



(12)发明专利

(10)授权公告号 CN 103827830 B

(45)授权公告日 2017.06.09

(21)申请号 201280047483.X

(22)申请日 2012.09.24

(65)同一申请的已公布的文献号

申请公布号 CN 103827830 A

(43)申请公布日 2014.05.28

(30)优先权数据

61/541,054 2011.09.29 US

13/415,670 2012.03.08 US

(85)PCT国际申请进入国家阶段日

2014.03.28

(86)PCT国际申请的申请数据

PCT/US2012/056945 2012.09.24

(87)PCT国际申请的公布数据

W02013/048970 EN 2013.04.04

(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 T·里托 金永顺 E·尼尤

(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 边海梅

(51)Int.Cl.

G06F 9/54(2006.01)

(56)对比文件

CN 1792076 A,2006.06.21,

CN 101227379 A,2008.07.23,

US 7281030 B1,2007.10.09,

Alessandro Forin等.High-performance

distributed objects over system area

networks.《Proceedings of 3rd USENIX

Windows NT Symposium》.1999,

审查员 熊沐阳

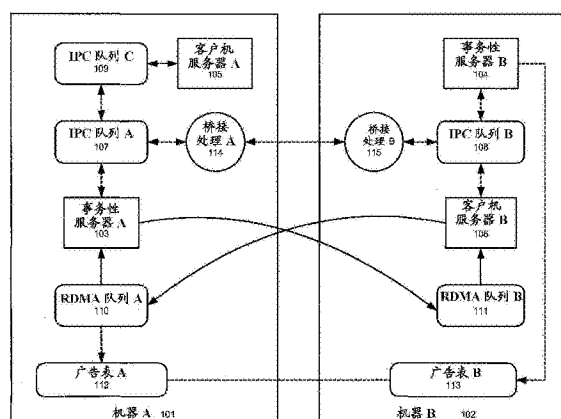
权利要求书4页 说明书8页 附图6页

(54)发明名称

用于在事务性中间件机器环境中防止单点瓶颈的系统和方法

(57)摘要

事务性中间件系统可通过使用远程直接存储器访问(RDMA)协议在本地机器与远程机器之间交换消息以通过类似于本地消息传送的方式实现较短的延时。事务性中间件机器环境可防止单点瓶颈。事务性中间件机器环境包含与第一事务性机器中的服务器相关的第一消息队列,其中,服务器进行操作以通过使用第一消息队列中的队列地址被访问。事务性中间件机器环境还包括与第二事务性机器中的客户机相关的第二消息队列。客户机进行操作以通过使用第一消息队列直接向服务器发送服务请求消息,并且服务器进行操作以从第一消息队列接收服务请求消息并且通过使用第二消息队列直接向客户机发送服务响应消息。



1. 一种用于在事务性中间件机器环境中提供直接消息传送的系统,包括:
与第一事务性机器中的事务性服务器相关的第一消息队列;
所述事务性中间件机器环境中的广告表,其中,所述事务性服务器进行操作以在所述广告表中公布所述第一消息队列;并且
其中,第二事务性机器中的客户机服务器进行操作以在所述广告表中寻找所述第一消息队列并通过使用所述第一消息队列建立与所述第一事务性机器中的事务性服务器的连接。
2. 根据权利要求1的系统,其中,所述客户机服务器与所述第二事务性机器中的第二消息队列相关,其中,所述客户机服务器进行操作以通过使用所述第二消息队列接收消息。
3. 根据权利要求1的系统,其中,所述客户机服务器进行操作以通过使用所述第一消息队列直接向所述事务性服务器发送服务器请求消息。
4. 根据权利要求3的系统,其中,所述事务性服务器进行操作以从所述第一消息队列接收所述服务请求消息,并且通过使用与所述第二事务性机器中的客户机服务器相关的第二消息队列直接向所述客户机服务器发送服务响应消息。
5. 根据权利要求1~4中的任一项的系统,其中,在所述事务性中间件机器环境中的机器之间传送消息时,不存在单点瓶颈。
6. 根据权利要求1~4中的任一项的系统,其中,所述第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。
7. 根据权利要求6的系统,其中,所述网络协议是用于Infiniband (IB) 网络的远程直接存储器访问 (RDMA) 协议。
8. 根据权利要求1的系统,还包括与所述第一事务性机器中的事务性服务器相关的第二消息队列,其中,所述事务性服务器使用所述第二消息队列以用于与所述第一事务性机器中的本地客户机通信。
9. 根据权利要求8的系统,其中,所述第二消息队列是与本地桥接处理相关的处理间通信 (IPC) 队列。
10. 根据权利要求8或9的系统,其中,所述事务性服务器同时在不同的线程中收听第一消息队列和第二消息队列二者,并选择较快的路由来发送消息。
11. 一种用于在事务性中间件机器环境中提供直接消息传送的方法,包括:
提供与第一事务性机器中的事务性服务器相关的第一消息队列;
经由所述事务性服务器在事务性中间件机器环境中的广告表中公布第一消息队列;和
允许第二事务性机器中的客户机服务器进行以下操作:
在广告表中寻找所述第一消息队列;和
通过使用所述第一消息队列建立与所述第一事务性机器中的事务性服务器的连接。
12. 根据权利要求11的方法,还包括使所述客户机服务器与第二事务性机器中的第二消息队列关联,其中,所述客户机服务器进行操作以通过使用所述第二消息队列接收消息。
13. 根据权利要求11的方法,还包括允许所述客户机服务器通过使用所述第一消息队列直接向所述事务性服务器发送服务请求消息。
14. 根据权利要求13的方法,还包括经由所述事务性服务器从所述第一消息队列接收服务请求消息并经由所述事务性服务器通过使用与所述第二事务性机器中的客户机服务

器相关的第二消息队列直接向所述客户机服务器发送服务响应消息。

15. 根据权利要求11~14中的任一项的方法,还包括防止在所述事务性中间件机器环境中的机器之间传送消息时的单点瓶颈。

16. 根据权利要求11~14中的任一项的方法,还包括允许第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。

17. 根据权利要求16的方法,其中,所述网络协议是用于Infiniband (IB) 网络的远程直接存储器访问 (RDMA) 协议。

18. 根据权利要求11的方法,还包括提供与所述第一事务性机器中的事务性服务器相关的第二消息队列,其中,所述事务性服务器使用所述第二消息队列以用于与所述第一事务性机器中的本地客户机通信。

19. 根据权利要求18的方法,其中,所述第二消息队列是与本地桥接处理相关的处理间通信 (IPC) 队列。

20. 根据权利要求18或19的方法,还包括经由所述事务性服务器同时在不同的线程中收听所述第一消息队列和第二消息队列二者,并选择较快的路由来发送消息。

21. 一种用于在事务性中间件机器环境的第一事务性机器中使用的事务性服务器,包括:

