(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 101447989 B
(45)授权公告日 2013.05.22

(21)申请号 200810178177.5
(22)申请日 2008.11.25
(30)优先权数据
07291417.9 2007.11.28 EP
(73)专利权人 阿尔卡特朗讯公司
地址 法国巴黎
(72)发明人 G·克里斯塔洛 N·詹森斯
J·M·C·莫雷尔
(74)专利代理机构 北京市中咨律师事务所 11247
代理人 杨晓光 于静
(51)Int.Cl.
H04L 29/06(2006.01)
H04L 12/70(2013.01)

(56)对比文件
US 2007/0091874 A1,2007.04.26,
CN 1722664 A,2006.01.18,

审查员 李晓利

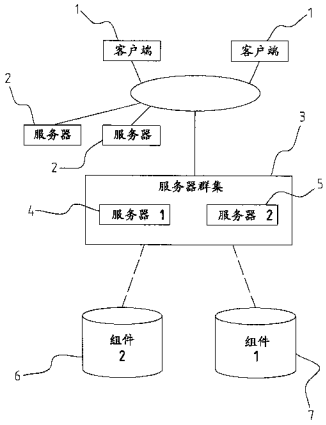
权利要求书1页 说明书4页 附图4页

(54)发明名称

用于改进的高可用性组件实现的系统和方法

(57)摘要

本发明涉及一种用于通过传输连接上的会话进行高可用性处理的、用在具有至少两个节点的群集中的计算机系统和方法。所述系统包括协议组件；具有至少两个节点的群集，其被安排用于运行所述协议组件；和服务器，其被安排用于维持传输连接上的与所述群集中的一个节点的协议会话。所述群集被安排用于维持所述至少两个节点中的每一个上的所述协议组件的一个实例，以使得至少两个实例是激活的；所述服务器被安排用于同时维持与每个实例的协议会话。



CN 101447989 B

1. 一种用于进行高可用性处理的计算机系统，包括：

- 信令协议组件 (6、7)，其中，所述信令协议组件使用在传输连接上被承载的信令协议会话，

- 具有至少两个节点 (4、5) 的群集 (3)，其被安排用于运行所述信令协议组件，和

- 服务器，其被安排用于维持所述传输连接上的与所述群集中一个节点的信令协议会话，

其特征在于，

所述群集被安排用于维持所述至少两个节点中每一个上的所述信令协议组件的一个实例 (31、32、33)，以使得同一个信令协议组件的至少两个实例是激活的；并且所述服务器被安排用于同时维持与所述至少两个实例中的每个实例的信令协议会话 (34、35、36)。

2. 根据权利要求 1 所述的计算机系统，其特征在于，所述系统还被安排用于在所述至少两个实例中的一个实例失败的情况下将信令协议会话上发送的业务重新分配给利用所述至少两个实例中的另一个实例来维持的另一个信令协议会话。

3. 根据权利要求 1 所述的计算机系统，其特征在于，所述传输连接上的所述信令协议会话使用下列协议中的一个：XMPP、DIAMETER、或任何其他适合用在传输连接上的信令协议。

4. 根据权利要求 1 所述的计算机系统，其特征在于，所述服务器被配备以用于按照预定算法而将协议分组分配给多个实例中的每一个的模块 (42)。

5. 根据权利要求 4 所述的计算机系统，其特征在于，所述模块能够访问与所述信令协议组件的激活信令协议会话的列表 (41)，并且所述服务器被安排用于递送协议分组至所述模块。

6. 一种用于通过传输连接上的会话来进行高可用性处理的方法，其用在具有至少两个节点 (4、5) 的群集 (3) 中，所述群集被安排用于运行信令协议组件 (6、7)，其中，所述信令协议组件使用在传输连接上被承载的信令协议会话，

