(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 103125097 B
(45)授权公告日 2016.11.16

(21)申请号 201180039807.0

(22)申请日 2011.09.16

(65)同一申请的已公布的文献号
申请公布号 CN 103125097 A

(43)申请公布日 2013.05.29

(30)优先权数据
61/384,228 2010.09.17 US
61/484,390 2011.05.10 US
61/493,330 2011.06.03 US
61/493,347 2011.06.03 US
61/498,329 2011.06.17 US

(85)PCT国际申请进入国家阶段日
2013.02.18

(86)PCT国际申请的申请数据
PCT/US2011/052021 2011.09.16

(87)PCT国际申请的公布数据
WO2012/037512 EN 2012.03.22

(73)专利权人 甲骨文国际公司
地址 美国加利福尼亚

(72)发明人 B-D·约翰逊 L·霍雷恩 R·安特森

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 李晓芳

(51)Int.Cl.
H04L 12/24(2006.01)
H04L 29/08(2006.01)

(56)对比文件
US 2005071382 A1,2005.03.31,
US 2005071382 A1,2005.03.31,
US 2002120720 A1,2002.08.29,
US 7409432 B1,2008.08.05,
CN 1567827 A,2005.01.19,

审查员 汪三骏

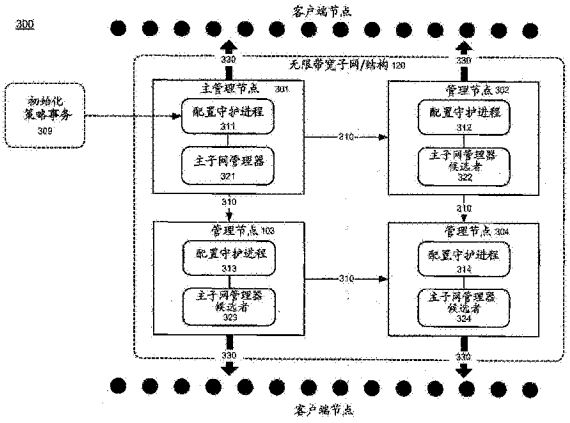
权利要求书5页 说明书9页 附图7页

(54)发明名称

在中间件机器中执行部分子网初始化

(57)摘要

一种系统和方法能够在中间件机器环境中执行子网初始化。根据一个实施例，中间件机器环境中的多个管理节点能够互连以使得与多个管理节点相关联的多个子网管理器能够使用带内通信协议彼此通信。多个子网管理器能够协商和推选主子网管理器，主子网管理器操作来配置和管理中间件机器环境。主子网管理器能够使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器。主子网管理器然后基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。



CN 103125097 B

1.一种用于在中间件机器中执行子网初始化的系统,包括:

一个或多个微处理器;

在中间件机器中在子网中互连的多个管理节点,其在所述一个或多个微处理器上运行,

与多个管理节点相关联的多个子网管理器,其中多个管理节点使得所述多个子网管理器能够使用带内通信协议彼此通信,其中所述多个子网管理器操作来彼此协商并且推选主子网管理器,以及

其中所述主子网管理器操作来:

使用所述带内通信协议将一个或多个子网初始化策略复制到中间件机器中的其它子网管理器,

使用配置之下的所述子网作为管理网络来将所述一个或多个子网初始化策略分布到多个客户端节点,以及

基于所述一个或多个子网初始化策略设置中间件机器中的所述多个客户端节点。

2.根据权利要求1所述的系统,其中:

所述多个管理节点是无限带宽(IB)子网的一部分。

3.根据权利要求1或2所述的系统,其中:

所述多个管理节点包括一个或多个网络交换机,其中每个所述子网管理器驻留在网络交换机上。

4.根据权利要求1或2所述的系统,还包括:

与所述多个管理节点连接的单独的存储系统。

5.根据权利要求1或2所述的系统,还包括:

与所述多个管理节点连接的多个主机服务器。

6.根据权利要求1或2所述的系统,还包括:

能够由客户机访问的一个或多个网关实例。

7.根据权利要求1或2所述的系统,还包括:

多个策略守护进程,其中每个所述策略守护进程与子网管理器相关联。

8.根据权利要求7所述的系统,其中:

与主子网管理器并置的所述策略守护进程负责使用一个或多个策略来配置和管理中间件机器。

9.根据权利要求1或2所述的系统,其中:

所述带内通信协议除了提供中间件机器中的以太网联网之外,还提供通信冗余。

10.根据权利要求1或2所述的系统,其中:

一个或多个划分策略能够定义中间件机器中的划分配置,并且其中所述一个或多个划分策略能够通过初始化策略事务被提供给中间件机器。

11.根据权利要求1或2所述的系统,其中:

当没有指定划分策略时,所述主子网管理器能够使用默认划分策略来初始化。

12.根据权利要求1或2所述的系统,其中:

所述主子网管理器操作来通过最初使用默认划分策略进行启动来同步结构级别的配置策略,并且在结构级别的同步已被实现时,接着在中间件机器中安装划分配置。

13.根据权利要求1或2所述的系统，其中：

所述主子网管理器操作来对于已经操作的节点保持当前划分成员资格不变，其中所述主子网管理器保证对于新的管理节点设置预定义的划分，使得所有管理节点经由通信协议通信和同步。

14.根据权利要求1或2所述的系统，其中：

子网管理器操作来在划分策略对于所述管理节点已被完全同步时，将所有末端节点完全初始化为具有划分策略中定义的完全划分成员资格。

15.一种用于在中间件机器中执行子网初始化的方法，包括：

在中间件机器中在子网中将运行在一个或多个微处理器上的多个管理节点互连，其中所述多个管理节点使得与多个管理节点相关联的多个子网管理器使用带内通信协议彼此通信；

在多个子网管理器之间协商并且推选主子网管理器，其中所述主子网管理器操作来配置和管理中间件机器；

经由所述主子网管理器、使用带内通信协议将一个或多个子网初始化策略复制到中间件机器中的其它子网管理器；

使用配置之下的所述子网作为管理网络来将一个或多个子网初始化策略分布到多个客户端节点；以及

基于所述一个或多个子网初始化策略设置中间件机器中的所述多个客户端节点。

16.根据权利要求15所述的方法，其中：

所述多个管理节点是无限带宽(IB)子网的一部分。

17.根据权利要求15或16所述的方法，其中：

所述多个管理节点包括一个或多个网络交换机，其中每个所述子网管理器驻留在网络交换机上。

18.根据权利要求15或16所述的方法，还包括：

提供与所述多个管理节点连接的单独的存储系统。

19.根据权利要求15或16所述的方法，还包括：

提供与所述多个管理节点连接的多个主机服务器。

20.根据权利要求15或16所述的方法，还包括：

提供能够由客户机访问的一个或多个网关实例。

21.根据权利要求15或16所述的方法，还包括：

提供多个策略守护进程，其中每个所述策略守护进程与子网管理器相关联。

22.根据权利要求21所述的方法，其中：

与主子网管理器并置的所述策略守护进程负责使用一个或多个策略来配置和管理中间件机器。

23.根据权利要求15或16所述的方法，还包括：

所述带内通信协议除了提供中间件机器中的以太网联网之外，还提供通信冗余。

24.根据权利要求15或16所述的方法，还包括：

提供定义中间件机器中的划分配置的一个或多个划分策略，并且其中所述一个或多个划分策略能够通过初始化策略事务被提供给中间件机器。

25.根据权利要求15或16所述的方法，还包括：

当没有指定划分策略时，所述主子网管理器使用默认划分策略来初始化。

26.根据权利要求15或16所述的方法，还包括：

所述主子网管理器操作来通过最初使用默认划分策略进行启动来同步结构级别的配置策略，并且在结构级别的同步已被实现时，接着在中间件机器中安装划分配置。

27.根据权利要求15或16所述的方法，还包括：

所述主子网管理器操作来对于已经可操作的节点保持当前划分成员资格不变，并且仅仅保证对于新的管理节点设置预定义的划分。

28.根据权利要求15或16所述的方法，还包括：

子网管理器操作来在划分策略对于所述管理节点已被完全同步时，将所有末端节点完全初始化为具有划分策略中定义的完全划分成员资格。

29.一种用于在中间件机器中执行子网初始化的系统，包括：

用于在中间件机器中在子网中将运行在一个或多个微处理器上的多个管理节点互连的装置，其中所述多个管理节点使得与所述多个管理节点相关联的多个子网管理器使用带内通信协议彼此通信；

