



(12)发明专利

(10)授权公告号 CN 104620559 B

(45)授权公告日 2018.03.27

(21)申请号 201380046457.X

(22)申请日 2013.09.06

(65)同一申请的已公布的文献号

申请公布号 CN 104620559 A

(43)申请公布日 2015.05.13

(30)优先权数据

61/698,216 2012.09.07 US

61/701,453 2012.09.14 US

(85)PCT国际申请进入国家阶段日

2015.03.06

(86)PCT国际申请的申请数据

PCT/US2013/058605 2013.09.06

(87)PCT国际申请的公布数据

W02014/039890 EN 2014.03.13

(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 M·法尔科 A·格雷泽

(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 陈新

(51)Int.Cl.

H04L 29/08(2006.01)

G06F 9/50(2006.01)

G06F 9/455(2006.01)

审查员 王璐

权利要求书3页 说明书9页 附图12页

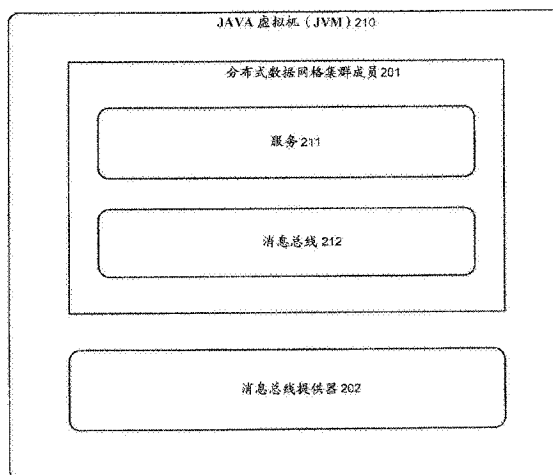
(54)发明名称

用于支持分布式数据网格集群中的可伸缩消息总线的系统和方法

(57)摘要

一种系统和方法可以在分布式数据网格中提供可伸缩消息总线。所述系统可以在分布式数据网格中提供多条消息总线,其中所述分布式数据网格包括多个集群成员并且提供多项服务。此外,所述系统可以将分布式数据网格中的每一项所述服务与一条所述消息总线相关联,并且使用所述多条消息总线来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输。此外,所述系统可以使用数据报层来支持分布式数据网格中的群集,并且绕过分布式数据网格中的数据报层以进行数据传输。

200



1. 一种用于支持分布式数据网格中的可伸缩消息总线的方法,所述分布式数据网格包括各自在一个或多个微处理器上运行并且各自提供多项服务的多个集群成员,所述方法包括:

在分布式数据网格中的所述多个集群成员中的每个集群成员上提供多条消息总线;

将所述多条消息总线中的每条消息总线与由所述多个集群成员中的所述每个集群成员提供的所述多项服务中的单独的服务相关联;以及

使用所述多条消息总线来支持由所述多个集群成员中的所述每个集群成员提供的所述多项服务与所述分布式数据网格中的不同集群成员上的不同服务之间的经由网络的数据传输,并且

其中使用所述多条消息总线进行的所述经由网络的数据传输绕过分布式数据网格的数据报层。

2. 根据权利要求1所述的方法,还包括:

使用所述数据报层来支持分布式数据网格中的所述多个集群成员的群集。

3. 根据权利要求1所述的方法,其中

所述多项服务至少包括高速缓存服务和调用服务。

4. 根据权利要求1所述的方法,还包括:

在所述多个集群成员中的所述每个集群成员上提供消息总线提供器来支持所述多个集群单元中的所述每个集群成员上的所述多条消息总线。

5. 根据权利要求4所述的方法,其中

提供消息总线提供器包括提供支持多种可插式消息总线传输协议的可插式消息总线提供器。

6. 根据权利要求5所述的方法,还包括:

提供用于所述消息总线提供器的配置文件,所述消息总线提供器使用单个开关从所述多种可插式消息总线协议中进行选择。

7. 根据权利要求1所述的方法,还包括:

当所述多个集群成员中的所述每个集群成员上的所述多条消息总线中的消息总线失效时触发所述每个集群成员的死亡检测。

8. 根据权利要求1所述的方法,还包括:

将每个集群成员的所述多条消息总线中的每条消息总线配置为利用线程共用集合来对所述经由网络的数据传输的消息进行并行处理。

9. 根据权利要求1所述的方法,其中

使用所述多条消息总线进行的所述经由网络的数据传输绕过所述分布式数据网格的数据报层并且使用远程直接存储器存取(RDMA)协议。

10. 根据权利要求1所述的方法,还包括:

允许网络转而依赖于用户数据报协议(UDP)。

11. 一种设备,包括用于实现根据权利要求1所述的方法的各个步骤的装置。

12. 一种用于支持分布式数据网格中的可伸缩消息总线的系统,包括:

多个计算机服务器,每个计算机服务器均包括微处理器和存储器;

连接所述多个计算机服务器的网络;

在所述多个计算机服务器上运行的多个集群成员,其中所述多个集群成员中的每个集群成员提供多项服务;

在所述分布式数据网络中的所述多个集群成员中的每个集群成员上的多条消息总线,其中所述多条消息总线中的每条消息总线与由所述多个集群成员中的每个集群成员提供的所述多项服务中的单独的服务相关联;并且

其中所述多条消息总线被用来支持由所述多个集群成员中的所述每个集群成员提供的所述多项服务与所述分布式数据网络中的不同集群成员上的不同服务之间的经由网络的数据传输,并且

其中使用所述多条消息总线进行的所述经由网络的数据传输绕过分布式数据网络的数据报层。

13. 根据权利要求12所述的系统,其中:

分布式数据网络被配置为使用数据报层来支持分布式数据网络中的所述多个集群成员的群集。

14. 根据权利要求12所述的系统,其中:

所述多项服务至少包括高速缓存服务和调用服务。

15. 根据权利要求12所述的系统,还包括:

所述多个集群成员中的所述每个集群成员上的消息总线提供器,其中所述消息总线提供器支持所述多个集群单元中的所述每个集群成员上的所述多条消息总线。

16. 根据权利要求12所述的系统,还包括:

所述多个集群成员中的所述每个集群成员上的可插式消息总线提供器,其中所述可插式消息总线提供器支持所述多个集群成员中的所述每个集群成员上的所述多条消息总线;并且

其中所述可插式消息总线提供器支持多种可插式消息总线传输协议。

17. 根据权利要求12所述的系统,还包括:

所述多个集群成员中的所述每个集群成员上的可插式消息总线提供器,其中所述可插式消息总线提供器支持所述多个集群成员中的所述每个集群成员上的所述多条消息总线,其中所述可插式消息总线提供器支持多种可插式消息总线传输协议;以及

与所述消息总线提供器相关联的配置文件,所述消息总线提供器使用单个开关从所述多种可插式消息总线协议中进行选择。

18. 根据权利要求12所述的系统,其中:

当所述多个集群成员中的所述每个集群成员上的所述多条消息总线中的消息总线失效时,分布式数据网络被配置为触发所述每个集群成员的死亡检测。

19. 根据权利要求12所述的系统,其中:

每个集群成员的所述多条消息总线中的每条消息总线被配置为利用线程共用集合来对所述经由网络的数据传输的消息进行并行处理。

20. 根据权利要求12所述的系统,其中:

使用所述多条消息总线进行的所述经由网络的数据传输使用远程直接存储器存取(RDMA)协议。

21. 根据权利要求19所述的系统,其中:

分布式数据网格被配置为允许网络转而依赖于用户数据报协议 (UDP)。

22. 一种其上存储有指令的非瞬时性机器可读存储介质, 用于支持分布式数据网格中的可伸缩消息总线, 所述分布式数据网格包括各自在一个或多个微处理器上运行并且各自提供多项服务的多个集群成员, 所述指令当被所述多个集群成员中的每个集群成员执行时使每个集群成员执行以下步骤:

在分布式数据网格中的所述多个集群成员中的每个集群成员上提供多条消息总线;

将所述多条消息总线中的每条消息总线与由所述多个集群成员中的所述每个集群成员提供的所述多项服务中的单独的服务相关联; 以及

使用所述多条消息总线来支持由所述多个集群成员中的所述每个集群成员提供的所述多项服务与所述分布式数据网格中的不同集群成员上的不同服务之间的经由网络的数据传输, 并且

其中使用所述多条消息总线进行的所述经由网络的数据传输绕过分布式数据网格中的数据报层。

用于支持分布式数据网格集群中的可伸缩消息总线的系统和 方法

[0001] 版权提醒

[0002] 本专利文献的公开内容的一部分包含受到版权保护的资料。版权所有人不反对任何人按照其在专利商标局专利文件或记录中所出现的那样对专利文献或专利公开内容进行复制再现,但是在其他方面保留所涉及的所有版权权利。

技术领域

[0003] 本发明总体上涉及计算机系统,并且特别涉及分布式数据网格。

背景技术

[0004] 现今的计算系统,特别是由大型组织和企业所采用的那些计算系统的尺寸和复杂度都不断增加。在例如因特网应用之类的领域中,预期应当有数以百万计的用户能够同时访问该应用,从而实际上导致由用户生成和消费的内容数量以及涉及该内容的事务的指数增加。这样的活动还导致对于数据库和元数据存储库的事务调用次数的相应增加,而所述数据库和元数据存储库适应该需求的容量有限。