其特征在于，

- 信令协议组件分布于所述至少两个节点中的至少两个之上，同一个信令协议组件的一个实例 (31、32、33) 运行在每个节点上，以及

- 服务器向所述信令协议组件在一个或多个传输连接上打开至少两个信令协议会话 (34、35、36)，以使得与每个实例的信令协议会话同时被维持。

7. 根据权利要求 6 所述的方法，其特征在于，基于预定算法而将协议分组分配给所述至少两个信令协议会话。

8. 根据权利要求 7 所述的方法，其特征在于，所述算法是下列算法中的一个：轮询方法、散列方法、随机方法、固定分配方法、基于会话标识符的方法、或所述算法中的一个或多个的组合。

9. 根据权利要求 7 所述的方法，其特征在于，所述服务器测量所述信令协议组件的每个实例上的处理负载，并且所述协议分组的分配是基于处理负载测量结果的。

10. 根据权利要求 6 所述的方法，其特征在于，在所述至少两个信令协议会话中的一个信令协议会话上被发送给至少两个实例中的失败实例的协议分组，被重新分配给所述至少两个信令协议会话中的另一个信令协议会话。

用于改进的高可用性组件实现的系统和方法

技术领域

[0001] 本发明涉及用于高可用性处理的系统，该系统包括协议组件、具有至少两个节点的群集和用于维持传输连接（例如 TCP 连接）上的与群集中的节点的协议会话的服务器，所述群集用于运行协议组件。本发明还涉及用于通过传输连接上的协议会话而进行高可用性处理的方法。

背景技术

[0002] 由于需要较高可靠性和较少停止时间的应用的不断增长，对于容错和高可用性处理系统的关注也不断增加。一种已知的解决方案是互连处理器群集组的网络。群集是基于硬件冗余的原理并且通常包括为增加可用性而用作单个系统的多个节点。

[0003] 高可用性群集的最常见的大小是两个节点，例如主节点和备用节点。目前，当销售商声称组件实现高度可用时，他们是指该组件运行在群集环境中并且该组件的一个实例维持与服务器的协议会话。主组件通常由备用组件支持，只要主组件启动并运行，备用组件就停止工作，但是当主组件出故障时，备用组件就变为可用。换句话说，当前的“高可用性”结构包括组件的几个实例（主组件和一个或多个备用组件）但是在主组件和服务器之间仅有一个协议会话。如果主组件出故障，则需要在服务器与备用组件之间开始新的协议会话。因此，服务器将察觉该服务在一段时间内的不可用性。实际上，组件上的处理负担及它出故障的可能性并没有减少。当然，由备用组件协助该组件并没有减少组件上的处理负担。具有或不具有备用组件不会影响组件变为不可用的可能性。此外，备用组件上的处理负担通常接近于主组件上的处理负担，以使得方案的总处理负担几乎是单个组件的处理负担的两倍。

发明内容

[0004] 本发明的目的是提供一种基于传输连接的高可用性组件实现，其降低了组件上的处理负担并且可被配置成减少可用资源上的负载，因而降低了组件的实例毁坏和 / 或出故障的可能性。

[0005] 为此，根据本发明的系统的区别在于，群集被安排用来维持所述至少两个节点中每一个上的一个协议组件实例，以使得至少两个实例是激活的，并且所述服务器被安排用来同时维持与每个实例的协议会话。现有技术问题被本发明方法解决，其特征在于，在至少两个节点上分配协议组件，同一协议组件的一个实例运行在每个节点上，并且利用该协议组件而打开一个或多个传输连接上的至少两个协议会话。

[0006] 本发明的系统和方法的优点在于，由于所述协议组件分布于至少两个节点上，所述协议可以通过降低协议组件使得处理负担而受益于真正的负载平衡机制，并且由于至少两个协议会话保持激活而使得至这些节点的业务可以连续地被适当调整。

[0007] 应当指出，术语“协议会话”应当以宽泛的含义来解释，其是指传输连接上的任何持续时间长的连接，例如 TCP（传输控制协议）连接上的 XMPP 会话，但是也包括例如 TCP 或