用于在多个子网管理器之间协商并且推选主子网管理器的装置，其中所述主子网管理器操作来配置和管理中间件机器；

用于经由所述主子网管理器、使用带内通信协议将一个或多个子网初始化策略复制到中间件机器中的其它子网管理器的装置；

用于使用配置之下的所述子网作为管理网络来将一个或多个子网初始化策略分布到多个客户端节点的装置；以及

用于基于所述一个或多个子网初始化策略设置中间件机器中的所述多个客户端节点的装置。

30.根据权利要求29所述的系统，其中：

所述多个管理节点是无限带宽(IB)子网的一部分。

31.根据权利要求29或30所述的系统，其中：

所述多个管理节点包括一个或多个网络交换机，其中每个所述子网管理器驻留在网络交换机上。

32.根据权利要求29或30所述的系统，还包括：

用于提供与所述多个管理节点连接的单独的存储系统的装置。

33.一种中间件机器中的子网管理器，包括：

通信模块，被配置为使用带内通信协议与其它子网管理器通信；

推选模块，被配置为与其它子网管理器协商并且推选主子网管理器；

复制模块，被配置为，响应于所述子网管理器被推选为主子网管理器，使用带内通信协议将一个或多个子网初始化策略复制到中间件机器中的其它子网管理器；和

设置模块，被配置为，响应于所述子网管理器被推选为主子网管理器，基于一个或多个子网初始化策略设置中间件机器中的多个客户端节点，

其中所述主子网管理器操作来使用配置之下的所述子网作为管理网络来将一个或多个子网初始化策略分布到所述多个客户端节点。

34.根据权利要求33所述的子网管理器，其中：

所述带内通信协议除了提供中间件机器中的以太网联网之外，还提供通信冗余。

35.根据权利要求33所述的子网管理器，其中：

一个或多个划分策略能够定义中间件机器中的划分配置，并且其中所述一个或多个划分策略能够通过初始化策略事务被提供给中间件机器。

36.根据权利要求33所述的子网管理器，还包括：

初始化模块，被配置为，响应于所述子网管理器被推选为主子网管理器，在没有指定划分策略时，使用默认划分策略来初始化。

37.根据权利要求33所述的子网管理器，还包括：

同步模块，被配置为，响应于所述子网管理器被推选为主子网管理器，通过最初使用默认划分策略进行启动来同步结构级别的配置策略，并且在结构级别的同步已被实现时，接着在中间件机器中安装划分配置。

38.一种中间件机器中的管理节点，包括根据权利要求33-37中的任何一个所述的子网管理器。

39.一种用于在中间件机器中执行子网初始化的系统，包括在中间件机器中互连的根据权利要求38所述的多个管理节点。

40.根据权利要求39所述的系统，其中：

所述多个管理节点是无限带宽(IB)子网的一部分。

41.根据权利要求39所述的系统，其中：

所述多个管理节点包括一个或多个网络交换机，其中每个所述子网管理器驻留在网络交换机上。

42.根据权利要求39所述的系统，还包括：

与所述多个管理节点连接的单独的存储系统。

43.根据权利要求39所述的系统，还包括：

与所述多个管理节点连接的多个主机服务器。

44.根据权利要求39所述的系统，还包括：

能够由客户机访问的一个或多个网关实例。

45.根据权利要求39所述的系统，还包括：

多个策略守护进程，其中每个所述策略守护进程与子网管理器相关联。

46.根据权利要求45所述的系统，其中：

与主子网管理器并置的所述策略守护进程负责使用一个或多个策略来配置和管理中间件机器。

47.根据权利要求39所述的系统，其中：

所述主子网管理器操作来对于已经可操作的节点保持当前划分成员资格不变，并且仅仅保证对于新的管理节点设置预定义的划分。

48.根据权利要求39所述的系统，其中：

子网管理器操作来在划分策略对于所述管理节点已被完全同步时，将所有末端节点完全初始化为具有划分策略中定义的完全划分成员资格。

49.一种用于在中间件机器中执行子网初始化的系统，包括：

在中间件机器的子网中互连的多个管理节点，

与所述多个管理节点相关联的多个子网管理器，其中所述多个管理节点使得所述多个子网管理器能够使用带内通信协议彼此通信，其中所述多个子网管理器彼此协商并且推选一个子网管理器作为主子网管理器，以及

其中所述主子网管理器操作来

使用带内通信协议将一个或多个子网初始化策略复制到中间件机器中的其它子网管理器，

使用配置之下的所述子网作为管理网络来将一个或多个子网初始化策略分布到所述多个客户端节点，以及

基于所述一个或多个子网初始化策略设置中间件机器中的所述多个客户端节点。

在中间件机器中执行部分子网初始化

[0001] 版权通知

[0002] 本专利文件的公开的一部分包含受版权保护的材料。版权所有者不反对任何人对专利文件或专利公开的传真复制，因为它出现在专利商标局专利文档或记录中，但在别的方面保留所有任何版权。

技术领域

[0003] 本发明一般涉及计算机系统和诸如中间件之类的软件，并且特别涉及支持中间件机器环境。

背景技术

[0004] 无限带宽(Infiniband)(IB)架构是支持用于一个或多个计算机系统的I/O和处理器间通信二者的通信和管理基础设施。IB架构系统的规模可以从具有几个处理器和几个I/O装置的小服务器到具有数百处理器和数千I/O装置的大规模并行设施。

[0005] IB架构定义交换式通信结构，使得许多装置在受保护的、远程管理的环境中以高带宽和低延迟同时通信。端结点可以通过多个IB架构端口通信并且可以通过IB架构结构利用多个路径。提供通过网络的许多IB架构端口和路径以用于容错和增大的数据传送带宽二者。

[0006] 这些一般是本发明的实施例预期针对的领域。

发明内容

[0007] 这里描述的是用于在中间件机器环境中执行子网初始化的系统和方法。根据一个实施例，中间件机器环境中的多个管理节点能够互连以使得与多个管理节点相关联的多个子网管理器能够使用带内通信协议彼此通信。多个子网管理器能够协商和推选主子网管理器，主 子网管理器操作来配置和管理中间件机器环境。主子网管理器能够使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器。主子网管理器然后基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

[0008] 在一个方面，提供一种中间件机器环境中的子网管理器，包括：通信模块，被配置为使用带内通信协议与其它子网管理器通信；推选模块，被配置为与其它子网管理器协商并且推选主子网管理器；复制模块，被配置为，响应于子网管理器被推选为主子网管理器，使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器；和设置模块，被配置为，响应于子网管理器被推选为主子网管理器，基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

[0009] 在一些实施例中，所述带内通信协议除了提供中间件机器环境中的以太网联网之外，还提供通信冗余。

[0010] 在一些实施例中，一个或多个划分策略能够定义中间件机器环境中的划分配置，并且其中所述一个或多个划分策略能够通过初始化策略事务被提供给中间件机器环境。

[0011] 在一些实施例中，子网管理器还包括初始化模块，被配置为，响应于子网管理器被推选为主子网管理器，在没有指定划分策略时，使用默认划分策略来初始化。

[0012] 在一些实施例中，子网管理器还包括同步模块，被配置为，响应于子网管理器被推选为主子网管理器，在结构级别的同步已被实现时，通过最初使用默认划分策略起动，并且接着在中间件机器环境中安装划分配置，来同步结构级别的配置策略。

[0013] 在另一个方面中，提供一种中间件机器环境中的管理节点，包括上述子网管理器。

[0014] 在另一个方面中，提供一种用于在中间件机器环境中执行子网初始化的系统，包括在中间件机器环境中互连的多个上述管理节点。

[0015] 在一些实施例中，多个管理节点是无限带宽(IB)子网的一部分。

[0016] 在一些实施例中，多个管理节点包括一个或多个网络交换机，其中每个所述子网管理器驻留在网络交换机上。

[0017] 在一些实施例中，所述系统进一步包括与多个管理节点连接的单独的存储系统。

[0018] 在一些实施例中，所述系统进一步包括与多个管理节点连接的多个主机服务器。

[0019] 在一些实施例中，所述系统进一步包括能够由客户机访问的一个或多个网关实例。

[0020] 在一些实施例中，所述系统进一步包括多个策略守护进程，其中每个所述策略守护进程与子网管理器相关联。

[0021] 在一些实施例中，与主子网管理器并置的所述策略守护进程负责使用一个或多个策略配置和管理中间件机器环境。

[0022] 在一些实施例中，所述主子网管理器操作来对于已经操作的节点保持当前划分成员资格不变，并且仅仅保证对于新的管理节点设置预定义的划分。

[0023] 在一些实施例中，子网管理器操作来在划分策略已被对于所述管理节点完全同步时，利用划分策略中定义的完全划分成员资格完全初始化所有末端节点。

[0024] 在另一个方面，提供一种用于在中间件机器环境中执行子网初始化的系统，包括：多个管理节点，在中间件机器环境中互连；多个子网管理器，与多个管理节点相关联，其中多个管理节点使得多个子网管理器能够使用带内通信协议彼此通信，其中多个子网管理器彼此协商并且推选一个子网管理器作为主子网管理器，以及其中所述主子网管理器操作来使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器，并且基于所述一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