公布单元,被配置为在事务性中间件机器环境中的广告表中公布与第一事务性机器中的事务性服务器相关的第一消息队列;以及

建立连接单元,被配置为响应于第二事务性机器中的客户机服务器进行操作以寻找广告表中的第一消息队列,通过使用第一消息队列建立与第二事务性机器中的客户机服务器的连接。

22. 根据权利要求21的事务性服务器,其中,客户机服务器与第二事务性机器中的第二消息队列相关,并且,客户机服务器进行操作以通过使用第二消息队列接收消息。

23. 根据权利要求21的事务性服务器,客户机服务器进行操作以直接通过使用第一消息队列向事务性服务器发送服务器请求消息。

24. 根据权利要求23的事务性服务器,还包括第一通信单元,被配置为从第一消息队列接收服务请求消息,并且通过使用与第二事务性机器中的客户机服务器相关的第二消息队列直接向客户机服务器发送服务响应消息。

25. 根据权利要求21的事务性服务器,其中,在事务性中间件机器环境中的机器之间传送消息时,不存在单点瓶颈。

26. 根据权利要求21的事务性服务器,其中,第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。

27. 根据权利要求21的事务性服务器,其中,网络协议是用于Infiniband (IB) 网络的远程直接存储器访问 (RDMA) 协议。

28. 根据权利要求21的事务性服务器,还包括第二通信单元,被配置为使用与第一事务性机器中的事务性服务器相关的第二消息队列,以与第一事务性机器中的本地客户机通信。

29. 根据权利要求28的事务性服务器,其中,第二消息队列是与本地桥接处理相关的处理间通信 (IPC) 队列。

30. 根据权利要求28的事务性服务器,还包括收听单元,被配置为同时在不同的线程中收听第一消息队列和第二消息队列二者,并选择快速的路由以发送消息。

31. 一种用于事务性中间件机器环境中的第二事务性机器中的客户机服务器,包括:

寻找单元,被配置为在事务性中间件机器环境中的广告表中寻找与第一事务性机器中的事务性服务器相关并且通过其公布的第一消息队列;以及

建立连接单元,被配置为通过使用第一消息队列建立与第一事务机器中的事务性服务器的连接。

32. 根据权利要求31的客户机服务器,还包括接收单元,被配置为通过使用与第二事务性机器中的客户机服务器相关的第二消息队列接收消息。

33. 根据权利要求31的客户机服务器,还包括发送单元,被配置为通过使用第一消息队列直接向事务性服务器发送服务请求消息。

34. 根据权利要求33的客户机服务器,其中,所述事务性服务器进行操作以从第一消息队列接收服务请求消息,并且通过使用与第二事务性机器中的客户机服务器相关的第二消息队列直接向客户机服务器发送服务响应消息。

35. 根据权利要求31的客户机服务器,其中,在事务性中间件机器环境中的机器之间传送消息时,不存在单点瓶颈。

36. 根据权利要求31的客户机服务器,其中,第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。

37. 根据权利要求31的客户机服务器,其中,网络协议是用于Infiniband (IB) 网络的远程直接存储器访问 (RDMA) 协议。

38. 根据权利要求31的客户机服务器,其中,事务性服务器使用与第一事务性机器中的事务性服务器相关的第二消息队列,以与第一事务性机器中的本地客户机通信。

39. 根据权利要求38的客户机服务器,其中,第二消息队列是与本地桥接处理相关的处理间通信 (IPC) 队列。

40. 根据权利要求38的客户机服务器,其中,事务性服务器同时在不同的线程中收听第一消息队列和第二消息队列二者,并选择快速的路由来发送消息。

41. 一种用于在事务性中间件机器环境中提供直接消息传送的设备,包括:

用于提供与第一事务性机器中的事务性服务器相关的第一消息队列的装置;

用于经由所述事务性服务器在事务性中间件机器环境中的广告表中公布第一消息队列的装置;和

用于允许第二事务性机器中的客户机服务器进行以下操作的装置:

在广告表中寻找所述第一消息队列;和

通过使用所述第一消息队列建立与所述第一事务性机器中的事务性服务器的连接。

42. 根据权利要求41的设备,还包括用于使所述客户机服务器与第二事务性机器中的第二消息队列关联的装置,其中,所述客户机服务器进行操作以通过使用所述第二消息队列接收消息。

43. 根据权利要求41的设备,还包括用于允许所述客户机服务器通过使用所述第一消息队列直接向所述事务性服务器发送服务请求消息的装置。

44. 根据权利要求43的设备,还包括用于经由所述事务性服务器从所述第一消息队列

接收服务请求消息的装置,以及用于经由所述事务性服务器通过使用与所述第二事务性机器中的客户机服务器相关的第二消息队列直接向所述客户机服务器发送服务响应消息的装置。

45.根据权利要求41~44中的任一项的设备,还包括用于防止在所述事务性中间件机器环境中的机器之间传送消息时的单点瓶颈的装置。

46.根据权利要求41~44中的任一项的设备,还包括用于允许第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据的装置。

47.根据权利要求46的设备,其中,所述网络协议是用于Infiniband (IB) 网络的远程直接存储器访问 (RDMA) 协议。

48.根据权利要求41的设备,还包括用于提供与所述第一事务性机器中的事务性服务器相关的第二消息队列的装置,其中,所述事务性服务器使用所述第二消息队列以用于与所述第一事务性机器中的本地客户机通信。

49.根据权利要求48的设备,其中,所述第二消息队列是与本地桥接处理相关的处理间通信 (IPC) 队列。

50.根据权利要求48或49的设备,还包括经由所述事务性服务器同时在不同的线程中收听所述第一消息队列和第二消息队列二者的装置,以及用于选择较快的路由来发送消息的装置。

用于在事务性中间件机器环境中防止单点瓶颈的系统和方法

[0001] (版权声明)

[0002] 本专利文件的公开的一部分包含受到版权保护的材料。当它出现于专利商标事务所专利文件或记录中时,版权所有人不反对任何人对专利文件或专利公开的传真再现,但另外保留所有版权。

技术领域

[0003] 本发明一般涉及计算机系统和诸如中间件的软件,特别是涉及支持事务性中间件机器环境。

背景技术

[0004] 事务性中间件系统或面向事务的中间件包含可处理组织内的各种事务的企业应用服务器。伴随诸如高性能网络和多处理器计算机的新技术的发展,需要进一步提高事务性中间件的性能。这些一般是本发明的实施例针对的领域。