SCTP（流控制传输协议）连接上的 Diameter 协议连接。原则上，可以设想采用任何其他使用协议会话并且适于使用传输连接的协议。

[0008] 从属权利要求中公开了有利的实施例。

[0009] 优选地，所述系统被安排用来当实例失败时将协议会话上发送的业务重新分配给另一个激活的协议会话。这可以确保永久的可用性。

[0010] 根据优选实施例，所述服务器配备有用于根据预定算法而将协议分组可选地分配给多个实例中的每一个的模块。所述模块通常能够访问具有协议组件的激活协议会话的列表，并且所述服务器被安排用来将协议分组递送给按照预定算法发送这些分组的模块。本领域技术人员应当理解，可以使用许多不同的算法，例如轮询方法、散列方法、随机方法、固定分配方法、基于会话标识符的方法，所述方法中一个或多个的组合，等等。将参考图 4 进一步说明这些方法。

[0011] 根据本发明方法的另一方面，服务器测量协议组件的每个实例上的处理负载，其中协议分组的分配是基于处理负载测量结果的。以该方式，可以在实例之间执行相当高效的负载平衡。

附图说明

[0012] 附图用于说明本发明的优选的、非限制性示例实施例。参考附图，通过阅读下面的详细描述，本发明的上述及其他有利特征和目的将变得更加明显并且能更好地理解本发明，其中：

[0013] - 图 1 示出了典型的计算机网络的示例，其中可以实现本发明的系统和方法；

[0014] - 图 2(A)-(B) 概略性地示出了现有技术的高可用性实现；

[0015] - 图 3(A)-(B) 概略性地示出了本发明的系统和方法的实施例；

[0016] - 图 4 概略性地示出了用于本发明的系统和方法的 XMPP 服务器。

具体实施方式

[0017] 图 1 示出了典型的计算机系统，其具有若干客户端 1、若干服务器 2 和包括第一服务器 4 和第二服务器 5 的服务器群集 3。注意，服务器群集 3 通常包括可运行于同一机器上或可位于不同机器上的若干服务器实例或过程。服务器群集的服务器实例被安排用于运行若干组件 6、7。

[0018] 图 2 说明了由具有所谓的高可用性 XMPP 组件实现的 XMPP（可扩展消息处理现场协议）组件销售商使用的现有技术方法。XMPP 是一种用于传输 XML 元素以交换消息并且几乎实时地呈现信息的协议。IETF 的 XMPP 工作组进一步将 Jabber 协议适配为 IETF 许可的即时消息传送 (IM) 和现场技术。所撰写的协议是可从 httD://www.ietf.org/rfc/rfc3920.txt 获得的 RFC3920(XMPP 核心）和可从 http://www.ietf.org/rfc/rfc3921.txt 获得的 RFC3921(XMPP 核心的 IM 和现场扩展），所述 RFC 文本在这里被引入作为参考。此外，Jabber 团体管理 jabber 扩展协议 (XEP)。

[0019] XMPP 使得可信组件能够连接到 XMPP 服务器，其中 XMPP 服务器和 XMPP 组件维持彼此之间的一个或几个 XMPP 会话。这种会话是建立在传输连接上的，特别是 TCP 连接。消息会话是作为 XML 流而在 TCP 连接上被承载的。

[0020] 如图 2(A) 所示，现有技术的高可用性实现在于使用备用 XMPP 组件 20 来在主 XMPP 组件 21 失败的情况下接管工作。为了运行 XMPP 组件，XMPP 服务器 22 仅维持与主组件 21 的单个会话。只要主组件 21 启动并运行，备用组件 20 就离线工作。在这种情况下，备用组件复制所有需要的来自主组件 21 的实时信息和 / 或配置，以使得在主组件变为不可用时，备用组件可以立即接管，如图 2(B) 所示。然而，由于缺乏存在的 XMPP 会话并且需要打开 XMPP 服务器与备用组件之间的新会话，用户将获知服务在一段时期内不可用。