附图说明

[0025] 图1示出了根据本发明的实施例的用于中间件机器的示范性配 置的例示。

[0026] 图2示出了根据本发明的实施例的中间件机器环境的例示。

[0027] 图3示出了根据本发明的实施例的用于在中间件机器环境中执行子网初始化的例示。

[0028] 图4示出了根据本发明的实施例的用于在中间件机器环境中执行子网初始化的示范性流程图。

[0029] 图5示出了根据本发明的实施例的具有不同的划分组的中间件机器环境的例示。

[0030] 图6示出了根据如上所述的本发明的一些实施例的中间件机器环境中的子网管理器的说明性的功能框图。

[0031] 图7示出了根据本发明的一些实施例的用于在中间件机器环境中执行子网初始化的系统的说明性的功能框图。

具体实施方式

[0032] 这里描述的是用于提供中间件机器或相似的平台的系统和方法。根据本发明的实施例，该系统包括高性能硬件(例如，64位处理器技术、高性能大容量存储器和冗余无限带宽和以太网联网)与诸如WebLogic套件之类的应用服务器或中间件环境的组合，以提供完整的Java EE应用服务器联合体，该Java EE应用服务器联合体包括整体上并行的存储器内网格，可以被快速提供，并且可以根据需要缩放。根据本发明的实施例，该系统可以被布置为完全的、一半或四分之一机架，或其它配置，这些配置提供应用服务器网格、存储区域网络和无限带宽(IB)网络。中间件机器软件可以提供应用服务器、中间件和诸如例如WebLogic Server、JRockit或Hotspot JVM、Oracle Linux或Solaris和Oracle VM之类的其它功能。根据本发明的实施例，该系统可以包括经由IB网络彼此通信的多个计算节点、一个或多个IB交换机网关和存储节点或单元。当被实现为机架配置时，机架的不使用的部分可以保持为空或被填充物占据。

[0033] 根据本发明的实施例，这里称为“Sun Oracle Exalogic”或“Exalogic”的系统是对于容纳诸如Oracle Middleware SW套件或Weblogic之类的中间件或应用服务器软件的容易布置的解决方案。如这里所述，根据实施例，系统是“盒装网格(grid in a box)”，其包括一个或多个服务器、存储单元、用于存储联网的IB结构、以及容纳中间件应用所需的所有其它组件。显著的性能可以通过使用例如Real Application Clusters和Exalogic Open存储器来平衡整体上并行的网格架构而对于所有类型的中间件应用分发。该系统利用线性I/O可伸缩性分发改善的性能、使用和管理起来简单、并且分发对任务关键的可用性和可靠性。

[0034] 图1示出了根据本发明的实施例的用于中间件机器的示范性配置的例示。如图1所示，中间件机器100使用单个机架配置，其包括两个网关网络交换机或连接到二十八个服务器节点的叶子网络交换机102和103。另外，可以存在对于中间件机器的不同的配置。例如，可以存在一半机架配置，其包含服务器节点的一部分，并且也可以存在多机架配置，其包含大量服务器。

[0035] 如图1所示，服务器节点可以连接到由网关网络交换机提供的端口。如图1所示，每个服务器机器可以具有单独地到两个网关网络交换机102和103的连接。例如，网关网络交换机102连接到服务器1-14106的端口1和服务器15-28 107的端口2，并且网关网络交换机103连接到服务器1-14 108的端口2和服务器15-28 109的端口1。

[0036] 根据本发明的实施例，每个网关网络交换机可以具有用于与不同的服务器连接的多个内部端口，并且网关网络交换机也可以具有用于与诸如现有数据中心服务网络之类的外部网络连接的外部端口。

[0037] 根据本发明的实施例，中间件机器可以包括通过网关网络交换机连接到服务器的单独的存储系统110。另外，中间件机器可以包括连接到两个网关网络交换机102和103的干

线网络交换机101。如图1所示，可以可选地存在从存储系统到干线网络交换机的两个链路。

[0038] IB结构/子网

[0039] 根据本发明的实施例，中间件机器环境中的IB结构/子网可以包含以胖树状拓扑互连的大量物理主机或服务器、交换机实例和网关实例。

[0040] 图2示出了根据本发明的实施例的中间件机器环境的例示。如图2所示，中间件机器环境200包括与多个末端节点连接的IB子网或结构220。IB子网包括多个子网管理器211-214，每个子网管理器驻留在多个网络交换机201-204中的一个上。子网管理器可以使用带内通信协议210彼此通信，带内通信协议诸如基于管理数据报(MAD)/子网管理分组(SMP)的协议或诸如基于IB的互联网协议(IPoIB)之类的其它协议。

[0041] 根据本发明的实施例，可以在IB结构上构造单个IP子网，IB结构使得交换机在相同的IB结构中彼此安全地通信(即，所有交换机之间的完全连接)。当在两个交换机之间存在具有操作链路的至少一个路线时，基于该结构的IP子网可以提供任何一对交换机之间的连接。如果通过重新路由而存在可替换的路线，则可以实现从链路故障的恢复。

[0042] 交换机的管理以太网接口可以连接到提供所有交换机之间的IP级别的连接的单个网络。每个交换机可以由两个主要IP地址标识：一个用于外部管理以太网并且一个用于基于该结构的IP子网。每个交换机可以使用两个IP地址监视到所有其它交换机的连接，并且可以使用任一操作地址用于通信。另外，每个交换机可以具有到该结构上的每个直接连接的交换机的点对点IP链路。因此，可以存在至少一个附加的IP地址。

[0043] IP路由设置使得网络交换机能够使用该结构的IP子网、外部管理以太网网络和交换机对之间的一个或多个结构级别的点对点IP链路的组合经由中间交换机将业务路由到另一个交换机。IP路由使得对网络交换机的外部管理访问能够经由网络交换机上的外部以太网端口以及通过该结构上的专用路由服务来路由。

[0044] IB结构包括具有对管理网络的管理以太网访问的多个网络交换 机。存在该结构中的交换机之间的带内物理连接。在一个示例中，当IB结构不退化时，在每一对交换机之间存在一个或多个跳的至少一个带内路由。对于IB结构的管理节点包括连接到IB结构的网络交换机和管理主机。

[0045] 子网管理器可以经由它的私有IP地址中的任何一个被访问。子网管理器也可以经由浮动IP地址访问，当子网管理器起主子网管理器的作用时，浮动IP地址被配置用于主子网管理器，并且当子网管理器被从该角色明确地释放时，子网管理器被解配置。可以对于外部管理网络以及对于基于该结构的管理IP网络二者定义主IP地址。对于点对点IP链路，不需要定义特殊的主IP地址。

[0046] 根据本发明的实施例，可以使用基于虚拟机的客户机将每个物理主机虚拟化。可以每个物理主机同时存在多个客户机，例如每个CPU核一个客户机。另外，每个物理主机可以具有至少一个双端口主机通道适配器(HCA)，其可以被虚拟化并且在客户机之间共享，以使得虚拟化的HCA的结构图是单个双端口HCA，正如非虚拟化的/共享的HCA一样。

[0047] IB结构可以分为由IB划分实现的动态集的资源域。IB结构中的每个物理主机和每个网关实例可以是多个划分的成员。此外，相同的或不同的物理主机上的多个客户机可以是相同的或不同的划分的成员。对于IB结构的IB划分的数目可以由P_Key表大小限制。