发明内容

[0005] 这里描述的是用于通过使用远程直接存储器访问(RDMA)协议在事务性中间件机器环境中支持本地机器与远程机器之间的消息交换的系统和方法。事务性中间件机器环境可防止单点瓶颈并以类似于本地消息传送的方式实现较短的延时。事务性中间件机器环境包含第一消息队列和第二消息队列。第一消息队列与第一事务性机器中的服务器相关,并且该服务器可通过使用第一消息队列中的队列地址被访问。第二消息队列与第二事务性机器中的客户机相关。客户机可通过使用第一消息队列直接向服务器发送服务请求消息,并且服务器可从第一消息队列接收服务请求消息并且通过使用第二消息队列直接向客户机发送服务响应消息。

附图说明

[0006] 图1表示根据本发明的实施例的可提供直接消息传送的事务性中间件机器环境的示图。

[0007] 图2示出根据本发明的实施例的用于在事务性中间件机器环境中提供直接消息传送的示例性流程图。

[0008] 图3表示根据本发明的实施例的在事务性中间件机器环境中设置绕过桥接处理特征的示图。

[0009] 图4表示根据本发明的实施例的事务性中间件机器环境中的功能配置。

[0010] 图5是根据本发明的实施例的事务性服务器的功能框图。

[0011] 图6是根据本发明的实施例的客户机服务器的功能框图。

具体实施方式

[0012] 这里描述的是用于支持可利用具有多个处理器的快速机器的诸如Tuxedo的事务性中间件系统和高性能网络连接的系统和方法。事务性中间件系统可通过使用远程直接存储器访问 (RDMA) 协议在本地机器与远程机器之间交换消息,以通过类似于本地消息传送的方式实现较短的延时。事务性中间件机器环境包含与第一事务性机器中的服务器相关的第一消息队列,其中,服务器进行操作以通过使用第一消息队列中的队列地址被访问。事务性中间件机器环境还包括与第二事务性机器中的客户机相关的第二消息队列。客户机进行操作以通过使用第一消息队列直接向服务器发送服务请求消息,并且服务器进行操作以从第一消息队列接收服务请求消息并且通过使用第二消息队列直接向客户机发送服务响应消息。

[0013] 根据本发明的实施例,该系统包括高性能硬件的组合,例如,64位处理器技术、高性能大存储器和冗余InfiniBand和Ethernet联网,连同应用服务器或中间件环境,诸如WebLogic套件,以提供包括可被迅速提供并且可根据需要缩放的大规模并行存储器内网格的完整的Java EE应用服务器综合体。根据实施例,该系统可被部署为提供应用服务器网格、存储区域网络和InfiniBand (IB) 网络的全、半或四分之一框架或其它的配置。中间件机器软件可提供应用服务器、中间件和其它的功能,诸如,例如,WebLogic服务器、JRockit或Hotspot JVM、Oracle Linux或Solaris和Oracle VM。根据实施例,该系统可包含经由IB网络相互通信的多个计算机节点、IB交换机网关和存储节点或单元。当实现为框架配置时,框架的未使用部分可被空置或者被过滤器占据。

[0014] 根据这里称为“Sun Oracle Exalogic”或“Exalogic”的本发明的实施例,该系统是用于托管诸如Oracle Middleware SW套件或Weblogic的中间件或应用服务器软件的容易部署方案。如这里描述的那样,根据实施例,系统是包含一个或更多个服务器、存储单元、用于存储联网的IB结构和托管中间件应用需要的所有其它部件的“箱内网络”。可通过用例如Real Application Clusters和Exalogic Open存储而利用大规模并行网络结构,从而对所有类型的中间件应用递送显著的性能。该系统通过线性I/O可缩放性递送改进的性能,易于使用和管理,并且递送任务关键的可用性和可靠性。

[0015] 根据本发明的实施例,Tuxedo是使得能够构建、执行和管理高性能分布式商业应用的一组软件模块,并且已被大量的多层应用开发工具用作事务性中间件。Tuxedo是可用于管理在分布式计算环境中的分布式事务处理的中间件平台。它是得到证明的用于解锁企业陈旧 (legacy) 应用并将它们扩展到面向服务的结构、同时递送不受限制的可缩放性和基于标准的互用性的平台。

[0016] 根据本发明的实施例,诸如Oracle系统的事务性中间件系统可利用诸如Exalogic中间件机器的具有多个处理器的快速机器和诸如InfiniBand (IB) 网络的高性能网络连接。

[0017] 根据本发明的实施例,事务性中间件系统可通过使用远程直接存储器访问 (RDMA) 协议在本地机器与远程机器之间交换消息,以通过类似于本地消息传送的方式实现较短的延时。

[0018] 提供直接消息传送

[0019] 根据本发明的实施例,诸如Exalogic中间件机器环境的中间件机器环境可以是紧密耦合且在逻辑上均匀的用于诸如Tuxedo应用的事务应用的环境。当在事务性中间件机器环境中的机器之间传送消息时不存在单点瓶颈。

[0020] 图1是根据本发明的实施例的可提供直接消息传送的事务性中间件机器环境的示图。如图1所示,事务性中间件机器环境包括多个事务性中间件机器:机器A101和机器B102。各事务性中间件机器包含广告表112或113,例如,Tuxedo中的公告板(BB)。广告表可被不同的机器共享。

[0021] 如图1所示,机器B102上的客户机B可查找针对提供该客户机B需要的服务的服务器的广告表。该目标服务器可以是诸如机器B102上的服务器B的本地服务器或者诸如机器A101上的服务器A的远程服务器。

[0022] 如果服务处于本地服务器上,那么,客户机B可通过诸如机器B102上的处理间通信(IPC)队列B108的一个或更多个系统V IPC队列向目标服务器发送服务请求消息并接收服务响应消息。

[0023] 另一方面,如果服务处于诸如机器A101上的服务器A的远程服务器上,那么客户机B可使用RDMA消息收发队列直接交换消息,即使它们在物理上分布于多个机器上。RDMA队列是可发送和接收消息的库,其基于在诸如IB网络的高性能网络上运行的RDMA协议。

[0024] 如图1所示,客户机服务器B106可将服务请求消息直接置于与事务性服务器A103相关的机器A101上的RDMA队列A110中。事务性服务器A103可从RDMA队列A110得到服务请求消息并处理服务请求。然后,事务性服务器A103可直接将服务响应消息置于与客户机B相关的机器B102上的RDMA队列B111中,使得客户机B可得到服务响应。