[0021] 本发明的主要思想是在提供备用组件之后，一方面在服务器和备用组件之间而另一方面在所述组件和备用组件之间提供备用协议会话。

[0022] 图 3(A) 和 3(B) 说明了所述概念的一种可能的实施例。在这个示例中使用 XMPP 协议。然而，本领域技术人员应当理解，所示出的系统和方法也可以用通常使用 TCP 的传输连接上所承载的其他协议来执行。XMPP 的可选方案的一个例子是 Diameter。Diameter 基础协议旨在为例如网络接入或 IP 移动性的应用提供认证、授权和计费 (AAA) 框架，并且在可从 http://www.ietf.org/rfc/rfc3588.txt 获得的 RFC3588 中被定义。

[0023] 在图 3(A) 的示例中，XMPP 组件分布于三个节点的群集上。在每个节点上存在同一 XMPP 组件的一个实例 31、32、33，并且每个实例 31、32、33 维持与服务器 30 的单个 XMPP 会话 34、35、36。所有这些实例 31-33 和它们相应的会话 34-36 是同时激活的，以使得业务可以在所有 XMPP 会话上被高效地分割。换言之，同一 XMPP 组件的总处理负担可以均匀分布于在群集中为激活的不同实例上。在这种均匀分布的情况下，n 个节点的群集的每个实例的处理负担近似地等于除以因子 n。

[0024] 然而，本领域技术人员应当理解，在一些系统中采用非均匀分配可能是优选的。这例如是当一个组件具有比另一组件更高的可用处理能力的情况。根据一种可能的变型，服务器可以被安排用于测量每个组件的负载，并且业务的分割可以是基于这种负载测量的。这将在下面参考图 4 进一步讨论。

[0025] 如果实例中的一个失败，例如图 3(B) 中的实例 3，会话 3 所发送到业务自动被重新分配给会话 1 和 / 或会话 2。

[0026] 朝向在群集中同时为激活的不同协议会话的业务分配以及失败情况下的自动重新分配可以由服务器中提供的专用模块来完成。现在将参考图 4 说明这种协议会话分配模块的实施例，其也称为容错 (FT) 模块。

[0027] 根据 FT 模块 42 的这个实施例被实现在 XMPP 服务器 40 中并且负责确定如何在可用 XMPP 会话的集合之中分割去往 XMPP 组件的业务，所述会话通常在一个或几个 TCP 连接之上打开。

[0028] 当 XMPP 会话集合与同一组件的会话开始时，XMPP 服务器 40 将通知所建立的 XMPP 会话的列表 41 的 FT 模块，所建立的 XMPP 会话通常用会话 ID 来标识。当分组应当被转发给 XMPP 组件时，FT 模块将判定使用哪个 XMPP 会话并且相应地发送所述分组。不同的算法可以被使用，例如 ：

[0029] - 轮询方法，其中依次使用每个 XMPP 会话。FT 模块记住最后一个会话 ID（或其相关变型）并且将下一分组发送给由下一会话 ID 标识的下一会话。这种算法在存在许多组件的情况下是有用的并且具有实施简单的优点。

[0030] - 散列方法 ：FT 模块首先通过在 XMPP 分组中的标识了“流”（例如与“线程 ID”（若

有的话）有关的“去往”和“来自”属性）的一些域上执行散列（例如 CRC16）来选择关键字。为每个会话 ID 分配关键字空间中的唯一区域。FT 模块使用关键字来判定需要向其发送分组的会话 ID。这种算法在专用组件需要被识别出的情况下特别合适，但是更加复杂。

[0031] - 散列方法和轮询方法或任何其他简单算法的组合：如果散列方法返回可能组件的列表，则另一方法应当被用来进行判定，这可以例如是轮询方法或下面列出的其他方法之一。