[0048] 根据本发明的实施例，客户机可以开启直接从客户机中的vNIC驱动器访问的两个

或更多个网关实例上的一组虚拟网络接口卡(vNIC)。客户机可以在物理主机之间迁移同时保持或具有更新的vNIC伙伴。

[0049] 根据本发明的实施例，交换机可以按照任何顺序启动并且可以根据例如IB指定的协商协议之类的不同的协商协议动态地选择主子网管理器。如果没有指定划分策略，则可以使用默认的使能划分的策略。另外，可以独立于任何附加的策略信息并且独立于主子网管理器是否知道完整的结构策略，来建立管理节点划分和基于结构的管理IP子 网。为了使得使用基于该结构的IP子网同步结构级别的配置策略信息，子网管理器可以最初使用默认划分策略启动。当已经实现结构级别的同步时，对于结构当前的划分配置可以由主子网管理器安装。

[0050] 部分子网初始化

[0051] 根据本发明的实施例，能够在中间件机器环境中执行子网初始化。中间件机器环境中的多个管理节点能够互连以使得与多个管理节点相关联的多个子网管理器能够使用带内通信协议彼此通信。多个子网管理器能够协商和推选主子网管理器，主子网管理器操作来配置和管理中间件机器环境。主子网管理器能够使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器。主子网管理器然后基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

[0052] 图3示出了根据本发明的实施例的用于在中间件机器环境中执行子网初始化的例示。如图3所示，中间件机器环境300包括管理多个客户端节点或末端节点的IB子网或结构320。IB子网包括多个管理节点301-304，其保持与多个配置守护进程311-314相关联的多个子网管理器321-324。管理节点能够是网络交换机或预定义的主机服务器。子网管理器能够使用诸如基于无限带宽的互联网协议(IPoIB)之类的带内通信协议310彼此通信。

[0053] 根据本发明的实施例，系统能够提供多个策略守护进程311-314，每个策略守护进程与子网管理器321-324相关联。在如图3所示的示例中，与主子网管理器301并置的策略守护进程311负责使用一个或多个策略配置和管理中间件机器环境中的末端节点。

[0054] 由中间件机器环境中的策略守护进程管理的一个示范性策略能够是划分配置策略。例如，包括末端节点A、B和C的中间件机器环境能够被分成两个组：包括节点A和B的组I和包括节点C的组II。划分配置策略能够定义如下划分更新：在将节点B增加到组II中之前需要从组I删除节点B。此划分配置策略可以要求主子网管理器在 不首先从组I删除节点B的情况下不会使得新的划分将节点B增加到组II中。此划分配置策略可以由主子网管理器使用策略守护进程来执行。

[0055] 带内通信连接除了中间件机器环境中的以太网联网之外，还提供通信冗余。系统可以利用中间件机器环境提供的通信冗余并且使用配置之下的IB子网作为用于分布结构初始化策略的管理网络以便完全配置中间件机器环境。

[0056] 根据本发明的实施例，IB子网中的子网管理器可以首先彼此协商并且推选负责配置和管理中间件机器环境的主子网管理器321。在初始化策略事务309的上下文内，主子网管理器可以使用带内通信协议将配置守护进程中的一个或多个结构初始化策略复制到中间件机器环境中的其它主子网管理器候选者322-324。然后，主子网管理器可以基于一个或多个子网初始化策略设置中间件机器环境中的多个末端节点。

[0057] 根据本发明的实施例，每个管理节点可以使用以太网网络连接330与中间件机器

环境之内的诸如主机服务器之类的一个或多个客户端节点或末端节点连接。网络交换机可以是直接与末端节点通信的叶子交换机或通过叶子交换机与末端节点通信的干线交换机。网络交换机可以经由网络交换机的交换机端口和主机服务器的主机端口与主机服务器通信。在IB网络中，可以定义划分以指定哪些末端端口能够与其它末端端口通信。中间件机器环境可以采用胖树状拓扑，其允许少量交换机位于胖树的顶层而保持大量末端节点作为树的叶子。

[0058] 图4示出了根据本发明的实施例的用于在中间件机器环境中执行子网初始化的示范性流程图。如图4所示，在步骤401，多个管理节点能够在中间件机器环境中互连以使得与管理节点相关联的多个子网管理器能够使用带内通信协议彼此通信。在步骤402，多个子网管理器可以首先在彼此之间协商并且推选主子网管理器。主子网管理器负责配置和管理中间件机器环境。此外，在步骤403，主子网管理器能够使用带内通信协议将一个或多个子网初始化策略复制到中间 件机器环境中的其它子网管理器。然后，在步骤404，主子网管理器可以基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

[0059] 利用划分配置策略的子网初始化

[0060] 图5示出了根据本发明的实施例的具有不同的划分组的中间件机器环境的例示。如图5所示，一个或多个划分策略519可以定义中间件机器环境500中的划分配置。可以通过初始化策略事务509将一个或多个划分策略提供给中间件机器环境。作为划分配置的结果，中间件机器环境可以分为三个单独的划分组：划分A 531、划分B 532和划分C 533。每个划分组可以与唯一的划分密钥相关联并且可以使用单独的连接541-543在逻辑上连接到IB子网520。

[0061] 根据本发明的实施例，位于IB网络中的网络交换机处的子网管理器可以经由网络交换机的交换机端口和主机服务器的主机端口与主机服务器通信。可以定义划分以指定哪些末端端口能够彼此通信。IB划分可以由十六位数字值的P_Key定义，其中较低的十五位定义划分数目，并且最高位定义相关的端口是该划分中的完全成员还是受限成员。受限成员只能够与完全成员通信，而完全成员可以在它们自己之间通信以及与相同划分的任何受限成员通信。

[0062] 每个末端端口可以是至少一个划分的成员以便能够与任何其它末端端口通信(即，在IB传输协议中，P_Key字段的使用不是可选的)。对于交换机端口，划分执行是由子网管理器控制的可选特征。

[0063] 子网管理器通过更新相关末端端口和交换机端口的划分表来控制划分成员资格。如果用于该结构或特定主机的信赖模型暗示主机或主机管理员可能不完全遵循子网管理器已经定义的划分，则可以使用基于交换机的执行以使得交换机输入/输出端口可以执行与子网管理器对于对应的主机端口已经设置的限制相同的限制。

[0064] 端口可以表示的划分的数目是作为端口属性反映到子网管理器的硬件实现参数(例如，在16-64/128范围中)。因此，虽然可以具 有高达32K个对于结构定义的各个P_Key值，但是每个端口可以是有限子集的成员，并且对于直接连接到末端端口的交换机结构输入/输出端口使能划分的交换机执行。

[0065] 默认划分密钥(0xFFFF-完全或0x7FFF-有限)可以通过技术规范预留并且用于访问子网管理器(默认)。在使用划分执行的系统中，末端节点是默认划分的有限的成员，以便

能够与子网管理器通信。如果不使用划分执行，则末端节点是默认划分的完全成员。零值可以被预留并且用于表示无效的P_Key值。

[0066] 为了在基于该结构的子网当中同步结构级别的配置策略信息，主子网管理器可以最初使用默认划分策略启动。当结构级别的同步已被实现时，主子网管理器可以接着在中间件机器环境中安装划分配置。

[0067] 根据本发明的实施例，对于仅仅是默认划分的受限成员的末端节点使用默认配置，这使得在可以实现正确的当前划分策略之前在末端节点之间没有非法的通信。主子网管理器可以设置管理节点作为预定义的划分的完全成员，由此使得策略守护进程使用IPolB通信，以使得当前动态划分策略的复制/同步可以经由IB结构发生。然而，这样的强迫末端节点变为默认划分的仅仅受限成员的方案具有如下缺点：当在当前主子网管理器接管子网时该结构对于末端节点之间的正在进行的通信已经操作时，中断了现有通信。

[0068] 根据本发明的实施例，为了避免正在进行的合法的通信的中断，并且为了在策略守护进程已经能够同步之前在所有指定的管理节点之间提供基于IPolB的通信，新的子网管理器可以对于已经操作的节点保持当前划分成员资格不变并且可以保证对于管理节点仅仅设置预定义的划分。当前不操作的非管理节点可以不被初始化(即，需要从主子网管理器的初始化)。或可替换地，非管理节点可以仅仅被设置具有默认划分中的受限成员资格。然后，当当前划分策略已被完全同步时，所有末端节点可以被完全初始化为具有当前划分策略定义的完全划分成员资格。