[0025] 作为替代方案,如图1所示,可存在机器A101上的桥接处理A114和机器B102上的桥接处理B115。通过使用这些桥接处理,当机器B102上的客户机B使消息瞄准远程机器A101时,客户机B可首先向与本地桥接处理即桥接处理B115相关的系统V IPC队列B108发送服务请求消息。然后,桥接处理B115可经由网络将该消息转送到远程目标机器A101上的另一桥接处理即桥接处理A114。最后,远程目标机器A101处的桥接处理A114可将消息发送到与目标事务性服务器A103相关的系统V IPC队列A107。但是,该过程可能是耗时的,并可引入相对较长的延时。并且,桥接处理中的每一个在重负载下可能太繁忙并且变为可影响系统的总体吞吐量的单点瓶颈。

[0026] 如图1所示,事务性服务器A103可同时在不同的线程中收听系统V IPC队列A107和RDMA队列A110。诸如客户机B的消息发送方可选择较快的方式来向事务性服务器A103发送消息。一般地,RDMA队列可被用于远程消息传送,并且,系统V IPC队列可被用于本地消息传送,原因是对于本地消息传送RDMA队列可比系统V IPC队列慢。

[0027] 图2示出根据本发明的实施例的用于在事务性中间件机器环境中提供直接消息传送的示例性流程图。如图2所示,在步骤201中,第一消息队列被提供为与第一事务性机器中的事务性服务器相关。然后,在步骤202中,事务性服务器可在事务性中间件机器环境中的广告表中公布第一消息队列。最后,在步骤203中,第二事务性机器中的客户机服务器可寻找广告表中的第一消息队列,并通过使用第一消息队列建立与第一事务性机器中的事务性服务器的连接。

[0028] 通过使用RDMA协议绕过Tuxedo中的桥接处理

[0029] 根据本发明的实施例,事务性中间件系统可通过诸如Tuxedo中的MSGQ队列库的RDMA协议通过使用RDMA队列在本地机器与远程机器之间交换消息。RDMA协议可被用于跨着在高性能IB网络上运行的机器传送数据。通过使用RDMA协议,消息发送方和接收方即使当

在物理上分布于多个机器上时也可直接交换消息。

[0030] Tuxedo是具有类分布式客户机服务器结构的事务性中间件。各Tuxedo服务器可贯穿整个域在广告表中公布其MSGQ队列名。客户机服务器可在广告表中寻找目标服务器的队列名,并且如果它是第一次用于连接则建立与目标服务器的连接。然后,客户机服务器和目标服务器可根据该队列名建立连接并经由建立的连接发送消息。

[0031] 另外,Tuxedo可在域内的各机器中使用桥接处理,以跨着机器交换消息。通过使用桥接处理,瞄准远程机器的每条消息可被首先发送到与本地桥接处理相关的系统V IPC队列。然后,本地桥接处理可经由网络将该消息转送到远程目标机器上的远程桥接处理。最后,远程目标机器处的远程桥接处理可将消息发送到目标系统V IPC队列。

[0032] 根据本发明的实施例,可通过使用一系列的消息传送API以最小的代码变化实现RDMA队列库,该一系列的消息传送API是通过使用系统V样式实现的。这些消息传送API可提供与系统V IPC队列提供的功能类似的功能,例如,以优先级从队列接收消息的功能。

[0033] 图3表示根据本发明的实施例的在事务性中间件机器环境中设置绕过桥接处理特征的示图。如图3所示,在步骤301中,用户311可首先启动DAEMON处理。DAEMON处理313可以是与诸如Tuxedo应用的事务应用分开的应用。并且,在Tuxedo的例子中,对于各Tuxedo机器启动至少一个DAEMON处理。然后,在步骤302和303中,用户可分别配置应用服务器和事务应用(Tuxedo)。最后,在步骤304中,为了使用绕过桥接特征,用户可启动Tuxedo应用,并且,在步骤305中,用户允许用于在处理之间交换消息的通信。

[0034] 下表1表示用于在Tuxedo中执行DAEMON程序的不同选项。

[0035] 表1

[0036]

选项	描述
-i	Infiniband 接口的 IP 地址,其值等于 RDMADAEMONIP。
-p	DAEMON 从其接收请求的端口,其值等于 RDMADAEMONPORT。
-m	对于 DAEMON 分配的共享存储器尺寸,其值大于队列的

[0037]

	数量 * (通过使用 remotename 连接的 IP 的数量+1) *qsize
-k	将由 DAEMON 使用的共享存储器键 (key)

[0038] 例如,可通过使用以下的命令启动DAEMON程序:

[0039] Msgq_daemon-i192.168.10.4-p4040-m100000000-k9100

[0040] 另外,为了利用消息队列,对Tuxedo应用配置不同类型的参数。这些参数至少包括消息队列的参数和Tuxedo应用中的处理用于与DAEMON处理通信的参数。这些参数可被配置于诸如Tuxedo UBB文件的配置文件的MACHINES部分中。下表2表示这三个环境变量。

[0041] 表2

[0042]

UBB 参数	描述
RDMA DAEMONIP	用于与 DAEMON 通信的 IP 地址
RDMA DAEMONPORT	用于与 DAEMON 通信的端口
RDMA QSIZE	消息队列的最大尺寸，缺省值是 1Mbytes。该参数可被配置于 MACHINES 部分或 SERVERS 部分（使用服务器选项“-b”）中。其在 MACHINES 部分中的值配置所有队列的缺省值。其在 SERVERS 部分中的值仅配置由服务器产生的队列。
RDMA QENTRIES	消息队列的最大条目，缺省值是 1024。该参数可被配置于 MACHINES 部分或 SERVERS 部分（使用服务器选项“-c”）中。其在 MACHINES 部分中的值配置所有队列的缺省值。其在 SERVERS 部分中的值仅配置由服务器产生的队列。

[0043] 以下的清单1包含示例性的Tuxedo配置文件的几个部分。

[0044]

*MACHINES

slcag1906

LMID=cpu_0

APPDIR="/root/john/mnt/eniu/exalogic/cases/config/testcase"

TUXCONFIG="/root/john/mnt/eniu/exalogic/cases/config/testcase/tuxconfig"

TUXDIR="/root/john/mnt/eniu/TUX11gR1PS164rp/LC/bld"

UID=507

GID=506

RDMA DAEMONIP="192.168.10.4"

RDMA DAEMONPORT=9800

RDMA QSIZE=1048576

RDMA QENTRIES=1024

*SERVERS

simpsserv

CLOPT="-A -b 1000000 -c 1000 -j 1000000"

SRVGRP=GROUP1 SRVID=1

Simpsserv1

SRVGRP=GROUP1 SRVID=1

[0045] 清单1

[0046] 以下的清单2是Tuxedo配置文件中的RESOURCES部分的例子。

```
*RESOURCES
```

```
IPCKEY          213456
```

```
#Example:
```

```
#IPCKEY          123456
```