[0032] - 基于负载的方法：服务器可以被安排用于获得与每个组件上的处理负载有关的信息。分组然后可以被发送给具有最低负载的组件的会话。

[0033] - “随机”方法，其随机地发送分组至“任何”组件。

[0034] - “总是相同”方法，其中，总是在相同的会话上发送某些分组。

[0035] - 基于标识符的方法，例如将分组发送至具有最低标识符的组件，等。

[0036] 尽管上面已经结合特定实施例阐述了本发明的原理，然而应当清楚地理解，该描述只是作为示例而不是对由所附权利要求限定的保护范围的限制。

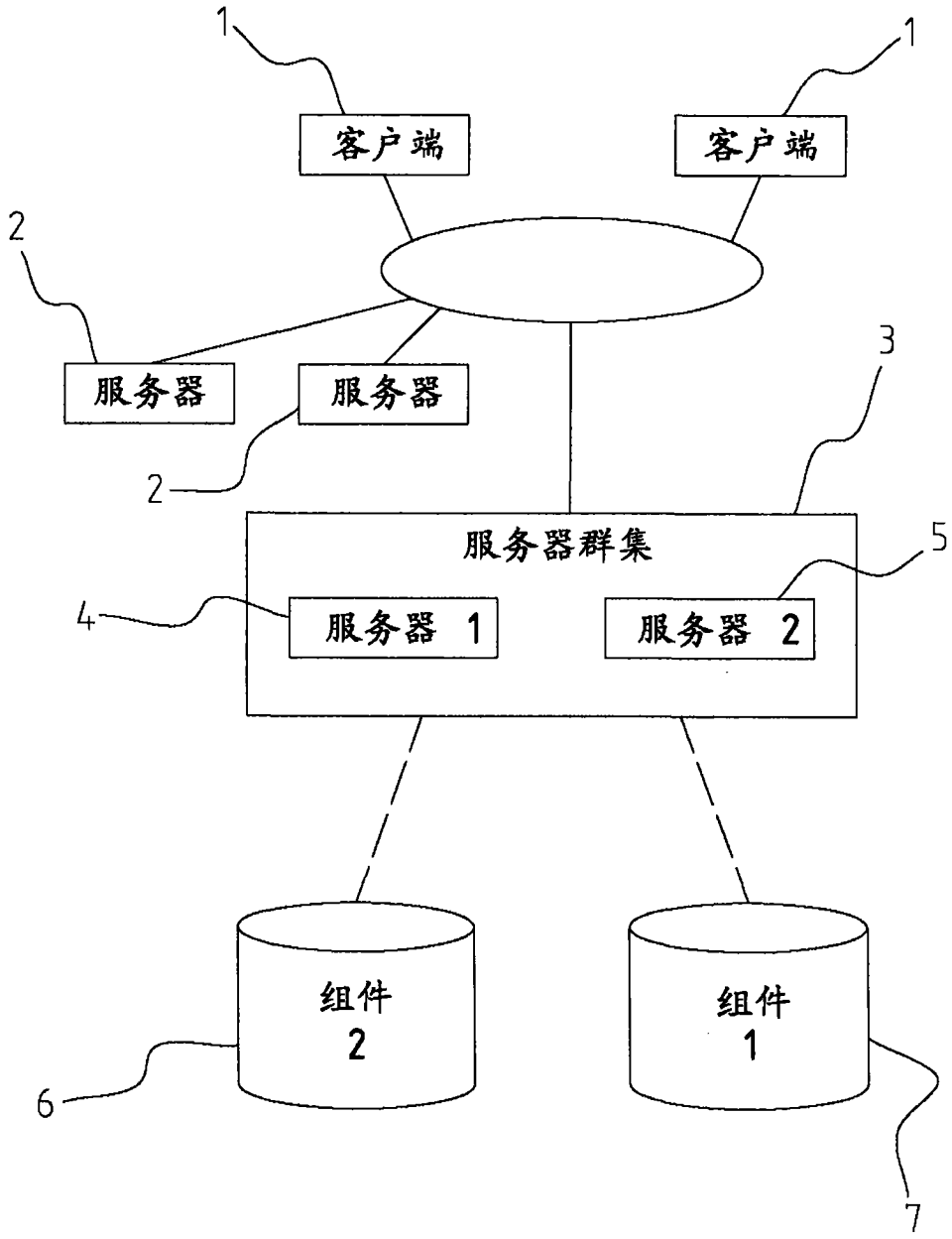


图 1

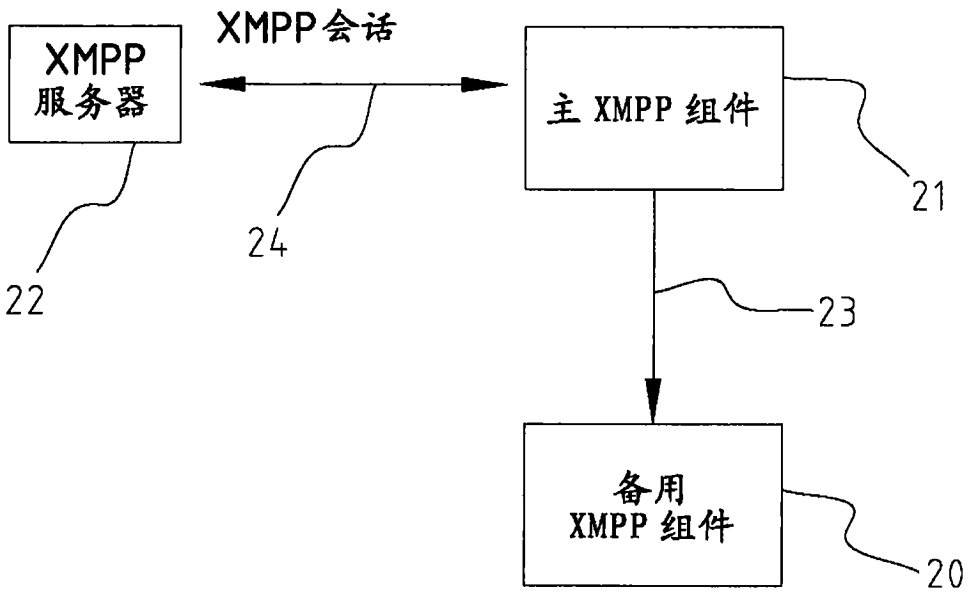

XMPP
服务器
XMPP会话
主 XMPP 组件
21
22
24
23
备用
XMPP 组件
20

图 2A

现有技术