[0005] 此外,现今的应用采用了水平伸缩(scale out)架构以作为针对成本、规模以及应用可靠性和可服务性方面的挑战的解决方案。这种方法相对于传统方法给出了许多优点,传统方法通常依赖于使用越来越大并且成本越来越高的高端服务器。但是这种方法通常受到一项持久存在的并且具有挑战性的限制:输入/输出(I/O)瓶颈。因此,现今的高度分布式的系统的性能和效率可能受到连接所有系统组件的通信机制的约束。

[0006] 这正是本发明的实施例意图解决的总体领域。

发明内容

[0007] 这里描述了可以支持分布式数据网格中的可伸缩消息总线的系统和方法。所述系统可以在分布式数据网格中提供多条消息总线,其中所述分布式数据网格包括多个集群成员并且提供多项服务。此外,所述系统可以将分布式数据网格中的每一项所述服务与一条所述消息总线相关联,并且使用所述多条消息总线来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输。此外,所述系统可以使用数据报层来支持分布式数据网格中的群集,并且绕过分布式数据网格中的数据报层以进行数据传输。

[0008] 本发明的一个实施例提供一种用于支持分布式数据网格中的可伸缩消息总线的设备,其包括:用于在分布式数据网格中提供多条消息总线的装置,其中所述分布式数据网格包括多个集群成员并且提供多项服务;用于将分布式数据网格中的每一项所述服务与一条所述消息总线相关联的装置;以及用于使用所述多条消息总线来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输的装置。

[0009] 在一个实例中,所述设备还包括:用于使用数据报层来支持分布式数据网格中的群集的装置。

[0010] 在一个实例中,所述设备还包括:用于绕过分布式数据网格中的数据报层以进行数据传输的装置。

[0011] 在一个实例中,所述设备还包括:用于使用消息总线提供器来支持多条消息总线的装置。

[0012] 在一个实例中,所述设备还包括:用于允许消息总线提供器是可插式提供器的装置。

[0013] 在一个实例中,所述设备还包括:用于使用单一开关从多种总线协议当中选择一种总线协议的装置。

[0014] 在一个实例中,所述设备还包括:用于在消息总线失效时触发集群成员上的死亡检测的装置。

[0015] 在一个实例中,所述设备还包括:用于通过网络连接分布式数据网格中的不同集群成员上的不同消息总线的装置。

[0016] 在一个实例中,所述设备还包括:用于允许网络基于远程直接存储器存取 (RDMA) 协议的装置。

[0017] 在一个实例中,所述设备还包括:用于允许网络转而依赖于用户数据报协议 (UDP) 的装置。

[0018] 本发明的另一个实施例提供一种用于支持分布式数据网格中的可伸缩消息总线的系统,其包括:被配置成在分布式数据网格中提供多条消息总线的消息总线提供器,其中所述分布式数据网格包括多个集群成员并且提供多项服务;被配置成将分布式数据网格中的每一项所述服务与一条所述消息总线相关联的关联单元;以及被配置成使用所述多条消息总线来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输的数据传输单元。

[0019] 在一个实例中,所述系统还包括被配置成支持分布式数据网格中的群集的数据报层。

[0020] 在一个实例中,所述系统还包括被配置成绕过分布式数据网格中的数据报层以进行数据传输的绕过单元。

[0021] 在一个实例中,所述消息总线提供器被用来支持多条消息总线。

[0022] 在一个实例中,所述消息总线提供器是可插式提供器。

[0023] 在一个实例中,所述系统还包括被配置成从多种总线协议当中选择一种总线协议的单一开关。

[0024] 在一个实例中,所述系统还包括被配置成在消息总线失效时触发集群成员上的死亡检测的触发单元。

[0025] 在一个实例中,所述系统还包括被配置成通过网络连接分布式数据网格中的不同集群成员上的不同消息总线的连接单元。

[0026] 在一个实例中,所述网络基于远程直接存储器存取 (RDMA) 协议。

[0027] 在另一个实例中,所述网络转而依赖于用户数据报协议 (UDP) 。

附图说明

[0028] 图1示出了基于分布式数据网格中的数据报层支持消息传输的图示。

- [0029] 图2示出了根据本发明的一个实施例在分布式数据网格中提供消息总线的图示。
- [0030] 图3示出了使用基于TCP/IP的传输层支持分布式数据网格中的消息传送的图示。
- [0031] 图4示出了根据本发明的一个实施例使用基于RDMA的传输层支持分布式数据网格中的消息传送的图示。
- [0032] 图5示出了根据本发明的一个实施例支持分布式数据网格中的每服务总线的图示。
- [0033] 图6示出了根据本发明的一个实施例的对应于支持分布式数据网格中的每服务总线的示例性流程图。
- [0034] 图7示出了根据本发明的一个实施例支持分布式数据网格中的并行消息去串行化的图示。
- [0035] 图8示出了根据本发明的一个实施例的对应于支持分布式数据网格中的并行消息去串行化的示例性流程图。
- [0036] 图9示出了根据本发明的一个实施例支持分布式数据网格中的消息预处理的图示。
- [0037] 图10示出了根据本发明的一个实施例的对应于支持分布式数据网格中的消息预处理的示例性流程图。
- [0038] 图11示出了根据本发明的一个实施例的系统的示意性功能方框图。
- [0039] 图12示出了根据本发明的一个实施例的用于支持分布式数据网格中的可伸缩消息总线的系统的功能方框图。

具体实施方式

[0040] 在附图中通过举例而非限制的方式示出了本发明,其中相同的附图标记表示类似的元件。应当提到的是,在本公开内容中提到“某一”、“一个”或“一些”实施例时并不一定都是指相同的实施例,而是意味着至少一个实施例。

[0041] 后面对于本发明的实施例的描述使用Coherence分布式数据网格作为分布式数据网格的一个实例。本领域技术人员将认识到,可以使用其他类型的分布式数据网格而不受限制。此外,后面对于本发明的描述使用Exabus消息传送机制作为消息传送机制的一个实例。本领域技术人员将认识到,可以使用其他类型的消息传送机制而不受限制。

[0042] 在这里描述了可以支持分布式数据网格中的可伸缩消息总线的系统和方法。可伸缩消息总线可以为每一项服务提供其自身的总线(传输引擎)。分布式数据网格可以利用输入/输出(I/O)线程的共用集合(pool)在把入站消息递送到所寻址的服务之前对其进行去串行化,并且可以缓解由于在可以识别出消息类型并且将其卸载到分布式数据网格中的线程共用集合之前通过单线程方式实施所有消息去串行化而导致的瓶颈。此外,分布式数据网格允许在用于可伸缩消息总线的I/O线程上对传入消息进行预处理。

[0043] 分布式数据网格

[0044] 根据一个实施例,这里所提到的“数据网格集群”或“数据网格”是包括在分布式或集群环境内一同工作来管理信息和相关操作(比如计算)的多台计算机服务器的系统。数据网格集群可以被用来管理在各台服务器上共享的应用对象和数据。优选的是,数据网格集群应当具有低响应时间、高吞吐量、可预测的可伸缩性、连续可用性以及信息可靠性。作为

这些能力的结果,数据网络集群非常适合于使用在计算密集的、有状态的中间层应用中。数据网络集群的一些实例(例如Oracle Coherence数据网络集群)可以把信息存储在内存中以获得更高的性能,并且可以采用冗余性来保持该信息的各份拷贝在多台服务器上同步的,从而确保系统的弹性以及服务器故障时的数据可用性。举例来说,Coherence在可靠的、高度可伸缩的对等群集协议的基础上提供复制的并且是分布式(分区式)的数据管理和高速缓存服务。

[0045] 内存中数据网络可以通过将数据分布在一同工作的若干台服务器上来提供数据存储和管理能力。数据网络可以是与应用服务器运行在同一层中或者运行在应用服务器内的中间件。其可以提供对于数据的管理和处理,并且还可以将所述处理推送到数据在网络中所处的位置。此外,当某一台服务器变为不可操作或者从网络断开时,内存中数据网络可以通过自动并且透明的故障切换和重新分配其群集的数据管理服务来消除单一故障点。在添加新的服务器时,或者在重新启动故障服务器时,其可以自动加入集群并且服务可以故障切换回到该处,从而透明地重新分配集群负载。数据网络还可以包括网络层级容错特征以及透明软重启能力。

[0046] 根据一个实施例,数据网络集群的功能基于使用不同的集群服务。集群服务可以包括根集群服务、分区高速缓存服务以及代理服务。在数据网络集群内,在提供和消费集群服务全部两方面,每一个集群节点可以参与若干项集群服务。每一项集群服务具有在数据网络集群内唯一地标识该服务的服务名称,以及定义该集群服务可以做什么的服务类型。除了运行在数据网络集群中的每一个集群节点上的根集群服务之外,还可以有每一种服务类型的多个命名实例。所述服务可以由用户配置,或者由数据网络集群提供以作为默认服务集合。

[0047] 图1示出了基于分布式数据网络中的数据报层支持消息传输的图示。如图1中所示,分布式数据网络100中的集群成员101可以包括一个或多个客户端/服务线程102。集群成员101上的客户端/服务线程102可以利用用户数据报协议(UDP)通过网络(例如以太网网络110)向分布式数据网络100中的其他集群成员发送消息。