[0047] DOMAINID simpapp

```
MASTER          simple
```

```
MAXACCESSERS     10
```

```
MAXSERVERS       5
```

```
MAXSERVICES      10
```

```
MODEL            SHM
```

```
LDBAL            N
```

```
OPTIONS           NO_XA,EXALOGIC,RDMA
```

[0048] 清单2

[0049] 如以上的例子所示,EXALOGIC和RDMA可被添加到Tuxedo配置文件的RESOURCES部分中的项目OPTIONS中。如果OPTIONS包含RDMA和EXALOGIC,那么绕过特征被激活并且本地桥接处理被绕过。否则,绕过特征被关闭。如果需要启用RDMA选项,那么EXALOGIC选项首先被启用。在启用RESOURCES部分中的RDMA选项之后,MACHINES部分的属性“TYPE”可能不被设定,原因是,在缺省状态下,MP模式中的任何机器是用于支持RDMA特征的Exalogic机器(具有相同的类型)。

[0050] 参照图4,描述根据本发明的实施例的用于在事务性中间件机器环境提供直接消息传送的系统400。图4表示系统400的功能配置。系统400包括提供单元410、公布单元420、存储器430、消息队列寻找器440和连接建立单元450。

[0051] 提供单元被配置为提供与诸如机器B103的第一事务性机器中的事务性服务器相关的第一消息队列。公布单元420被配置为经由事务性服务器在事务性中间件机器环境中的广告表中公布第一消息队列。存储器430被配置为存储广告表。消息队列寻找器440被配置为在广告表中寻找第一消息队列。连接建立单元450被配置为通过使用第一消息队列建立与第一事务性机器中的事务性服务器的连接。消息队列寻找器440和连接建立单元450可在客户机服务器中被实现。

[0052] 根据本发明的实施例,图5和图6表示根据以上描述的本发明的原理配置的事务性服务器500和客户机服务器600的功能框图。可通过硬件、软件或硬件和软件的组合实现事务性服务器和客户机服务器的功能框图,以实施本发明的原理。本领域技术人员可以理解,在图5和图6中描述的功能框图可组合或分离成子块以实现以上描述的本发明的原理。因此,这里的描述可支持这里描述的功能块的任何可能的组合或分离或其它定义。

[0053] 如图5所示,用于事务性中间件机器环境中的第一事务性机器中的事务性服务器500包含公布单元510和建立连接单元520。公布单元510被配置为在事务性中间件机器环境中的广告表中公布与第一事务性机器中的事务性服务器500相关的第一消息队列。建立连

接单元520被配置为响应于第二事务性机器中的客户机服务器进行操作以寻找广告表中的第一消息队列,通过使用第一消息队列建立与第二事务性机器中的客户机服务器的连接。

[0054] 在一些实施例中,客户机服务器与第二事务性机器中的第二消息队列相关,并且,客户机服务器进行操作以通过使用第二消息队列接收消息。

[0055] 在一些实施例中,客户机服务器进行操作以直接通过使用第一消息队列向事务性服务器500发送服务器请求消息。

[0056] 在一些实施例中,事务性服务器500还包括第一通信单元530。第一通信单元530被配置为从第一消息队列接收服务请求消息,并且通过使用与第二事务性机器中的客户机服务器相关的第二消息队列直接向客户机服务器发送服务响应消息。

[0057] 在一些实施例中,在事务性中间件机器环境中的机器之间传送消息时,不存在单点瓶颈。

[0058] 在一些实施例中,第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。

[0059] 在一些实施例中,网络协议是用于Infiniband (IB)网络的远程直接存储器访问(RDMA)协议。

[0060] 在一些实施例中,事务性服务器500还包括第二通信单元540。第二通信单元540被配置为使用与第一事务性机器中的事务性服务器500相关的第二消息队列,以与第一事务性机器中的本地客户机通信。

[0061] 在一些实施例中,第二消息队列是与本地桥接处理相关的处理间通信(IPC)队列。

[0062] 在一些实施例中,事务性服务器400还包括收听单元550。收听单元550被配置为同时不同的线程中收听第一消息队列和第二消息队列二者,并选择快速的路由以发送消息。

[0063] 如图6所示,用于事务性中间件机器环境中的第二事务性机器中的客户机服务器600包含寻找单元610和建立连接单元620。寻找单元610被配置为在事务性中间件机器环境中的广告表中寻找与第一事务性机器中的事务性服务器相关并且通过其公布的第一消息队列。建立连接单元620被配置为通过使用第一消息队列建立与第一事务消息中的事务性服务器的连接。

[0064] 在一些实施例中,客户机服务器600还包括被配置为通过使用与第二事务性机器中的客户机服务器相关的第二消息队列接收消息。在一些实施例中,客户机服务器600还包括被配置为通过使用第一消息队列直接向事务性服务器发送服务请求消息的发送单元640。

[0065] 在一些实施例中,事务性服务器进行操作以从第一消息队列接收服务请求消息,并且通过使用与第二事务性机器中的客户机服务器600相关的第二消息队列直接向客户机服务器600发送服务响应消息。

[0066] 在一些实施例中,在事务性中间件机器环境中的机器之间传送消息时,不存在单点瓶颈。

[0067] 在一些实施例中,第一消息队列使用网络协议以用于跨着在高性能网络上运行的机器传送数据。

[0068] 在一些实施例中,网络协议是用于Infiniband (IB)网络的远程直接存储器访问

(RDMA) 协议。

[0069] 在一些实施例中,事务性服务器使用与第一事务性机器中的事务性服务器相关的第二消息队列,以与第一事务性机器中的本地客户机通信。

[0070] 在一些实施例中,第二消息队列是与本地桥接处理相关的处理间通信(IPC)队列。

[0071] 在一些实施例中,事务性服务器同时在不同的线程中收听第一消息队列和第二消息队列二者,并选择快速的路由来发送消息。

[0072] 可以常规的方式通过使用一个或更多个常规的通用或专用数字计算机、计算装置、机器或根据本公开的教导被编程的包含一个或更多个处理器、存储器和/或计算机可读存储介质的微处理器,实现本发明。基于本公开的教导,本领域技术人员可很容易地准备适当的软件代码,这对于本领域技术人员来说是十分明显的。

[0073] 在一些实施例中,本发明包括作为上面/其中存储可用于将计算机编程以执行本发明的处理中的任一个的指令的存储介质或计算机可读介质(介质)的计算机程序产品。存储介质可包含但不限于包括软盘、光盘、DVD、CD-ROM、微驱动和磁光盘的任意类型的盘、ROM、RAM、EPROM、EEPROM、DRAM、VRAM、快擦写存储器装置、磁卡或光卡、纳米系统(包含分子存储器IC)或适于存储指令和/或数据的任意类型的介质或装置。

