 位（64 字节）的 SeviceName 字段 301 的前 24 字节被设置为相应

于该 FCP I/O 设备的 SRP 服务名称的前 24 个字节。这些字节后面跟着字符串 '.FCP'，然后接着是一系列的空字符以填充 SeviceName 字段中剩余的字节。字符串 '.FCP' 被附加在 SRP 服务名称上以便将其与用于不相应于一个 FCP I/O 设备的 SRP I/O 设备的 SRP 服务名称区分开。这样的服务名称将在相应的字符位置上包含空字符。

ServiceData 字段 302 的前 64 位被设置为提供到该 FCP I/O 设备的访问的 IOC 的 IOCGUID。ServiceData 字段中剩余的字节没有被本发明使用，因而可以被设置为任意值。这完成了为 FCP I/O 设备进行配置的步骤；为所有可以从 IB 子网访问的 FCP I/O 设备重复该配置步骤。

假定已经完成了上面的配置步骤，则通过执行图 4 中给出的步骤，现在主机 101 可以确定提供到具有给定 WWPN 的 FCP I/O 设备的访问的 IB 到 FC 适配器 105 的 IB 地址。

参考图 4，主机向子网管理器 102（SM）发送一个 SubnetAdminGet（ServiceRecord）请求（401）。该请求指示 SM 102 返回所有包括与在配置步骤中为该 FCP I/O 设备 107 注册的服务名称相应的服务名称的服务记录。例如，为了获得相应于具有 WWPN 为 x '5347 9899 5348 8888' 的 FCP I/O 设备的服务记录，主机请求所有包含 ServiceName 字段为 'SRP.T10:5347989953488888.FCP'、其后为空字符的服务记录。在参考文献 1 的 IB 规范中说明了发送这个请求的过程。

如果 IB 子网 104 中的 IB 到 FC 适配器 105 提供了到由 SubnetAdminGet（ServiceRecord）请求所标识的 FCP I/O 设备 107 的访问，则该响应包含至少一个相应于该 FCP I/O 设备的服务记录；如果该响应不包括任何服务记录，则该设备是不可访问的（402），并且过程终止。

假设响应包括至少一个服务记录，那么对于每个服务记录，主机 101 如下所述确定访问 FCP I/O 设备 107 所必须的 IB 寻址参数：

1. 在步骤 403，主机 101 通过将 ServiceGID 字段转换为 IB "路径" 来确定 IB 到 FC 适配器 105 的 IB 寻址参数。IB 规范中说明了执行该任务的过程。

2. 在步骤 404，主机 101 通过从 ServiceData 字段提取 64 位的 IOCGUID 来确定提供对 FCP I/O 设备 107 的访问的 IOC 201。注意，IOCGUID 是在 FCP I/O 设备 107 的配置步骤中被存储在 ServiceData 字段 302 中的。

3. 在步骤 405，主机 101 通过拼接 IOCGUID 和 I/O 设备 107 的 WWPN 构造相应于 FCP I/O 设备 107 的 SRP 目的端口标识符。即，SRP 目的端口标识符被设置为 IOCGUID.WWPN。

4. 在步骤 406，主机 101 通过执行 SRP 规范（参考文献 4）中给出的过程访问 I/O 设备 107。

对 SubAdminGet（ServiceRecord）请求的响应中所返回的每个服务记录重复上面的步骤。对每个服务记录执行上述步骤序列会导致识别通过 IB 子网 104 和 FC 结构 106 到 FCP I/O 设备 107 的所有物理路径。所述物理路径可能包括通过不同的 IB 到 FC 适配器、每个适配器中不同的 IOC、FC 结构上不同的端口以及通过该结构的不同的路由的访问。

注意，上述的过程不需要主机在访问 I/O 设备之前轮询子网中的多个 IOU，而这是本发明之前的技术所需要的。此外，主机不需要轮询 IOU 中的所有 IOC 以便确定支持与 FCP I/O 设备相应的服务名称的 IOC。在本发明之前，主机需要在每个 IOU 中轮询高达 256 的 IOC 以便确定所希望的 IOC。不是执行上述所有的轮询操作，一个不可接受的长时间的过程，而是主机可以通过向子网管理数据库发送单个请求来确定到 FCP I/O 设备的所有物理路径。