[0048] 根据本发明的一个实施例,集群成员101可以采用不同的逻辑,比如分组化逻辑、分组重传逻辑以及Ack/Nack逻辑,以用于向分布式数据网络100中的另一个集群成员发送消息以及接收响应消息。

[0049] 此外,上述消息传送处理可以涉及多次情境切换(context switch)。如图1中所示,客户端/服务线程102可以首先向公布器103发送消息。随后,公布器103可以将消息转发到扬声器104,其负责将消息发送到网络110。

[0050] 此外,分布式数据网络100中的集群成员101可以利用一个或多个侦听器105接收响应消息,其可以将所接收到的消息转发到接收器106。随后,接收器106可以将所接收到的消息转发到客户端/服务线程102,并且可选地通知公布器103。

[0051] 可伸缩消息总线

[0052] 根据本发明的一个实施例,可以使用可伸缩消息总线来消除各个层级处的I/O瓶颈。

[0053] 图2示出了根据本发明的一个实施例在分布式数据网络中提供消息总线的图示。如图2中所示,集群成员201可以运行在分布式数据网络200中的虚拟机210上,比如JAVA虚

拟机。集群成员201可以涉及一项或多项服务211,其可以使用一条或多条消息总线212来进行消息传送。

[0054] 根据本发明的一个实施例,消息总线212可以基于二进制低层级消息传输层,其具有多点寻址和可靠的有序递送。此外,消息总线可以基于纯粹Java实现方式和/或原生实现方式,并且可以采用基于异步事件的编程模型。

[0055] 此外,可以利用联网的硬件和软件子系统,例如Oracle Exalogic工程系统中的Exabus,来支持消息总线212。消息总线不仅可以使得应用运行得更快,而且还可以使得应用运行得更加高效。此外,即使在具有数以千计的处理核心以及数以兆兆字节计的存储器的极大规模部署中并且对于几乎所有商务应用,消息总线也可以使得应用连贯地并且可预测地运行。

[0056] 根据本发明的一个实施例,每一条消息总线212可以是基于提供器的传输层,其可以通过使用虚拟机210中的消息总线提供器202而得到支持,比如JRockit/HotSpot。

[0057] 此外,消息总线提供器202可以基于以可插式提供器为基础的框架。举例来说,消息总线提供器202可以支持不同的消息总线,比如基于TCP/SDP的SocketBus以及基于Infiniband RDMA的InfiniBus。

[0058] 根据本发明的一个实施例,消息总线提供器可以使用单一开关从多种总线协议当中选择一种总线协议。例如在Coherence中,系统可以在以下配置文件中规定所述单一开关:

[0059]

```
Dtangosol.coherence.transport.reliable=protocol
```

```
<cluster-config>!
```

```
<unicast-listener>!
```

```
<reliable-transport>protocol</reliable-transport>
```

[0060] 此外,在Coherence中,系统可以使用所述单一开关选择以下总线的其中之一,比如:

[0061] ■ tmb:TCP消息总线

[0062] ■ sdmb:SDP消息总线

[0063] ■ imb:Infiniband消息总线

[0064] ■ 数据报:传统UDP(默认)

[0065] 因此,消息总线212可以改进分布式数据网格200中的节点内可伸缩性,并且可以使得分布式数据网格200对于协议是不可知的。举例来说,利用消息总线212,分布式数据网格200可以有效地利用大量核心、改进消息传送并发性并且提高吞吐量以及减少等待时间。此外,消息总线212允许分布式数据网格200最小化情境切换并且利用零拷贝。

[0066] 根据本发明的一个实施例,当某一条消息总线失效时,所述系统可以在集群成员上触发死亡检测。

[0067] 图3示出了利用基于TCP/IP的传输层来支持分布式数据网格中的消息传送的图示。如图3中所示,为了在分布式计算环境300中从应用301向应用302发送消息,所述消息可

能需要经过本地机器上的应用缓冲器303、TCP/IP传输层305和内核层306。随后可以在远程机器处在应用缓冲器304中并且经由远程机器中的内核层306和TCP/IP传输层305接收所述消息。

[0068] 图4示出了根据本发明的一个实施例使用基于RDMA的传输层来支持分布式数据网格中的消息传送的图示。如图4中所示,所述系统可以基于以RDMA为基础的传输层从本地机器上的应用401直接向远程机器上的应用401发送消息。

[0069] 每服务总线

[0070] 根据本发明的一个实施例,可伸缩消息总线可以为每一项服务提供其自身的总线(或传输引擎)。

[0071] 图5示出了根据本发明的一个实施例对于分布式数据网格中的各项服务支持可伸缩消息总线的图示。如图5中所示,分布式数据网格500可以包括多个集群成员,例如集群成员501-504。

[0072] 此外,每一个集群成员可以包括不同的服务,其中每一项服务可以与一条单独的消息总线相关联。举例来说,集群成员501可以包括分区高速缓存服务511-512和调用服务513,其可以与消息总线514-516相关联;集群成员502可以包括分区高速缓存服务521-522和调用服务523,其可以与消息总线524-526相关联;集群成员503可以包括分区高速缓存服务531-532和调用服务533,其可以与消息总线534-536相关联;并且集群成员504可以包括分区高速缓存服务541-542和调用服务543,其可以与消息总线544-546相关联。

[0073] 此外,网络510可以连接分布式数据网格500中的不同集群成员501-504上的不同消息总线。举例来说,网络510可以基于远程直接存储器存取(RDMA)协议。此外,网络510在必要时可以转而依赖于用户数据报协议(UDP)。

[0074] 根据本发明的一个实施例,所述系统可以使用多条消息总线来支持分布式数据网格中的不同集群成员之间的数据传输。此外,所述系统可以使用数据报层520来支持分布式数据网格中的群集,并且可以绕过分布式数据网格中的数据报层520以进行数据传输。

[0075] 因此,所述系统允许相对于由末端用户配置的服务数目提高CPU利用率。与传统的联网模型中不同,可以对于每项服务而不是每个集群节点提供单一传输引擎,从而当太多处理器尝试利用单一集群节点时使得分布式数据网格可以缓解瓶颈。

[0076] 图6示出了根据本发明的一个实施例的用于支持消息传输的示例性流程图,其中所述消息传输基于分布式数据网格中的以提供器为基础的传输层。如图6中所示,在步骤601处,所述系统可以在分布式数据网格中提供多条消息总线,其中所述分布式数据网格包括多个集群成员。此外,在步骤602处,所述系统可以把分布式数据网格中的每一项服务与一条所述消息总线相关联,并且在步骤603处,所述系统可以使用多条消息总线来支持分布式数据网格的不同集群成员之间的数据传输。

[0077] 并行消息去串行化

[0078] 根据本发明的一个实施例,可以使用线程共用集合来提供线程,比如输入/输出(I/O)线程,以用于驱动可伸缩消息总线在分布式数据网格(例如Coherence数据网格)中应对入站消息。此外,通过把消息卸载到分布式数据网格内的线程共用集合并且在I/O线程上预先对消息进行去串行化,系统可以最小化服务线程瓶颈的影响。

[0079] 图7示出了根据本发明的一个实施例支持分布式数据网格中的并行消息去串行化

的图示。如图7中所示,分布式数据网格700中的服务线程702可以与消息总线701相关联,消息总线701可以接收一条或多条传入消息,例如消息703-704。

[0080] 此外,消息总线701可以与线程共用集合710相关联,线程共用集合710可以包括一个或多个线程,比如I/O线程711-173。分布式数据网格700可以利用该线程共用集合710来缓解服务线程702处的性能瓶颈。

[0081] 根据本发明的一个实施例,分布式数据网格700可以使用线程共用集合710中的多个不同的I/O线程711-713来并行地处理传入消息703-704。因此,所述系统可以避免由于在可以识别出消息类型之前在单一线程中实施所有消息去串行化而导致的服务线程瓶颈。

[0082] 举例来说,在使用直接存储器存取 (RDMA) 协议时,在将传入消息703递送到服务线程702之前,消息总线701可以使用I/O线程711对消息703进行去串行化。此外,在将传入消息704递送到服务线程702之前,消息总线701可以使用I/O线程713对消息704进行去串行化。

[0083] 图8示出了根据本发明的一个实施例的用于支持分布式数据网格中的并行消息去串行化的示例性流程图。如图8中所示,在步骤801处,系统可以提供线程共用集合,以便提供操作为驱动可伸缩消息总线的多个输入/输出 (I/O) 线程。此外,在步骤802处,系统可以在多个I/O线程上接收一条或多条入站消息,并且在步骤803处,在将所述一条或多条入站消息递送到所寻址的服务之前,系统可以在多个I/O线程上对所述一条或多条入站消息进行去串行化。

[0084] 消息预处理

[0085] 根据本发明的一个实施例,可伸缩消息总线可以提供消息预处理能力,其允许在将所接收到的消息递送到服务线程之前,例如在输入/输出 (I/O) 线程上对所接收到的消息进行预处理。

[0086] 图9示出了根据本发明的一个实施例支持分布式数据网格中的消息预处理的图示。如图9中所示,分布式数据网格900中的服务线程902可以与消息总线901相关联。