[0074] 出于解释和描述的目的提供了本发明的以上的描述。这不是详尽的,或者要将本发明限于公开的确切形式。许多修改和变动对于本领域技术人员是十分明显的。变动可包含这里公开的两个或更多个特征的组合。为了最佳地解释本发明的原理及其实际应用,选择和描述了本发明,由此使得本领域技术人员能够理解本发明的各种实施例和适于设想的实际用途的各种修改。本发明的范围要由以下的权利要求和它们的等同限定。

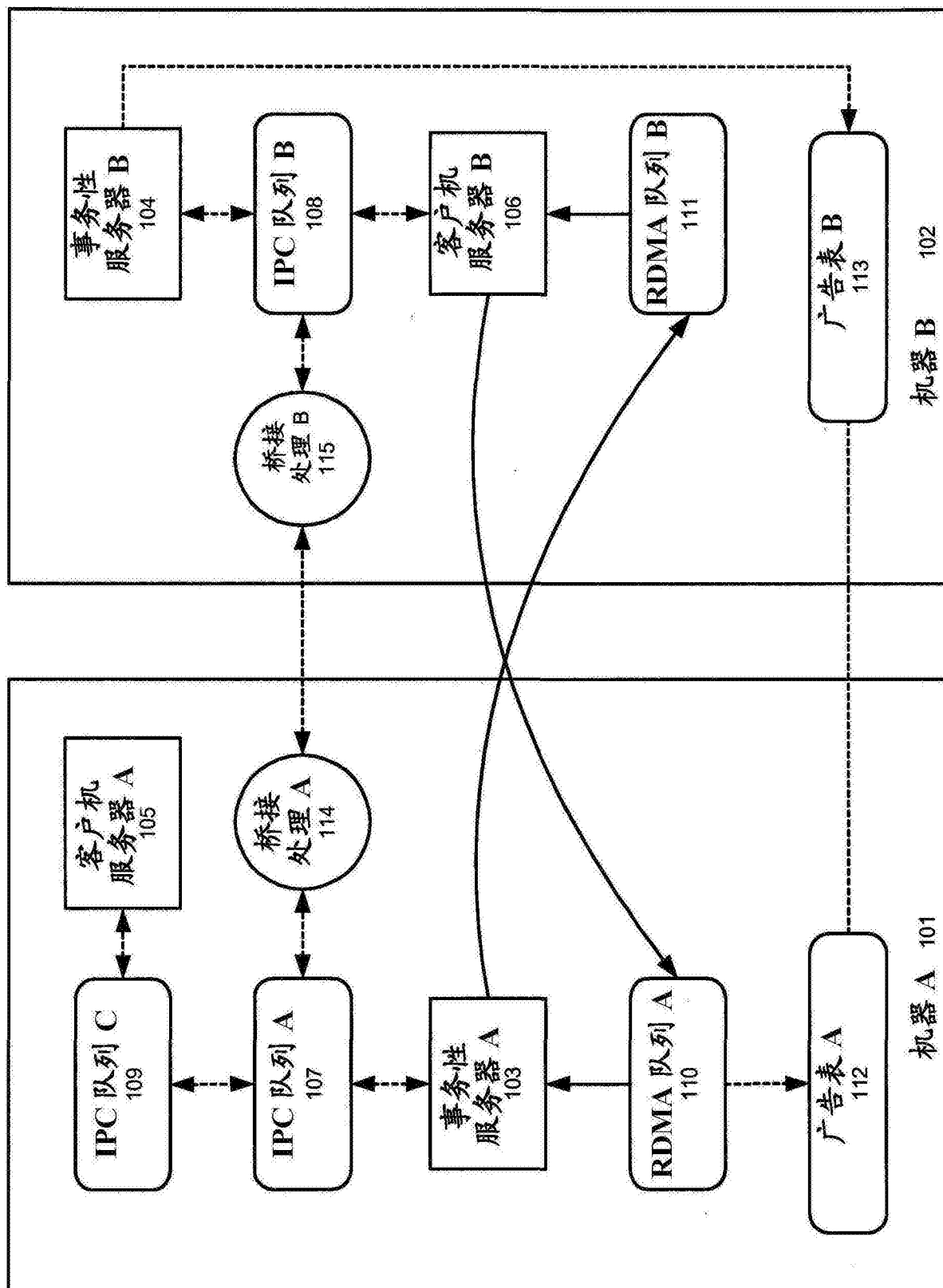


图1

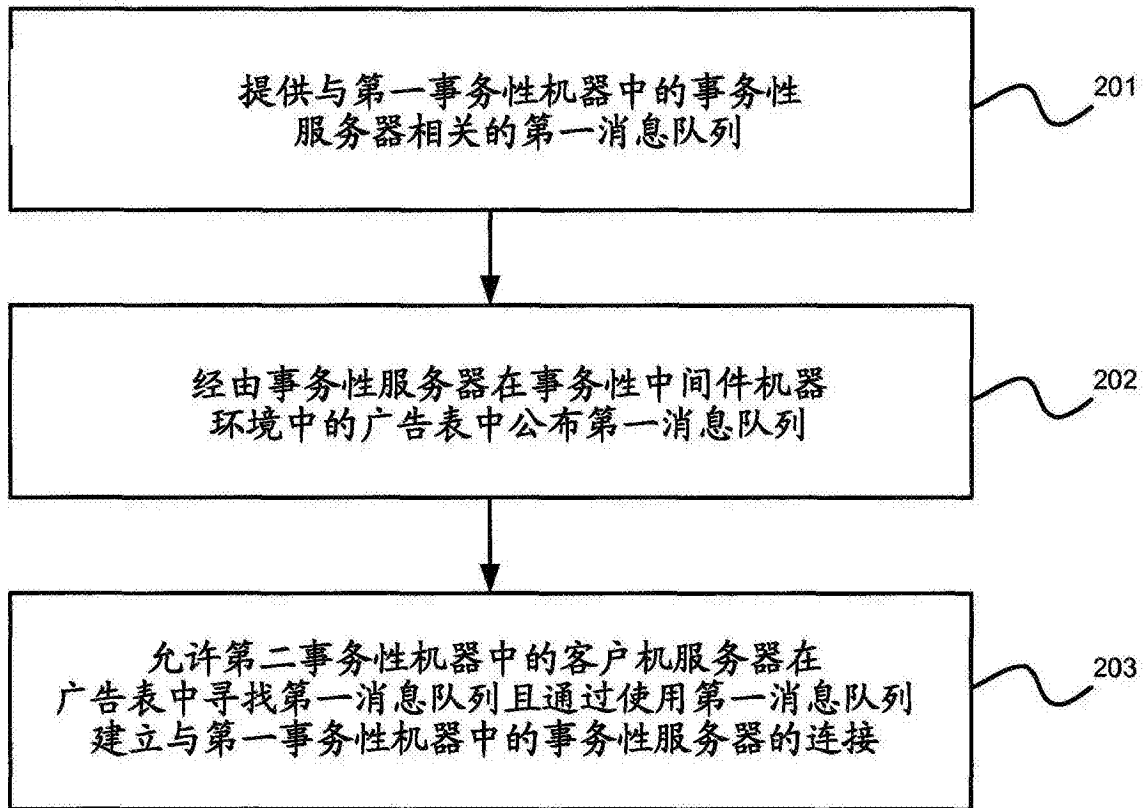


图2

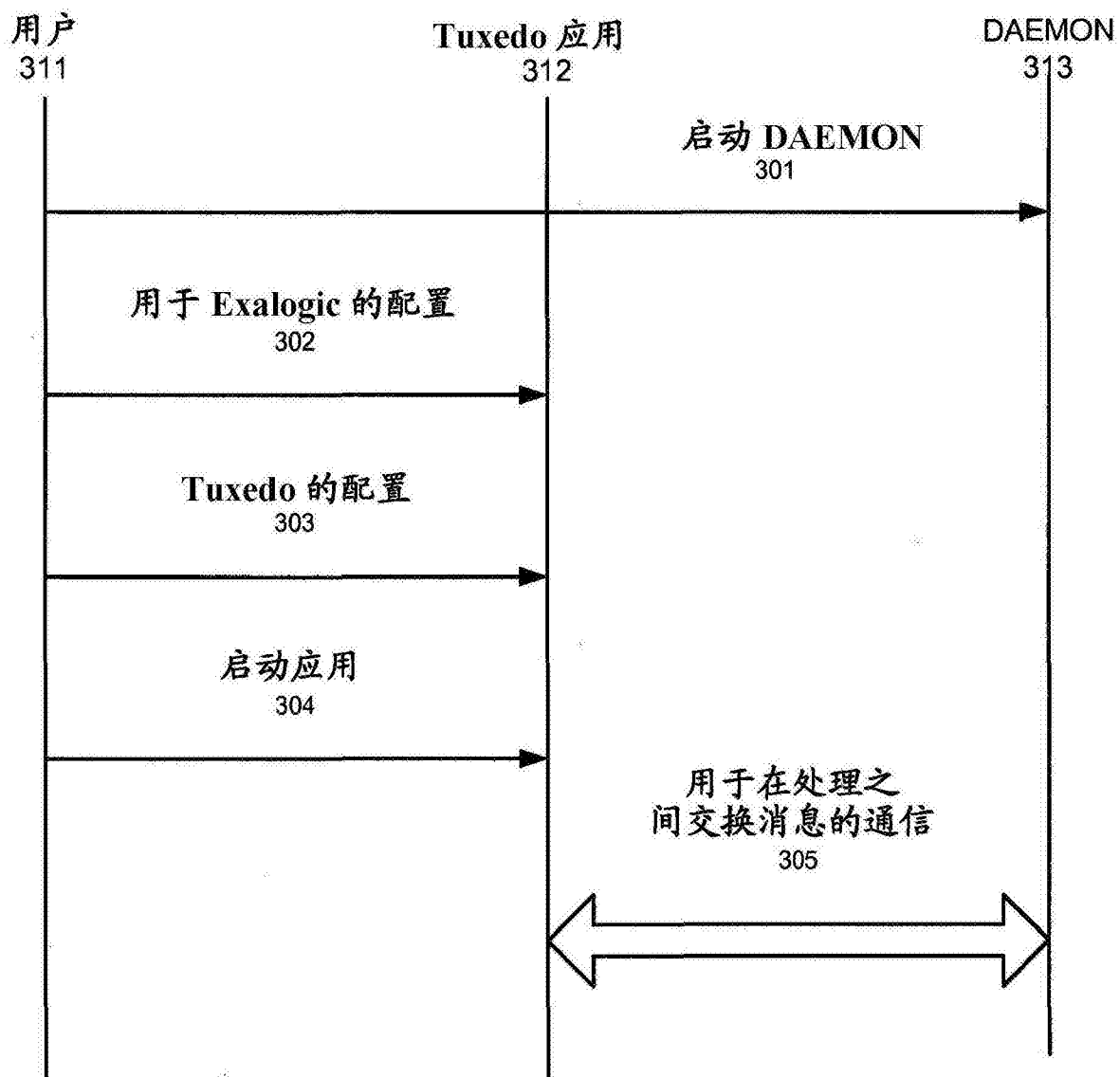