[0069] 由于任何当前主子网管理器可以初始化IB结构以遵照当前划分 策略，并且主子网管理器的故障可以将结构保持在它被完全初始化的状态下或处于当前策略已被部分实现的状态下，因此上述方案可以是有用的。这包括旧的主子网管理器在它完全去掉由旧的策略定义而在新的策略中不存在的成员资格之前发生故障的情况。

[0070] 根据本发明的实施例，策略守护进程之间的复制和同步可以保证法定数目或大多数的主子网管理器候选者在经历任何变化之前被更新为新的策略。然后，实际的去掉旧的划分成员资格并且添加新的成员资格可以由单个当前主子网管理器完全地实现，或者由两个或更多个主子网管理器实例在两个或更多个部分步骤中实现，作为子网中控制权的故障转移(fail-over)或移交的结果。每个这样的部分步骤可以或可以不改变IB子网的任何状态但是可以保证不引入任何中间状态，所述中间状态在没有混合两个版本的任何策略的机会的情况下既不是由旧的策略定义的成员资格的子集也不是新的策略的子集。此外，在一个或多个故障转移或移交事件的情况下，新的主子网管理器可以不取消由前一主子网管理器执行的操作，而是可以从基于新的主子网管理器动态地发现的IB子网的当前策略和当前状态的相关的下一步骤开始。

[0071] 根据一些实施例，图6示出了在中间件机器环境中根据如上所述的本发明的原理配置的子网管理器600的说明性的功能框图，图7示出了根据如上所述的本发明的原理配置的子网初始化系统700的说明性的功能框图，包括多个管理节点702，每个管理节点包括如图6所示的子网管理器600。子网管理器和子网初始化系统的功能块可以由硬件、软件或硬件和软件的组合实现以执行本发明的原理。本领域技术人员将理解，图6和7中描述的功能块可以被组合或分成子块以实现如上所述的本发明的原理。因此，这里的描述可以支持这里描述的功能块的任何可能的组合或分离或者进一步定义。

[0072] 如图6所示，子网管理器600可以包括通信模块602、推选模块604、复制模块606和设置模块608。通信模块603可以被配置为使用带内通信协议与其它子网管理器通信。推选模块604可以被配置为与 其它子网管理器协商并且推选主子网管理器。复制模块606可以被配置为，响应于子网管理器600被推选为主子网管理器，使用带内通信协议将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器。设置模块608可以被配置为，响应于子网管理器600被推选为主子网管理器，基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点。

[0073] 在一些实施例中，带内通信协议可以除了中间件机器环境中的以太网联网之外，还提供通信冗余。

[0074] 在一些实施例中，一个或多个划分策略可以定义中间件机器环境中的划分配置，并且一个或多个划分策略可以通过诸如图5中的初始化策略事务509之类的初始化策略事务提供给中间件机器环境。

[0075] 在一些实施例中，子网管理器600可以可选地包括初始化模块610。初始化模块610可以被配置为，响应于子网管理器600被推选为主子网管理器，在没有指定划分策略时使用默认划分策略用于初始化。

[0076] 在一些实施例中，子网管理器600可以可选地包括同步模块612。同步模块612可以被配置为，响应于子网管理器600被推选为主子网管理器，最初使用默认划分策略进行启动来同步结构级别的配置策略，并且在结构级别的同步已被实现时，接着在中间件机器环境中安装划分配置。

[0077] 如图7所示，子网初始化系统700可以包括多个管理节点702。每个管理节点702可以包括如图6所示的子网管理器600，并且管理节点702可以在中间件机器环境中互连。应当注意，虽然子网管理器600在图7中被示出为包括在管理节点702中，但是它可以另外与管理节点702相关联。例如，子网管理器600可以在物理上与管理节点702分离并且通过相关的管理节点与其它子网管理器通信。多个管理节点702可以是无限带宽(IB)子网的一部分。

[0078] 在一些实施例中，虽然在图7中未示出，但是多个管理节点702可以包括一个或多个网络交换机，其中每个子网管理器驻留在网络交 换机上。

[0079] 在一些实施例中，子网初始化系统700可以可选地包括单独的存储系统704。虽然未示出，但是存储系统704可以与管理节点702连接。

[0080] 在一些实施例中，子网初始化系统700可以可选地包括多个主机服务器706。虽然未示出，但是主机服务器706可以与管理节点702连接。

[0081] 在一些实施例中，子网初始化系统700可以可选地能够由客户机访问的一个或多个网关实例708。并且在一些实施例中，子网初始化系统700可以可选地包括多个策略守护进程710，其中每个所述策略守护进程710与子网管理器600相关联。与主子网管理器并置的策略守护进程710可以负责使用一个或多个策略配置和管理中间件机器环境。应当注意，虽然网关实例708和策略守护进程710在图7中被显示为与管理节点702分开的部件，但是它们也可以被布置在或驻留在管理节点702上。

[0082] 在一些实施例中，子网初始化系统700中的主子网管理器可以操作来对于已经操作的节点保持当前划分成员资格不变，并且仅仅保证对于新的管理节点设置预定义的划分。可替换地，此操作可以由主子网管理器的特殊的模块或由参考图6描述的模块602-612

中的一个执行。

[0083] 在一些实施例中，当划分策略对于管理节点702已被完全同步时，子网初始化系统700中的子网管理器600可以操作来将所有末端节点完全初始化为具有在划分策略中定义的完全划分成员资格。可替换地，此操作可以由子网管理器600的特殊的模块或由参考图6描述的模块602-612中的一个执行。

[0084] 可以使用一个或多个传统的通用或专门数字计算机、计算设备、机器或微处理器方便地实现本发明，包括一个或多个处理器、存储器和/或根据本公开的教导编程的计算机可读存储介质。合适的软件编码能够容易地由熟练的程序员基于本公开的教导来准备，这对软件领域 的技术人员将是明显的。

[0085] 在一些实施例中，本发明包括计算机程序产品，其是上面或其中存储指令的存储介质或计算机可读介质(介质)，指令可以用于编程计算机来执行本发明的处理中的任何一个。存储介质能够包括以下中的一个或多个：任何类型的盘，包括软盘、光盘、DVD、CD-ROM、微驱动器、和磁光盘、ROM、RAM、EPROM、EEPROM、DRAM、VRAM、闪速存储器件、磁或光卡、纳米系统(包括分子记忆IC)或适合于存储指令和/或数据的任何类型的介质或设备，但是不局限于此。

[0086] 本发明的上面的描述是为了说明和描述的目的而提供的。它不意欲是穷举的或将本发明限制于公开的精确的形式。许多修改和变化能够对本领域的实践者是明显的。选择并且描述实施例以便最佳地说明本发明的原理和它的实际应用，从而使得本领域技术人员能够理解用于各个实施例中的并且具有适合于预期的特定使用的各种修改的本发明。预期本发明的范围由以下权利要求书和它们的等价物定义。