[0087] 消息总线901可以使用一个或多个I/O线程(例如I/O线程903)来接收一条或多条传入消息,例如消息905。此外,消息总线901可以使用I/O线程903对传入消息905进行去串行化。

[0088] 一旦将传入消息905去串行化,在将其递送到服务线程902之前,消息总线901可以对传入消息905进行预处理。随后,服务线程902可以进一步完成对于经过预处理的传入消息905的处理,并且在必要时可以向发送传入消息905的服务请求者发送响应907。

[0089] 根据本发明的一个实施例,传入消息905可以提供预处理方法906。消息总线901可以在I/O线程903上执行与传入消息905相关联的预处理方法906,以便部分地或者完全地处理传入消息905。此外,对于各种消息类型,有可能完全应对消息执行(在预处理阶段期间到达结束/响应907)并且完全避免服务线程。

[0090] 根据本发明的一个实施例,可伸缩消息总线的消息预处理能力在被用在分布式数据网格900中时可能是有益的。首先,通过利用消息预处理能力,系统可以避免使得服务线程超负荷,这是因为服务线程可能是分布式数据网格中的瓶颈。其次,利用消息预处理,系统避免了在I/O线程903与服务线程902之间移动消息时可能需要的情境切换。例如在基于远程直接存储器存取 (RDMA) 的传输的情况下,这样的情境切换可能导致总体请求等待时间

的较大百分比。第三,如果可伸缩消息总线具有多个IO线程(比如在基于RDMA的总线的情况下),则可伸缩消息总线允许并行地完全执行消息。

[0091] 此外,可伸缩消息总线可以将消息预处理能力与并行消息去串行化能力相组合,从而可以并行地对多条传入消息进行去串行化和预处理。图10示出了根据本发明的一个实施例的用于支持分布式数据网格中的消息预处理的示例性流程图。如图10中所示,在步骤1001处,系统可以把消息总线与分布式数据网格中的集群成员上的服务线程相关联。此外,在步骤1002处,系统可以利用输入/输出(I/O)线程在消息总线处接收一条或多条传入消息,并且在步骤1003处,在将每一条所述传入消息递送到分布式数据网格中的服务线程之前,系统可以在所述I/O线程上对所述一条或多条传入消息进行预处理。

[0092] 图11示出了根据本发明的一个实施例的用于支持分布式数据网格中的可伸缩消息总线的系统1100的示意性功能方框图。系统1100包括一个或多个微处理器,以及运行在所述一个或多个微处理器上的分布式数据网格。所述分布式数据网格包括多个集群成员,并且提供多个图11,系统1100包括提供器模块1110、关联模块1150以及数据传输模块1190。提供器模块被适配成提供多条消息总线,比如消息总线1120、1130、1140。关联模块被适配成将分布式数据网格中的每一项服务1160、1170、1180分别与消息总线1120、1130、1140相关联。数据传输模块1190被适配成使用多条消息总线(比如消息总线1120、1130、1140)来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输。

[0093] 图12示出了根据如前面所描述的本发明的原理的被配置成支持分布式数据网格中的可伸缩消息总线的系统1200的功能方框图。系统1200的功能方框可以通过硬件、软件或者硬件和软件的组合来实施,以便实现本发明的原理。本领域技术人员应当理解的是,图12中描述的功能方框可以被组合或分离成子方框,以便实施如前面所描述的本发明的原理。因此,这里的描述可以支持这里所描述的功能方框的任何可能的组合或分离或进一步定义。

[0094] 如图所示,系统1200包括消息总线提供器1202、关联单元1204和数据传输单元1206。消息总线提供器1202被配置成在分布式数据网格中提供多条消息总线。所述分布式数据网格可以包括多个集群成员并且提供多项服务。关联单元1204被配置成将分布式数据网格中的每一项服务与一条消息总线相关联。数据传输单元1206被配置成使用所述多条消息总线来支持分布式数据网格中的不同集群成员上的不同服务之间的数据传输。

[0095] 在一个实例中,系统1200还可以包括被配置成支持分布式数据网格中的群集的数据报层1208。

[0096] 在一个实例中,系统1200还可以包括被配置成绕过分布式数据网格中的数据报层1208以进行数据传输的绕过单元1210。

[0097] 在一个实例中,消息总线提供器1202被用来支持消息总线。

[0098] 在一个实例中,消息总线提供器1202是可插式提供器。

[0099] 在一个实例中,系统1200还可以包括被配置成从多种总线协议当中选择一种总线协议的单一开关1212。

[0100] 在一个实例中,系统1200还可以包括被配置成在消息总线失效时触发集群成员上的死亡检测的触发单元1214。

[0101] 在一个实例中,系统1200还可以包括被配置成通过网络连接分布式数据网格中的

不同集群成员上的不同消息总线的连接单元1216。

[0102] 在一个实例中,所述网络基于远程直接存储器存取 (RDMA) 协议。在另一个实例中,所述网络转而依赖于用户数据报协议 (UDP)。

[0103] 本发明可以方便地利用一个或多个传统的通用或专用数字计算机、计算设备、机器或微处理器来实施,其中包括一个或多个处理器、存储器以及/或者根据本公开内容的教导编程的计算机可读存储介质。本领域技术人员将认识到,熟练的程序员可以根据本公开内容的教导很容易地准备适当的软件编码。

[0104] 在一些实施例中,本发明包括作为其上/其中存储有指令的存储介质或(多种)计算机可读介质的计算机程序产品,所述指令可以被用来对计算机进行编程以便实施本发明的任何处理。所述存储介质可以包括(但不限于)任何类型的盘(其中包括软盘、光盘、DVD、CD-ROM、微驱动器和磁光盘)、ROM、RAM、EPROM、EEPROM、DRAM、VRAM、闪存设备、磁性或光学卡、纳米系统(其中包括分子存储器IC)或者适合于存储指令和/或数据的任何类型的介质或设备。

[0105] 前面提供的关于本发明的描述是用于说明和描述的目的。其并不意图进行穷举或者将本发明限制到所公开的确切形式。本领域技术人员将认识到许多修改和变型。所述修改和变型包括所公开的各项特征的任意组合。选择和描述所述实施例是为了最好地解释本发明的原理及其实际应用,从而使得本领域其他技术人员能够理解适合于所设想的具体用途的本发明的各个实施例和各种修改。本发明的范围应当由所附权利要求书及其等效表述限定。