图3

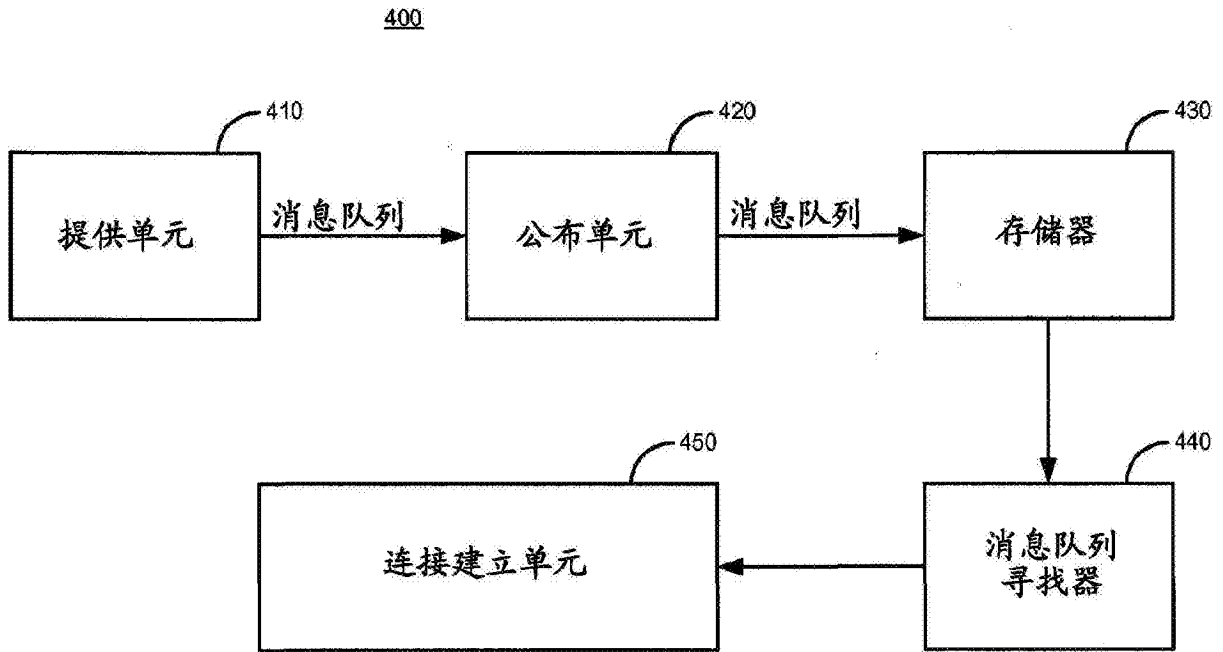


图4

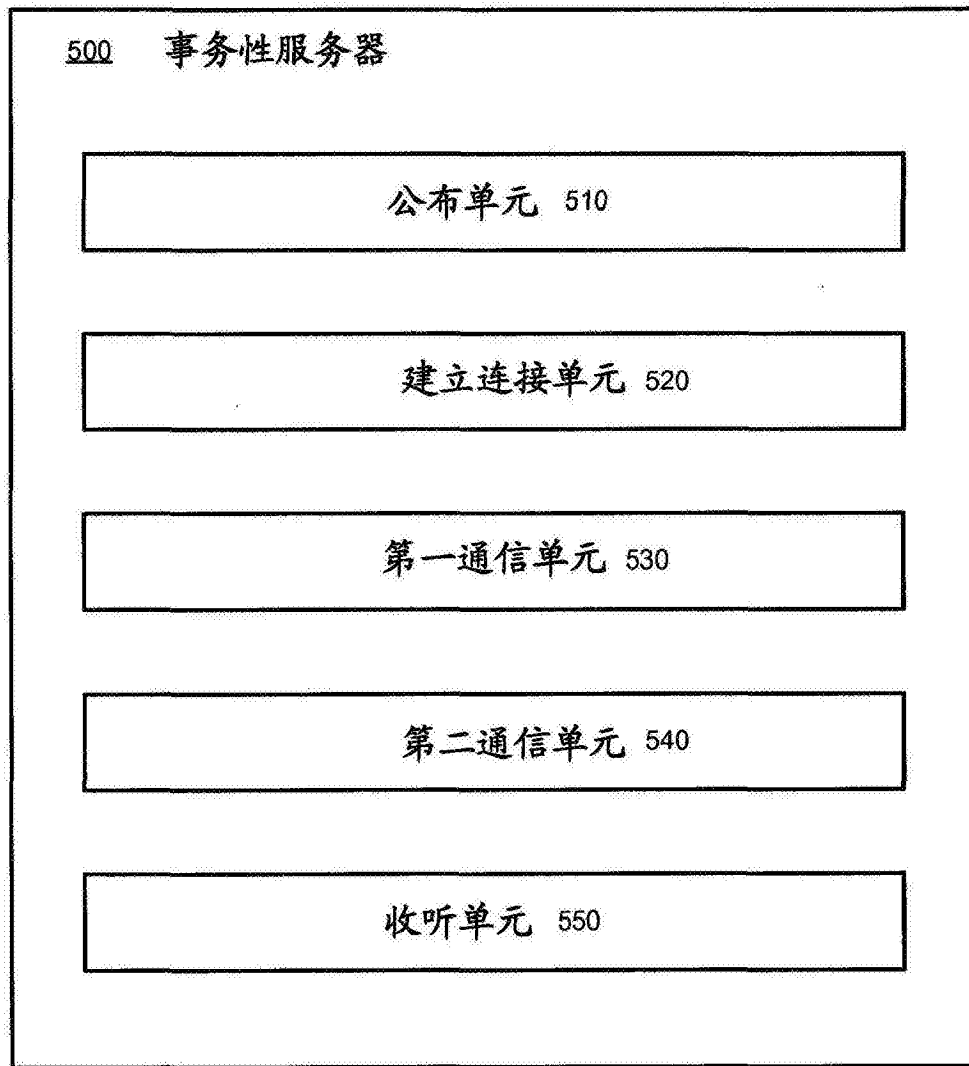


图5

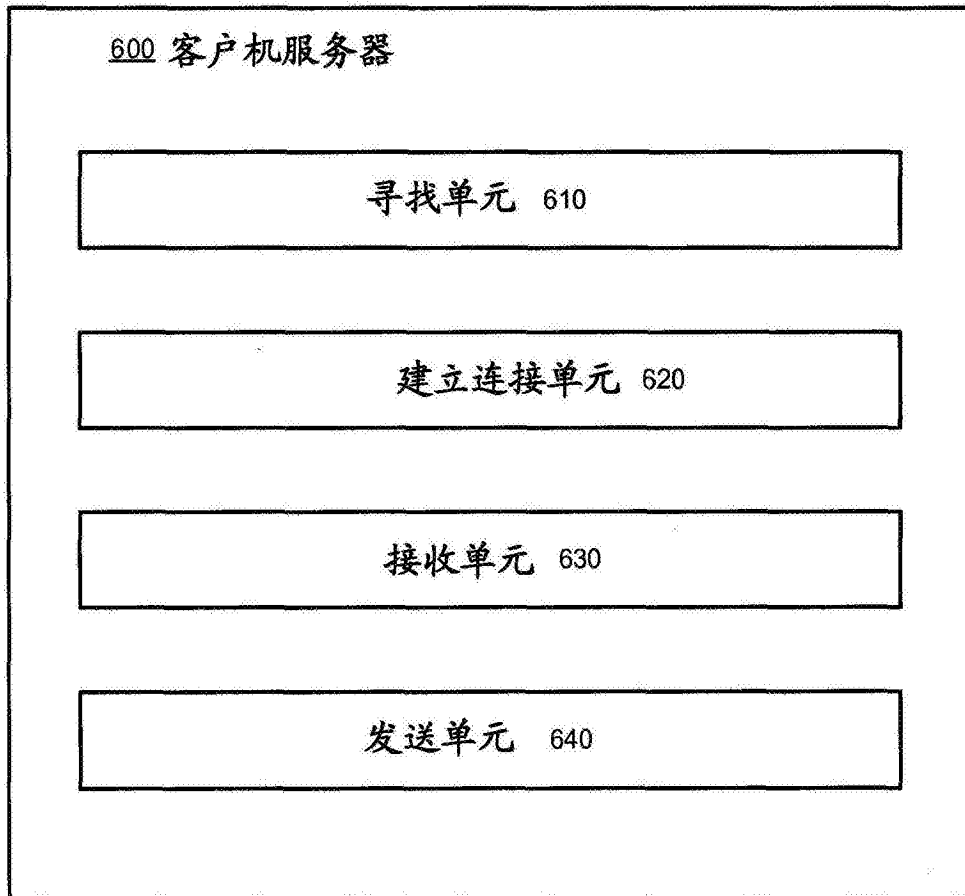


图6