本发明的功能可以用软件、固件、硬件或它们的某些组合实现。

作为一个例子，本发明的一个或多个方面可以被包括在一个具有例如计算机可用介质的制品（例如，一个或多个计算机程序产品）中。所述的介质上嵌入有用于提供并有助于实现本发明的功能的计算机可读的程序代码手段。所述的制品可以被包括作为计算机系统的一部分或是单独出售。

此外，可以提供至少一个机器可读的程序存储装置，该装置有形地体现至少一个可以被该机器所执行以便执行本发明的功能的指令程序。

此处给出的流程图仅仅是例子。可以有描述于此的这些示意图或步骤（或操作）的许多变形而不脱离本发明的精神。例如，各个步骤可以不同的顺序执行，或者可以添加、删除或修改步骤。所有这些变形都被认为是所要求保护的发明的一部分。

虽然已经说明了本发明的优选实施例，本领域的技术人员应当理解，不论是现在还是未来，可以对本发明做出各种落在下面的权利要求之内的改进和增强。这些权利要求应当被解释为主张了对前面说明的发明的适当的保护。

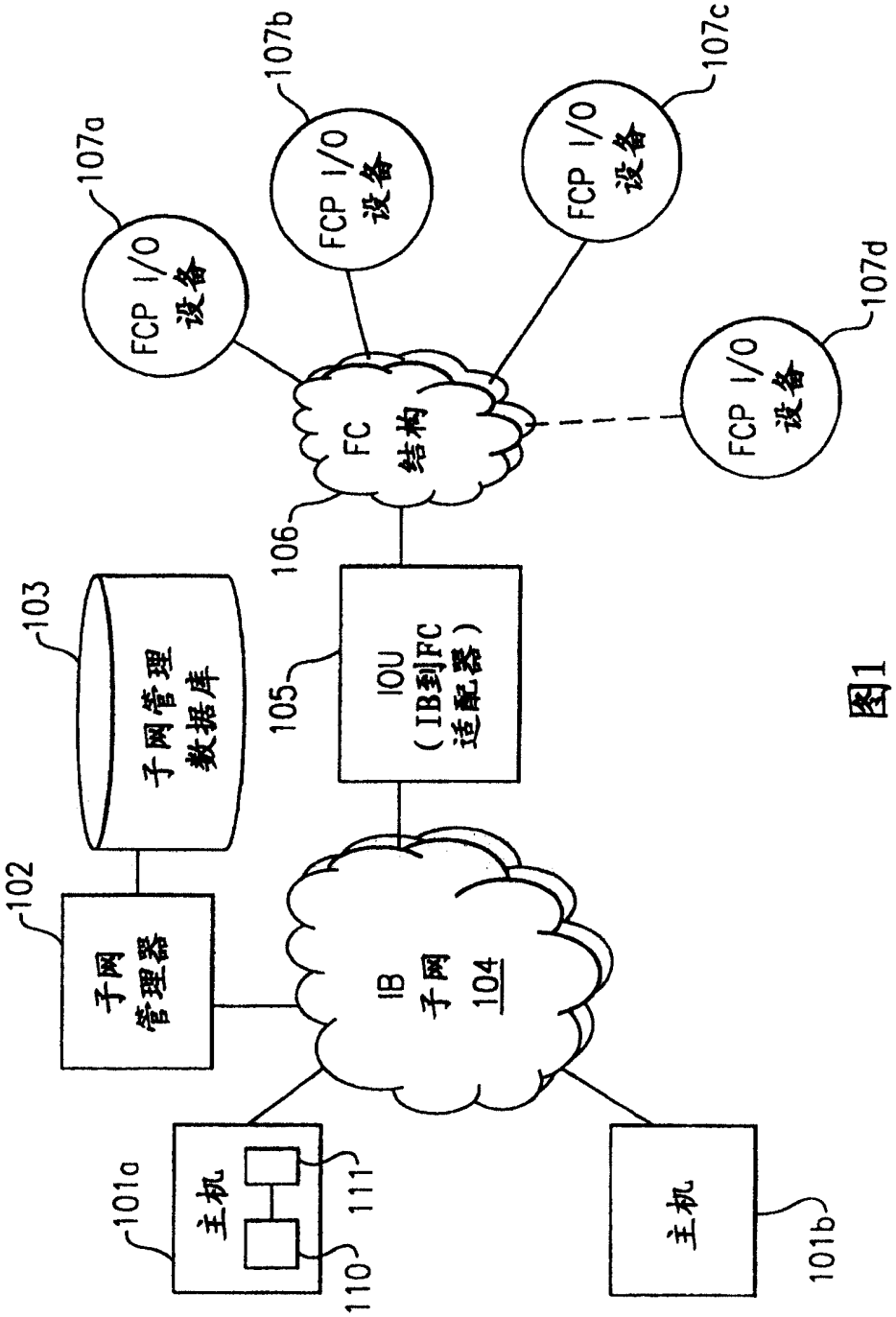

107a
FCP I/O 设备
107b
FCP I/O 设备
107c
FCP I/O 设备
107d
FCP I/O 设备
FC 结构
106
IOU (IB到FC 适配器)
105
103
子网管理 数据库
102
子网 管理器
IB 子网 104
101a
主机
111
110
主机
101b