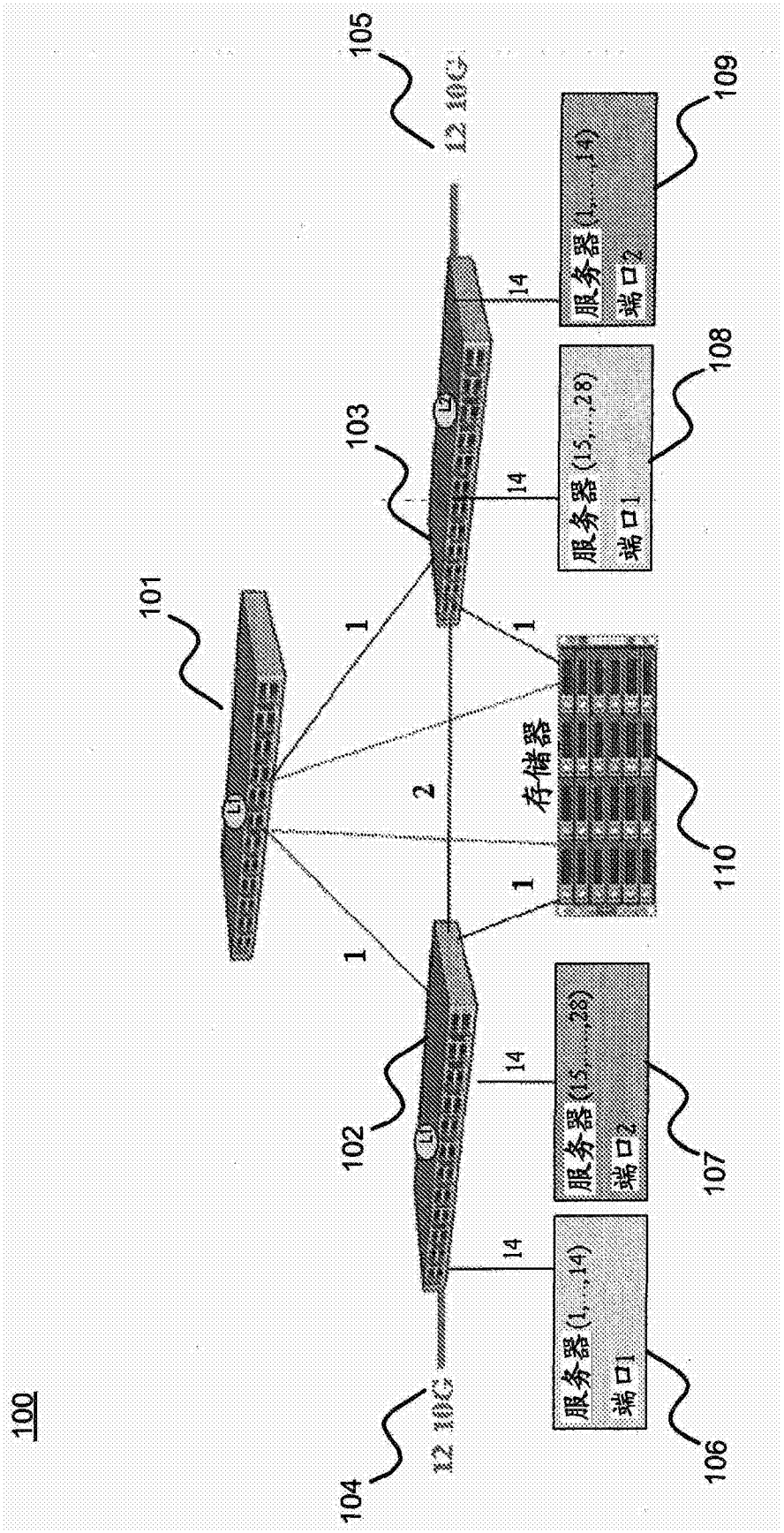

100
101
102
103
104
105
106
107
108
109
110
L1
L2
L3
12 10G
12 10G
1
2
14
存储器
服务器(1,...,14)
端口2
服务器(15,...,28)
端口1
服务器(15,...,28)
端口2
服务器(1,...,14)
端口1

图1

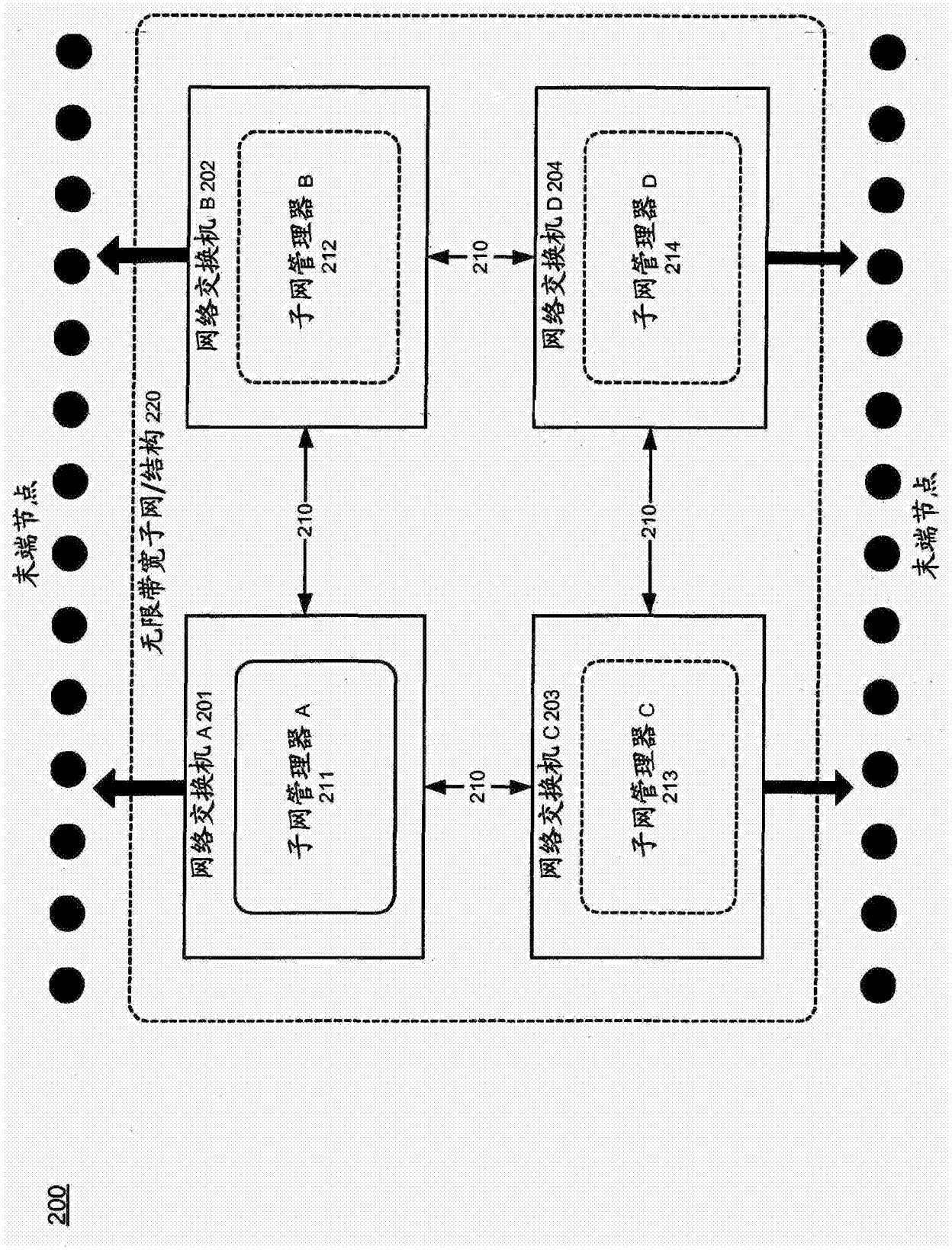

200
末端节点
无限带宽子网/结构 220
网络交换机 A 201
子网管理器 A 211
网络交换机 B 202
子网管理器 B 212
网络交换机 C 203
子网管理器 C 213
网络交换机 D 204
子网管理器 D 214
210
210
210
210
末端节点