100

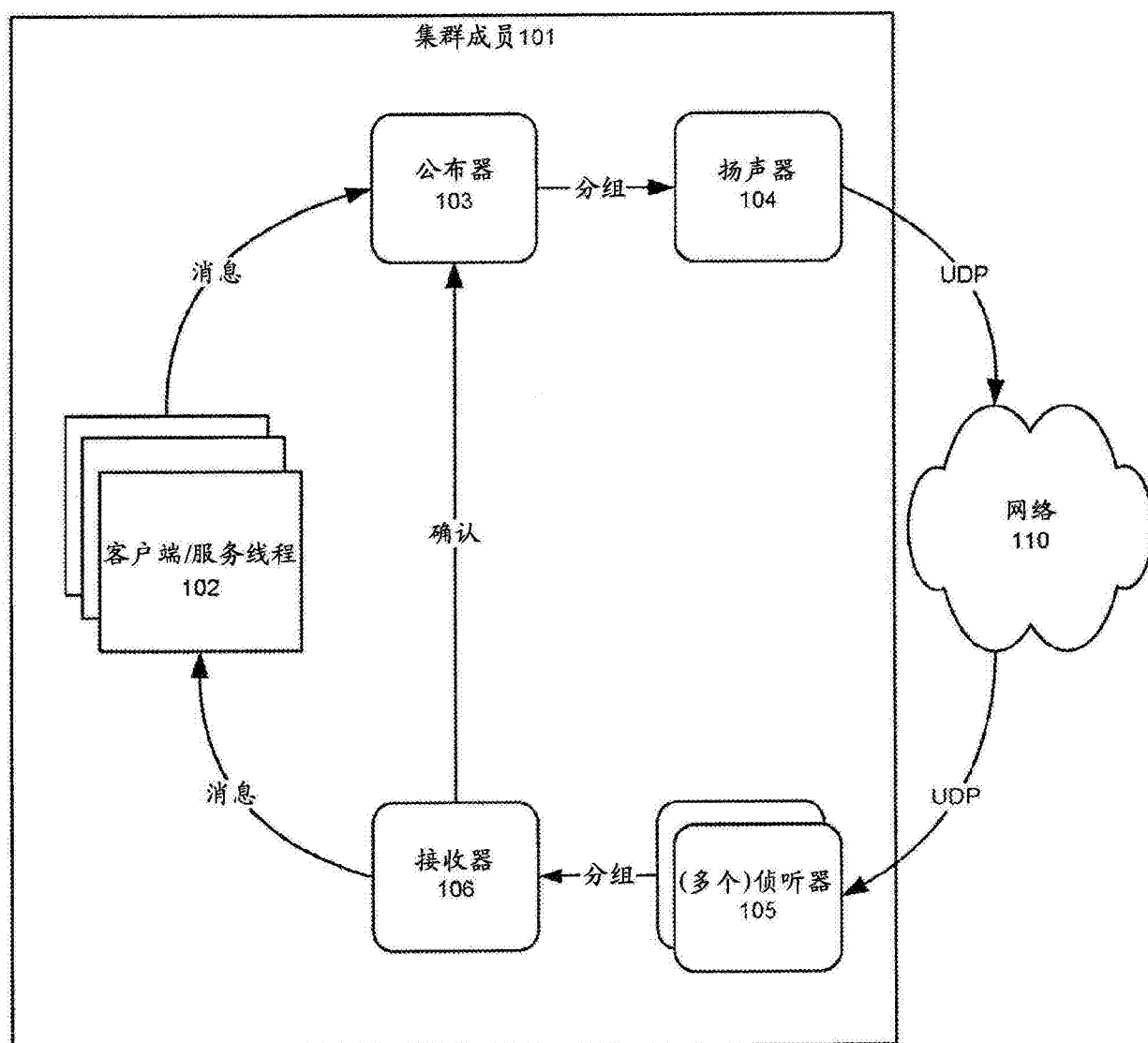


图1

200

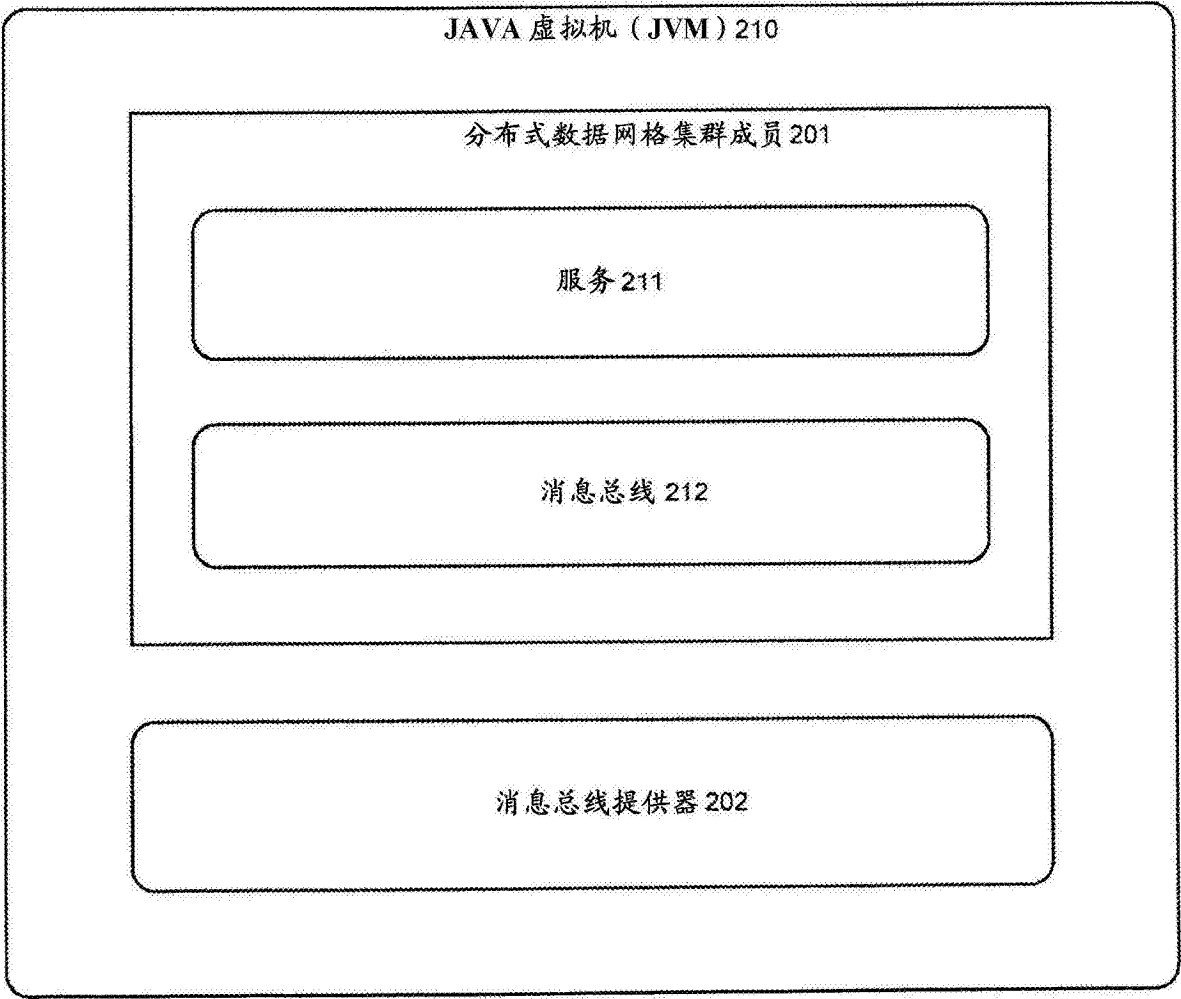


图2

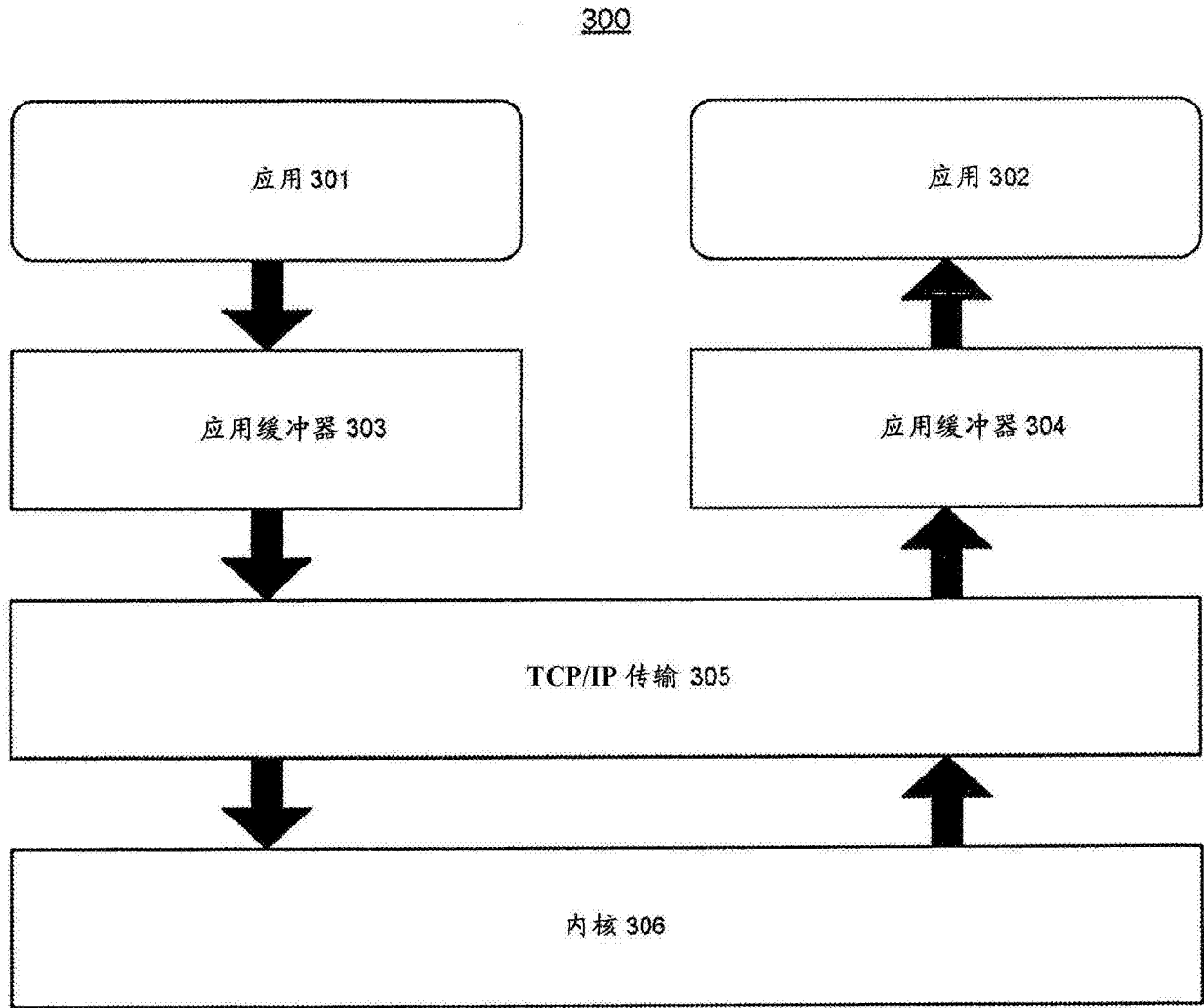