图1

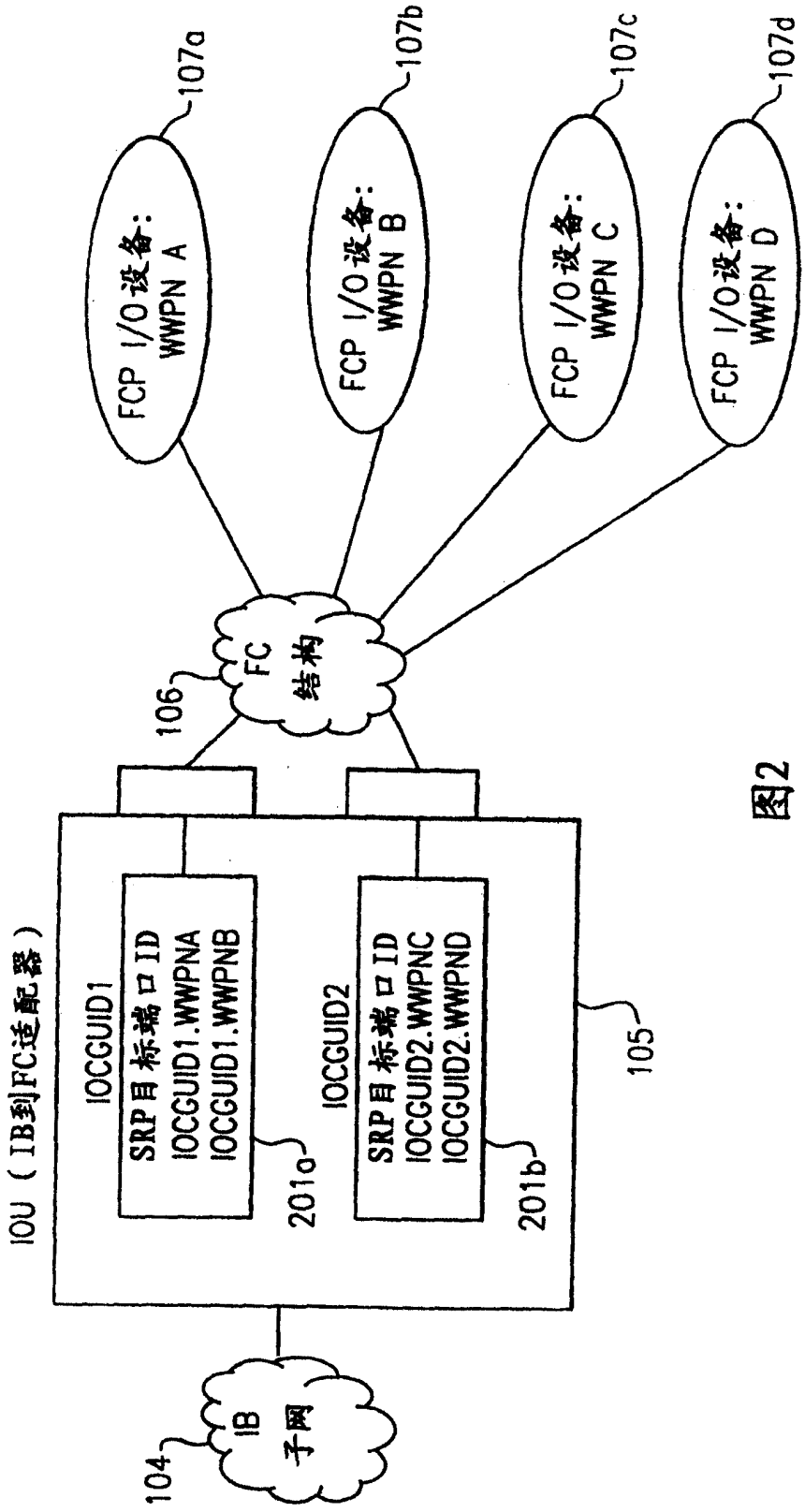

FCP I/O设备:
WWPN A
107a
FCP I/O设备:
WWPN B
107b
FCP I/O设备:
WWPN C
107c
FCP I/O设备:
WWPN D
107d
FC
结构
106
IOU（IB到FC适配器）
IOCGUID1
SRP目标端口ID
IOCGUID1.WWPNA
IOCGUID1.WWPNB
201a
IOCGUID2
SRP目标端口ID
IOCGUID2.WWPNC