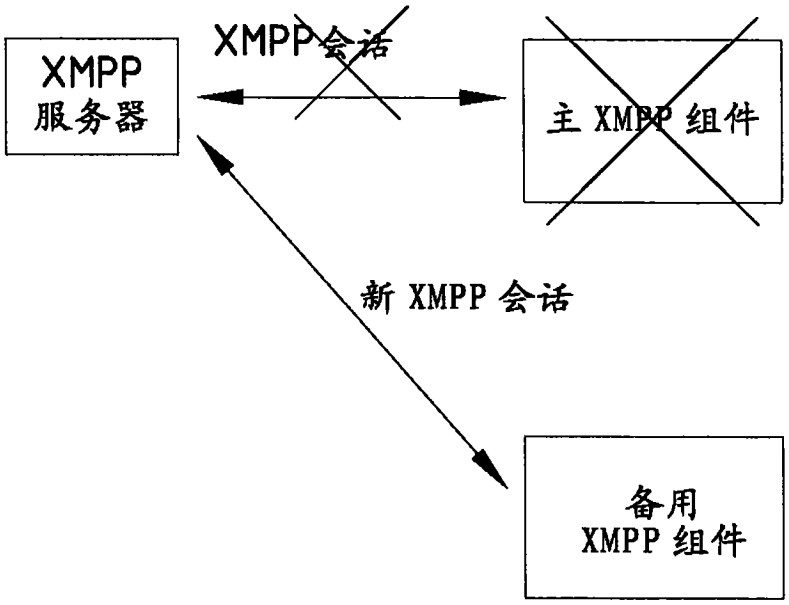

XMPP
服务器
XMPP会话
主 XMPP 组件
新 XMPP 会话
备用
XMPP 组件

图 2B

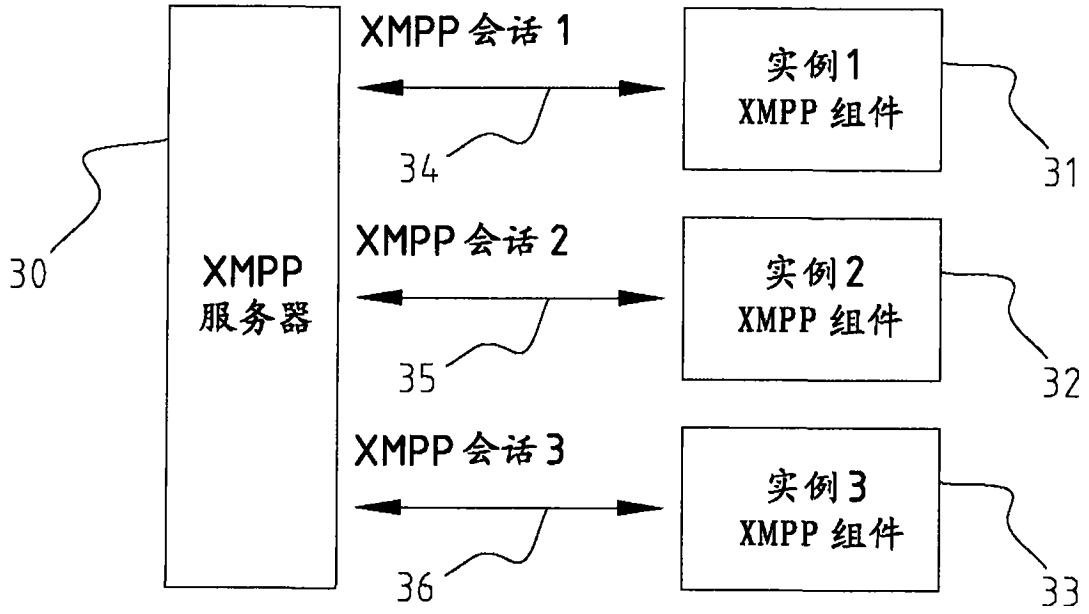

XMPP 会话 1
XMPP 会话 2
XMPP 会话 3
XMPP
服务器
实例 1
XMPP 组件
实例 2
XMPP 组件
实例 3
XMPP 组件
30
31
32
33
34
35
36