图3

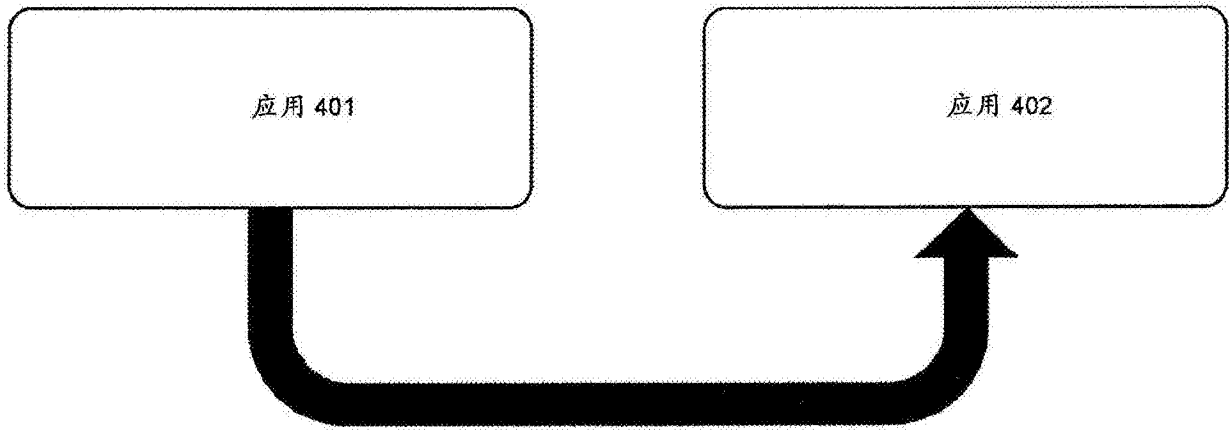
400

图4

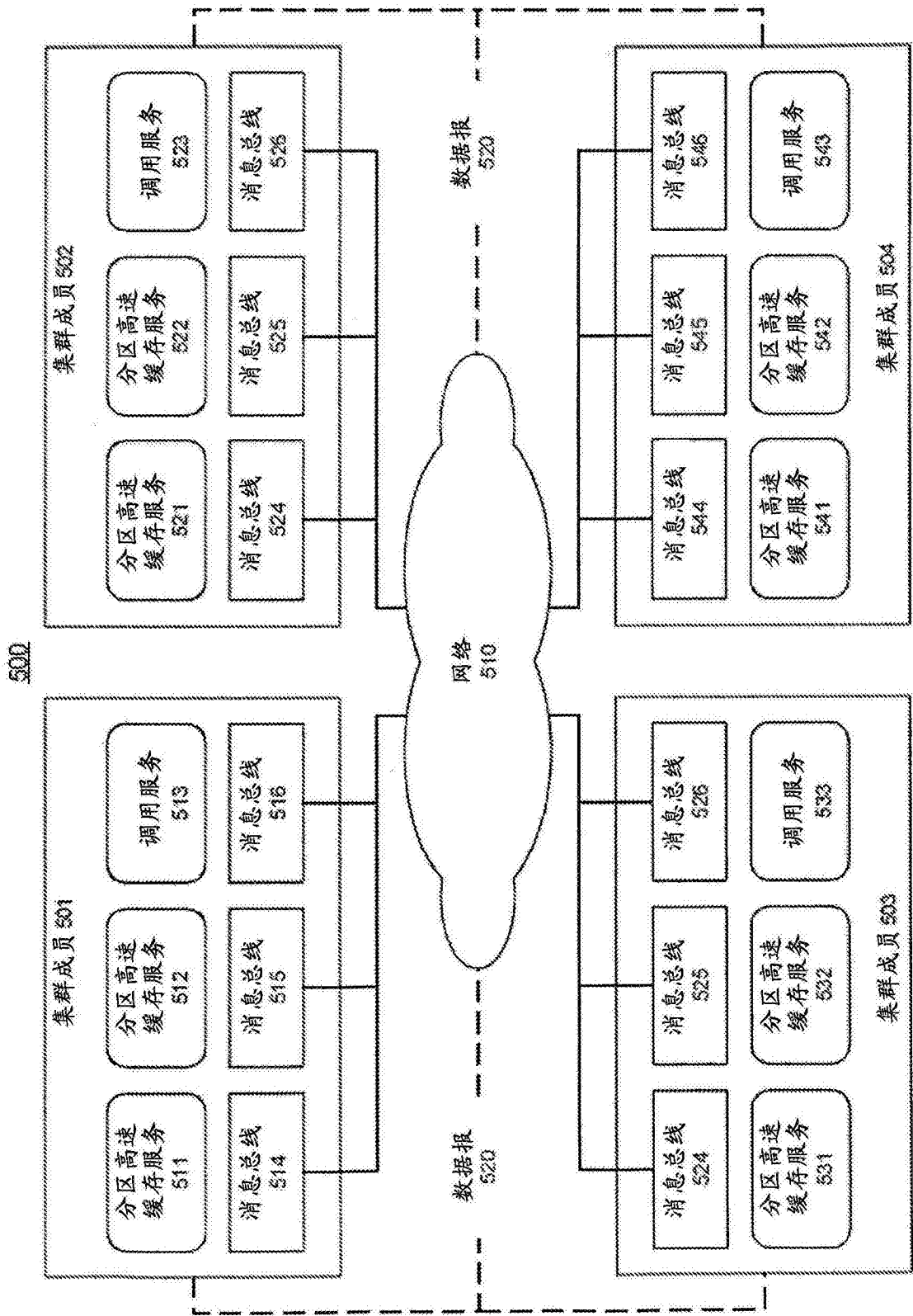


图5

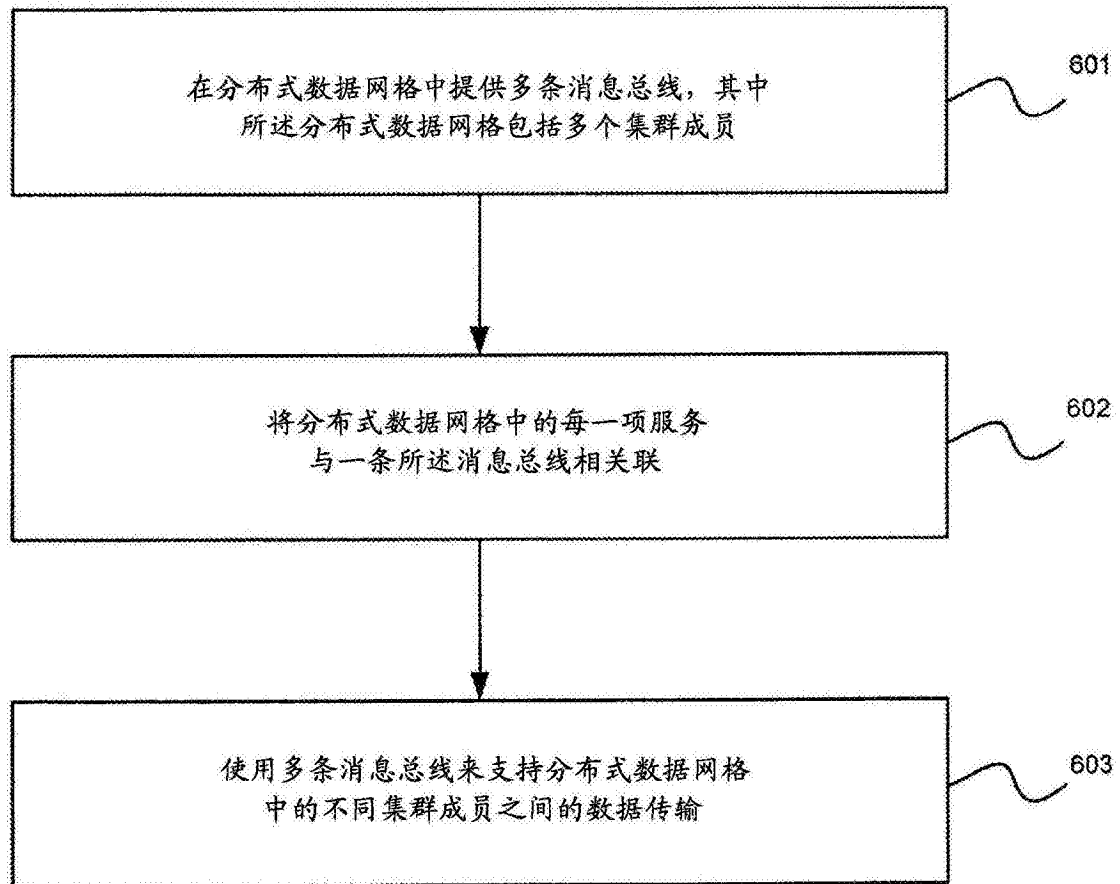