图2

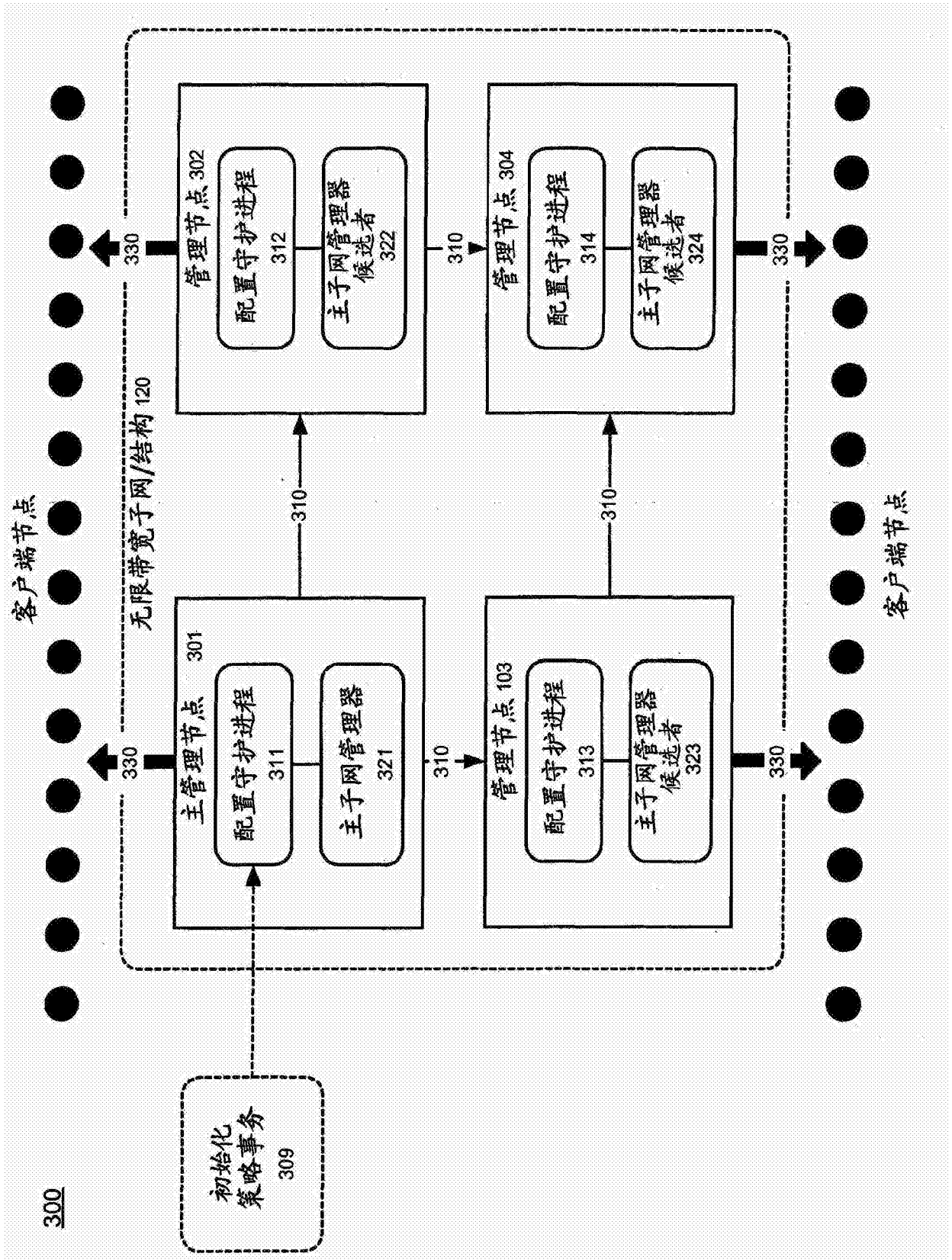

300
客户端节点
无限带宽子网/结构 120
客户端节点
主管理节点 301
配置守护进程 311
主子网管理器 321
管理节点 302
配置守护进程 312
主子网管理器候选者 322
管理节点 304
配置守护进程 314
主子网管理器候选者 324
管理节点 103
配置守护进程 313
主子网管理器候选者 323
初始化策略事务 309
310
330

图3

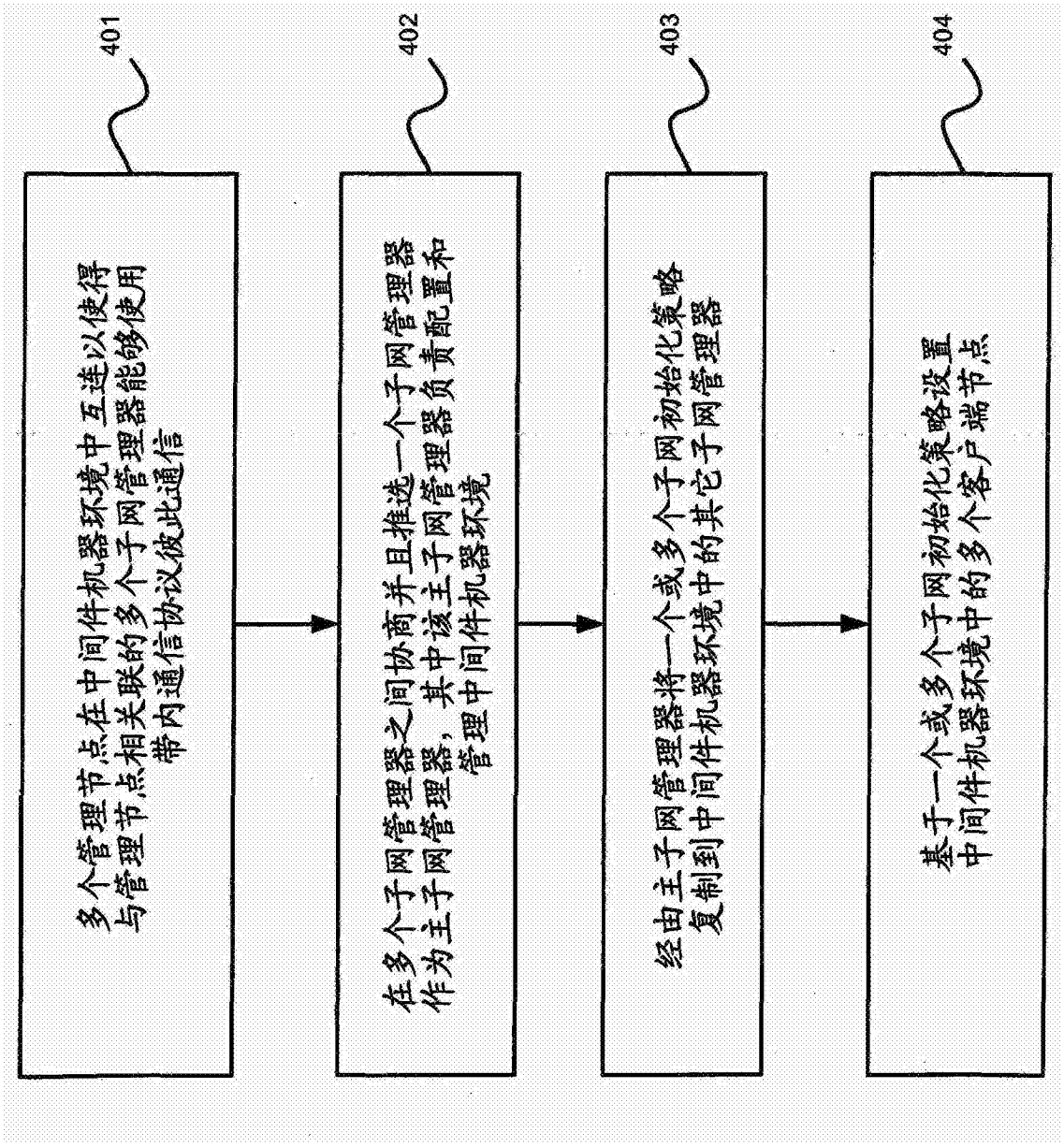

401
多个管理节点在中间件机器环境中互连以使得与管理节点相关联的多个子网管理器能够使用带内通信协议彼此通信
402
在多个子网管理器之间协商并且推选一个子网管理器作为主子网管理器，其中该主子网管理器负责配置和管理中间件机器环境
403
经由主子网管理器将一个或多个子网初始化策略复制到中间件机器环境中的其它子网管理器
404
基于一个或多个子网初始化策略设置中间件机器环境中的多个客户端节点