IOCGUID2.WWPND
201b
105
IB
子网
104

图2

IB 服务记录

# 位数	值	
64	SERVICEID	
128	SERVICEGID	
16	SERVICEP_KEY	
16	RESERVED	
32	SERVICELEASE	
128	SERVICEKEY	
512	SERVICENAME	301
512	SERVICEDATA	302

图 3

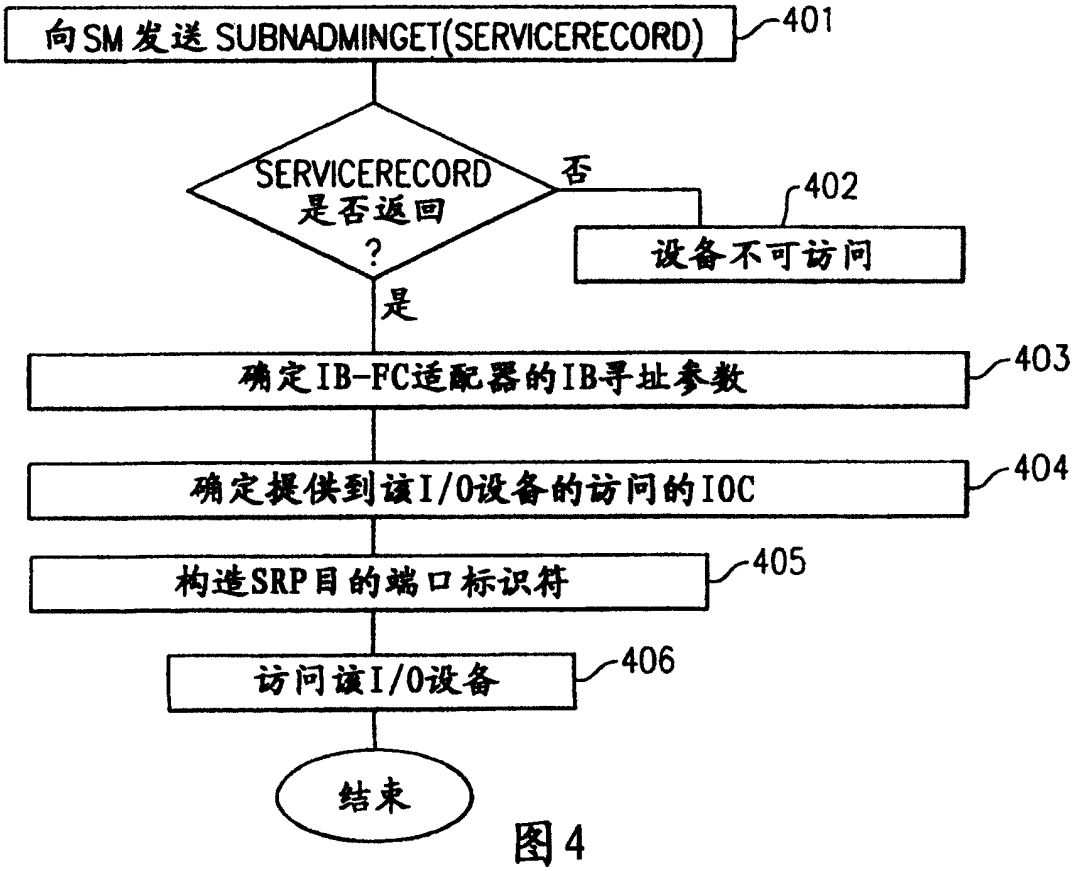


图 4