图6

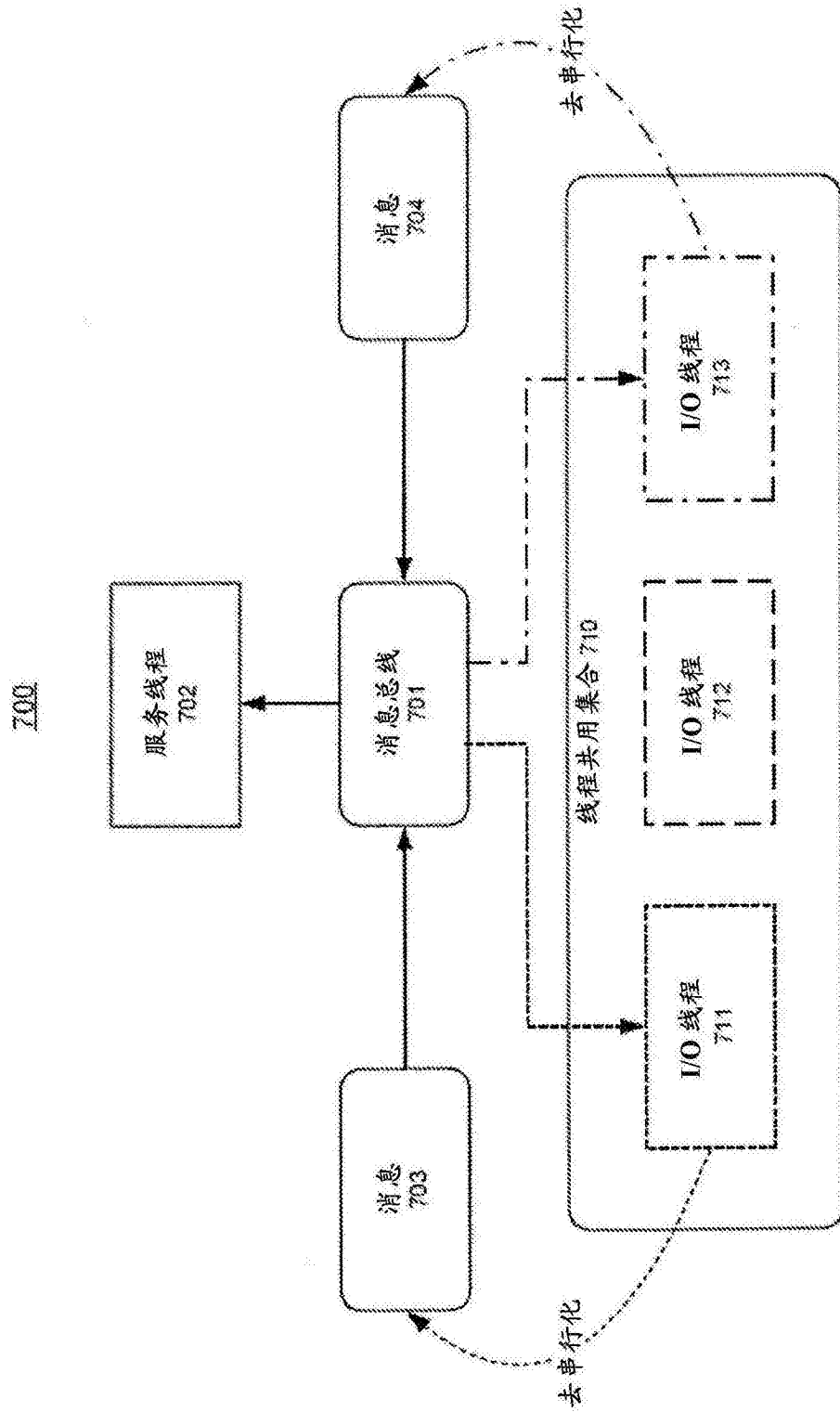


图7

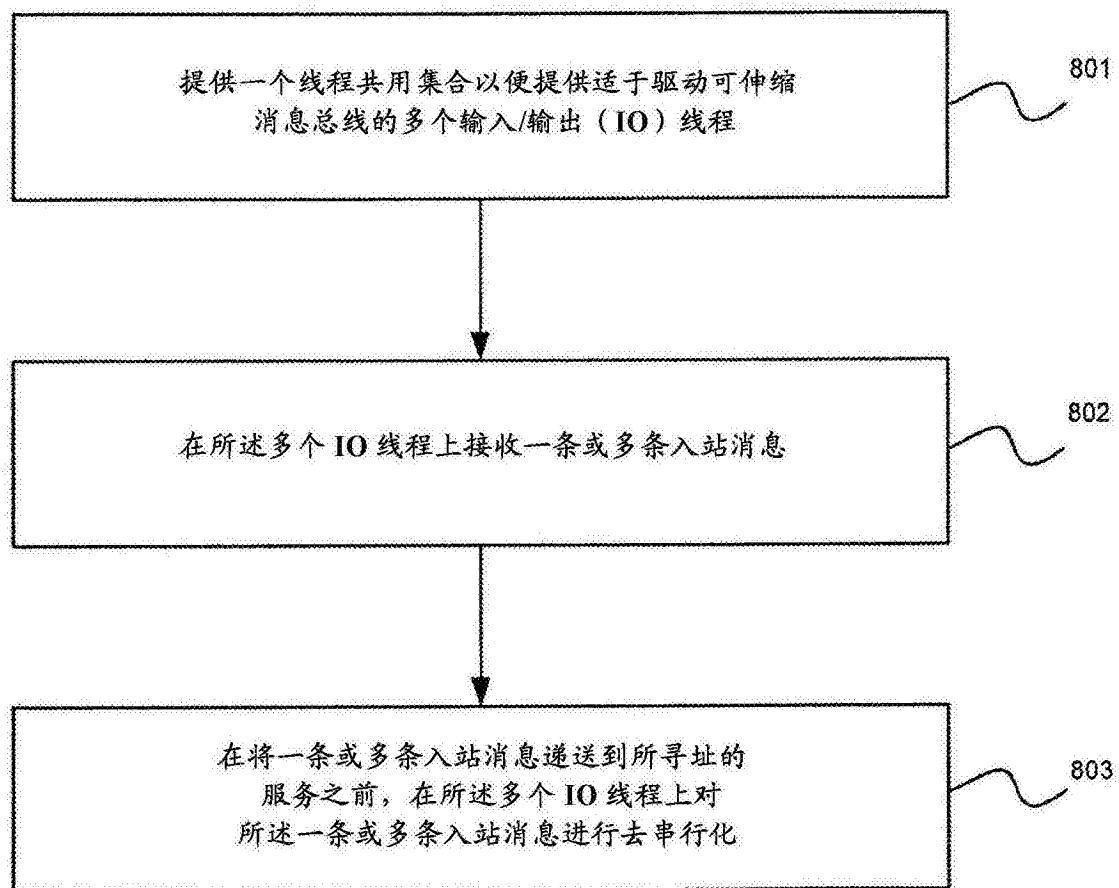


图8

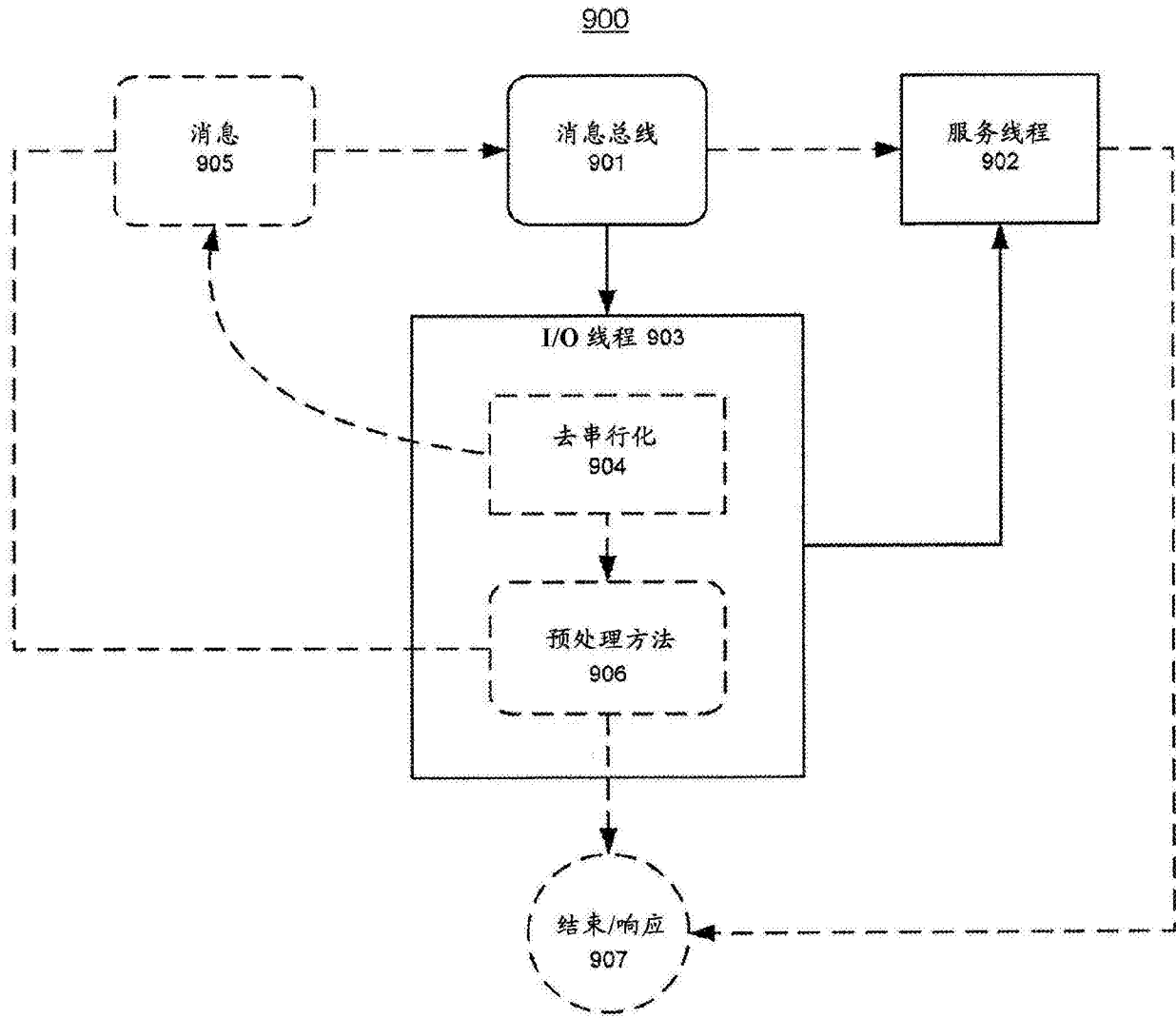