图4

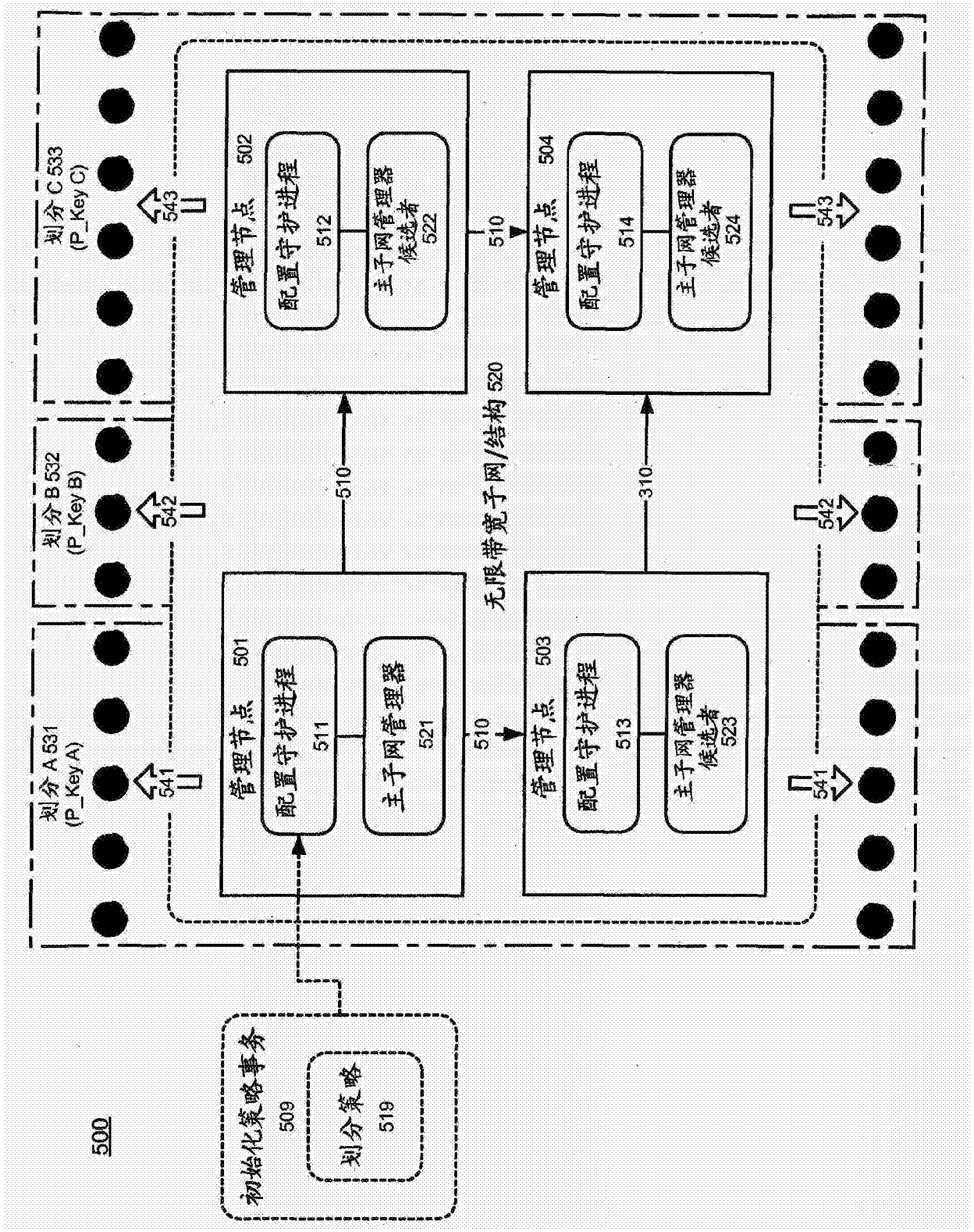

500
划分 A 531
(P_Key A)
划分 B 532
(P_Key B)
划分 C 533
(P_Key C)
541
542
543
管理节点 501
配置守护进程 511
主子网管理器 521
管理节点 502
配置守护进程 512
主子网管理器候选者 522
管理节点 503
配置守护进程 513
主子网管理器候选者 523
管理节点 504
配置守护进程 514
主子网管理器候选者 524
510
310
无限带宽子网/结构 520
初始化策略事务 509
划分策略 519

图5

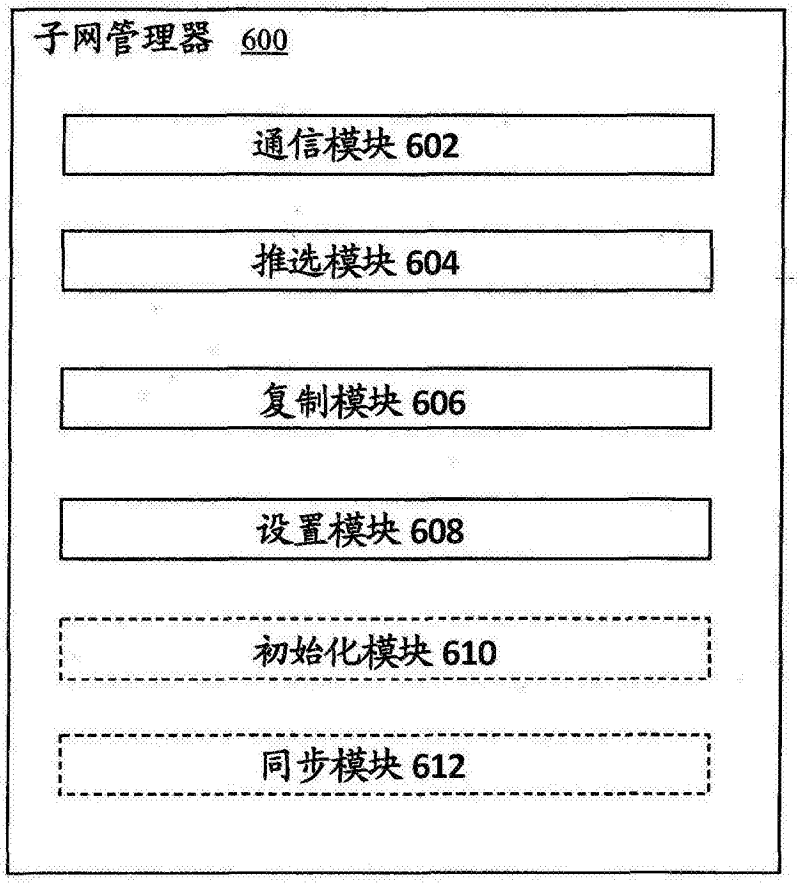

子网管理器 600
通信模块 602
推选模块 604
复制模块 606
设置模块 608
初始化模块 610
同步模块 612

图6

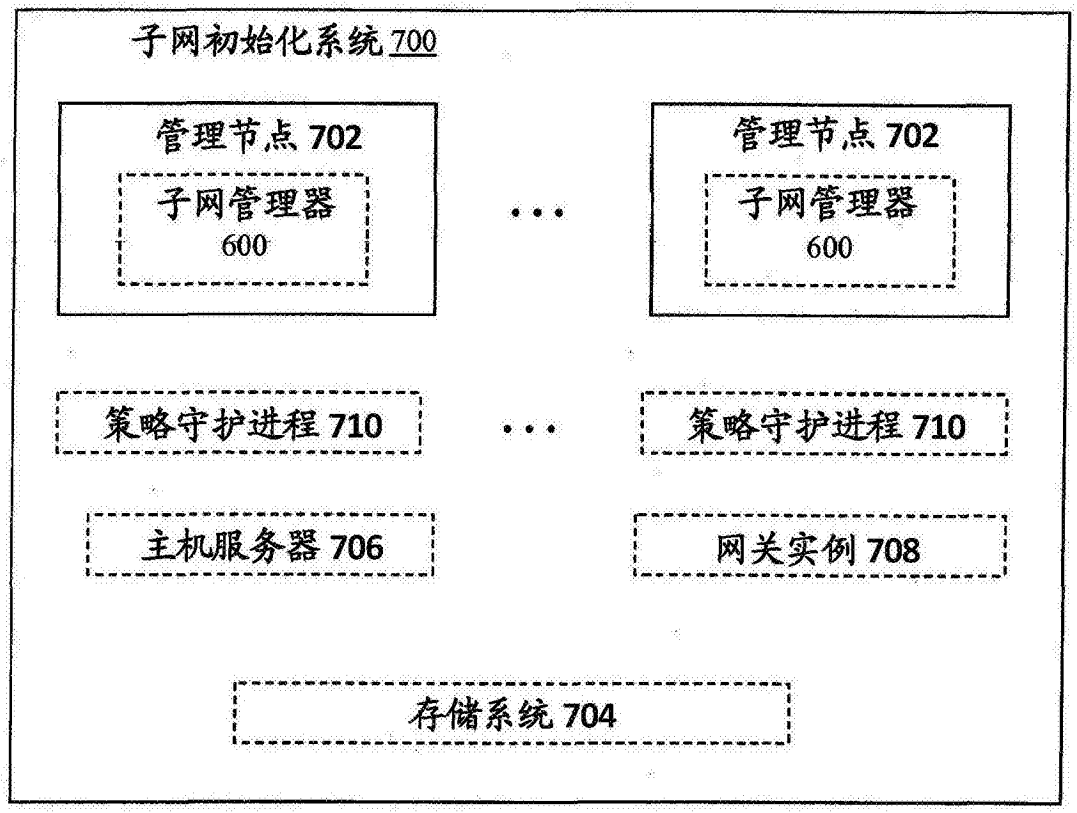

子网初始化系统 700
管理节点 702
子网管理器
600
管理节点 702
子网管理器
600
策略守护进程 710
策略守护进程 710
主机服务器 706
网关实例 708
存储系统 704

图7