图 3A

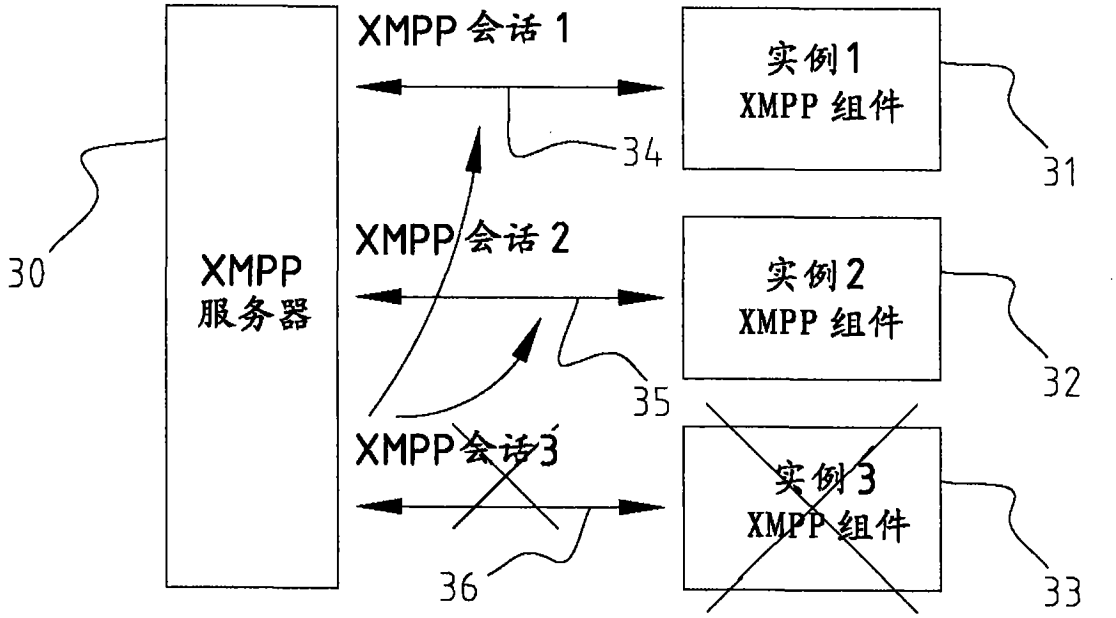

XMPP 会话 1
XMPP 会话 2
XMPP 会话 3
XMPP
服务器
实例 1
XMPP 组件
实例 2
XMPP 组件
实例 3
XMPP 组件
30
31
32
33
34
35
36

图 3B

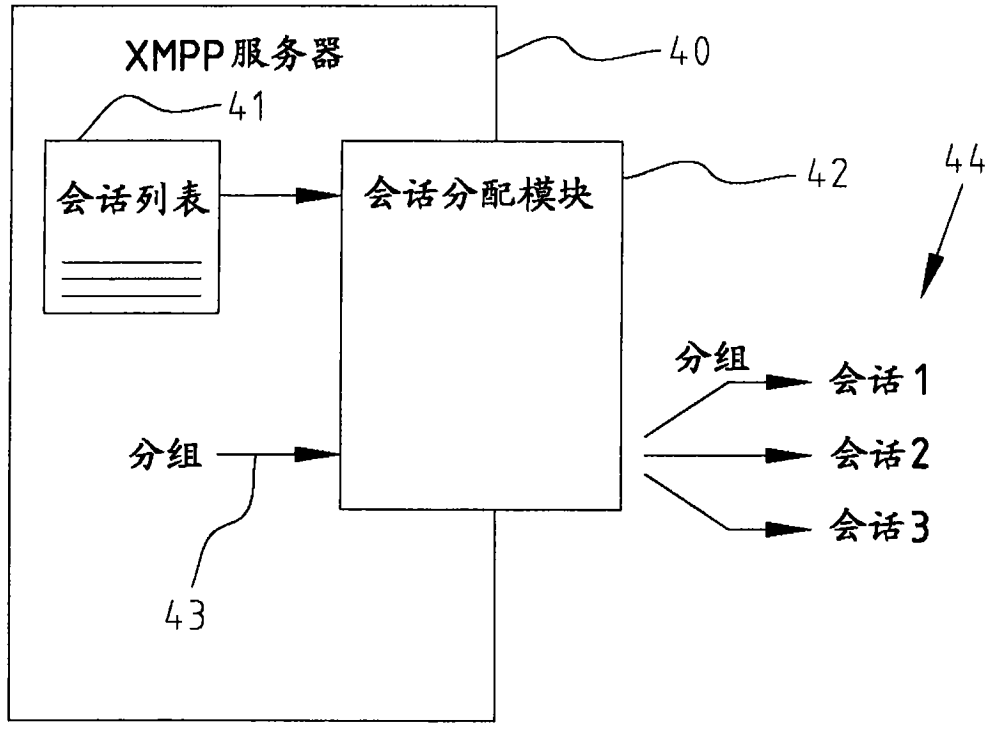


图 4