图9

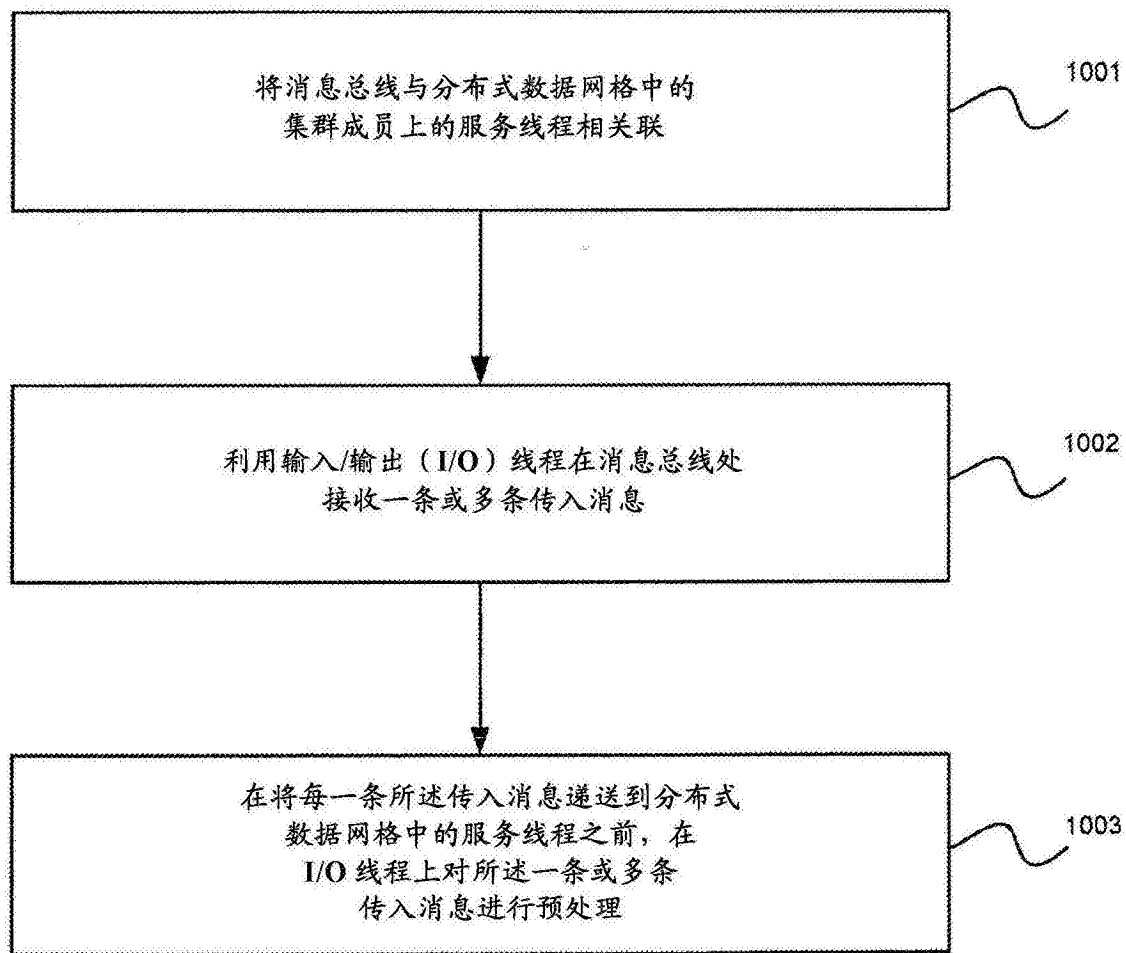


图10

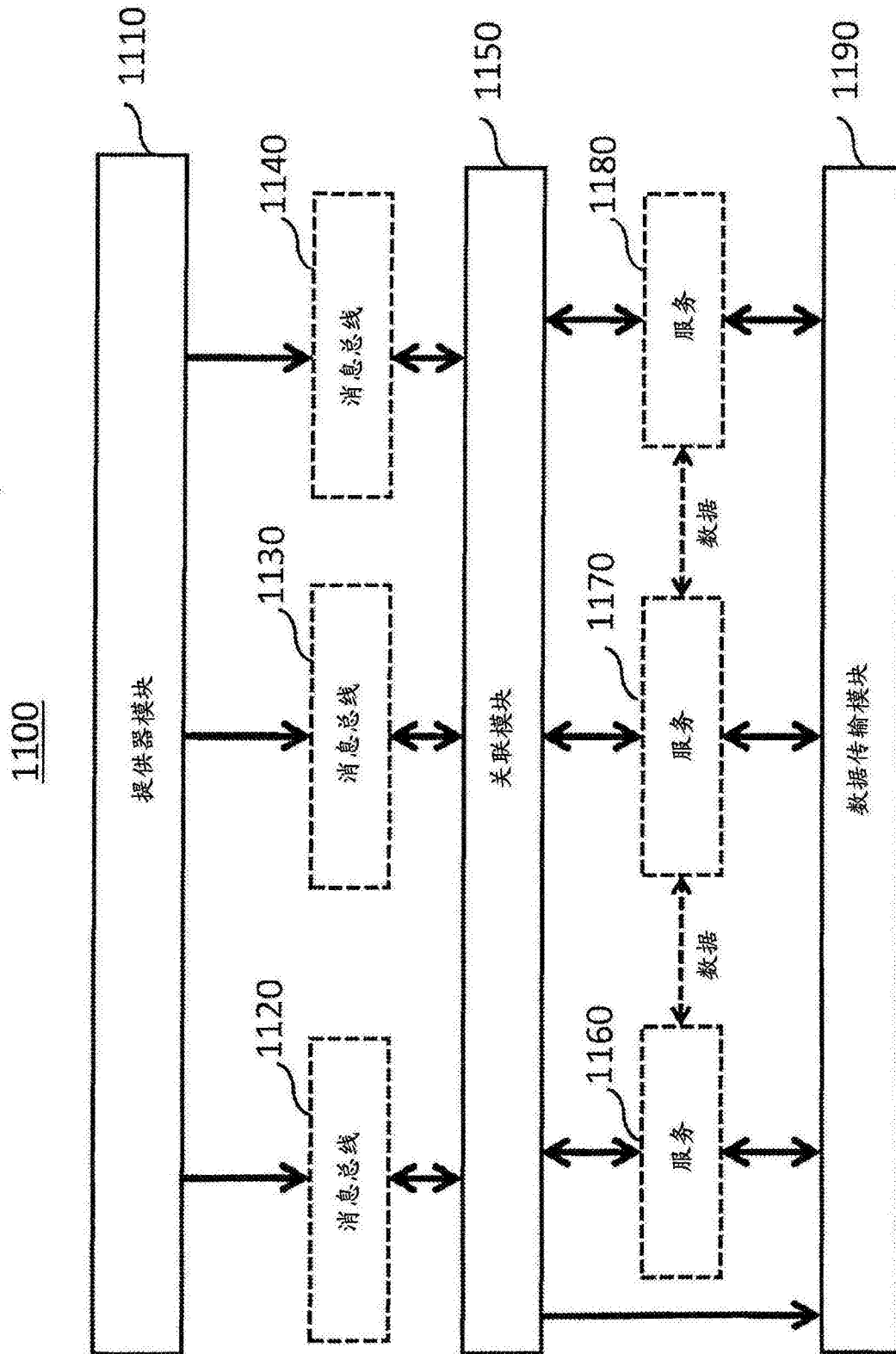


图11

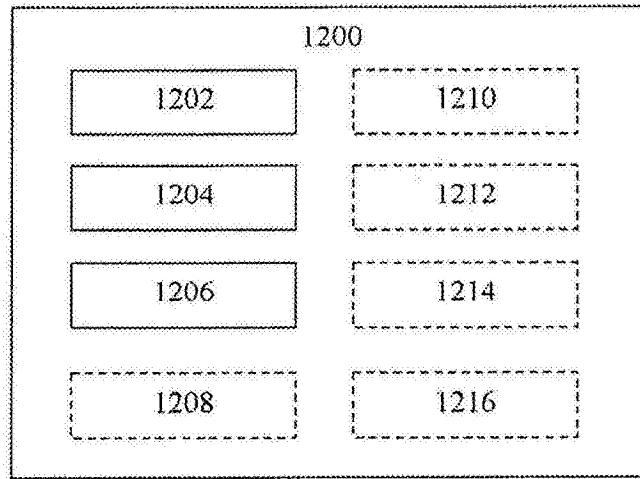


图12