



(12) 发明专利

(10) 授权公告号 CN 105830031 B

(45) 授权公告日 2020.12.11

(21) 申请号 201380081695.4

(22) 申请日 2013.12.18

(65) 同一申请的已公布的文献号
申请公布号 CN 105830031 A

(43) 申请公布日 2016.08.03

(85) PCT国际申请进入国家阶段日
2016.06.16

(86) PCT国际申请的申请数据
PCT/SE2013/051550 2013.12.18

(87) PCT国际申请的公布数据
W02015/094039 EN 2015.06.25

(73) 专利权人 瑞典爱立信有限公司
地址 瑞典斯德哥尔摩

(72) 发明人 托马斯·梅克林 托米·罗特

米连科·奥普塞尼查

约尼·梅恩佩

(74) 专利代理机构 北京德琦知识产权代理有限公司 11018

代理人 严芬 宋志强

(51) Int.Cl.
G06F 9/50 (2006.01)

(56) 对比文件
CN 103442034 A, 2013.12.11
CN 102571686 A, 2012.07.11
CN 101087403 A, 2007.12.12
US 2011307899 A1, 2011.12.15

审查员 张旭光

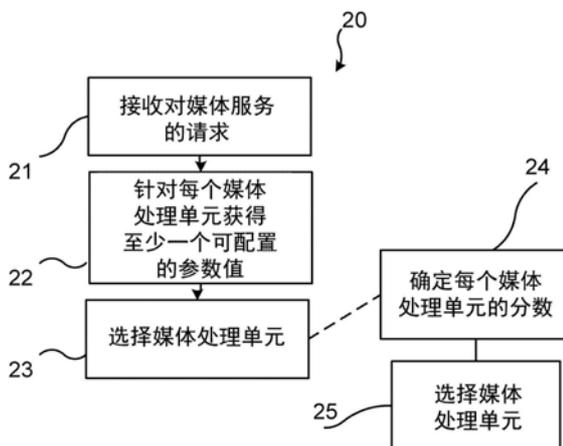
权利要求书4页 说明书15页 附图6页

(54) 发明名称

用于选择媒体处理单元的方法和网络节点

(57) 摘要

本公开涉及在分布式云2的网络节点5中执行的用于选择媒体处理单元3、3A、3B、3D; 4、4A、4B、4D; 10、11、12的方法20。分布式云2包括两个或更多个能够被配置为处理媒体服务所需要的媒体处理的处理单元3、3A、3B、3D; 4、4A、4B、4D; 10、11、12。方法20包括: 从通信设备6、6A、6B、6C、6D接收21对媒体服务的请求; 针对每个媒体处理单元3、3A、3B、3D; 4、4A、4B、4D; 10、11、12获得与媒体服务的处理有关的参数的至少一个可配置参数值; 以及, 基于该至少一个参数值来选择23用于处理通信设备6、6A、6B、6C、6D的所请求的媒体服务的媒体处理单元3、3A、3B、3D; 4、4A、4B、4D; 10、11、12。本公开还涉及对应的网络节点、计算机程序和计算机程序产品。



1. 一种在分布式云 (2) 的控制器节点 (5) 中执行的用于选择媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的方法 (20), 所述分布式云 (2) 包括能够被配置为处理相应媒体服务实例中的媒体服务所需要的媒体处理的两个或更多个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 每个媒体处理单元包括数据中心 (4、4A、4B、4D、10、11、12), 所述方法 (20) 包括:

- 从通信设备 (6、6A、6B、6C、6D) 接收 (21) 对媒体服务的请求,
- 针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 获得 (22) 与所述媒体服务的处理有关的参数的至少一个可配置参数值, 以及
- 基于所述至少一个参数值来选择 (23) 用于处理所述通信设备 (6、6A、6B、6C、6D) 所请求的媒体服务的媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 其中, 所述选择包括:
 - i. 针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 通过对相应媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的所有参数值求和来确定 (24) 分数, 相应的和构成每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的相应分数,
 - ii. 选择 (25) 具有最高分数的第一数据中心 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12),
 - iii. 确定在所选择的第一数据中心 (4、4A、4B、4D; 10、11、12) 中创建了最大数量的媒体服务器实例以及确定所有的媒体服务器实例满负荷,
 - iv. 选择具有第二高分数的第二数据中心 (4、4A、4B、4D; 10、11、12), 以及
 - v. 将所述通信设备 (6、6A、6B、6C、6D) 连接到所选择的第二数据中心 (4、4A、4B、4D; 10、11、12) 的媒体服务器实例, 其中, 所述媒体服务器实例被配置为处理媒体处理需要。

2. 根据权利要求1所述的方法 (20), 其中, 所述针对每个处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 确定 (24) 所述分数包括: 通过将具体媒体处理单元的相应参数值除以所有媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的所有对应参数值之和来将每个参数值归一化。

3. 根据权利要求1或2所述的方法 (20), 其中, 针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 获得 (22) 至少一个可配置参数值还包括: 针对每个参数获得与该参数的重要性相对应的权重, 以及所述确定 (24) 还基于所述权重。

4. 根据权利要求1所述的方法 (20), 所述方法 (20) 包括:

- 选择所选择的第二数据中心 (4、4A、4B、4D; 10、11、12) 的现有媒体服务器实例, 或者在所选择的第二数据中心 (4、4A、4B、4D; 10、11、12) 中创建新的媒体服务器实例, 以及
- 将所述通信设备 (6、6A、6B、6C、6D) 连接到所选择的现有媒体服务器实例或连接到创建的新媒体服务器实例。

5. 根据权利要求4所述的方法 (20), 其中, 对现有媒体服务器实例的所述选择基于以下各项中的一项或多项: 媒体服务器实例的负荷、媒体服务器实例的能力、媒体服务器群中的媒体服务器实例的数量、媒体服务器实例所服务的现有会话、其上正在运行媒体服务器实例的硬件的能力、连接性、带宽和操作状态。

6. 根据权利要求1或2所述的方法 (20), 其中, 所述方法包括: 通过在确定现有媒体服务器实例的负荷高于第一阈值时创建新的媒体服务器实例, 并且在确定现有媒体服务器实例的负荷低于第二阈值时移除媒体服务器实例, 将所述通信设备 (6、6A、6B、6C、6D) 连接到所述第二数据中心 (4、4A、4B、4D; 10、11、12) 的媒体服务实例, 并且基于业务负荷来适配媒体

服务器实例的数量。

7. 根据权利要求1或2所述的方法 (20), 其中, 对媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的选择 (23) 基于从以下各项中选择的至少一个参数的参数值: 传输质量、媒体服务处理、传输链路成本、传输链路容量、拓扑约束、当前系统负荷、媒体服务器能力、末端通信设备能力、安全性、数据中心中的媒体服务器的可用性、使用数据中心的成本、位置。

8. 根据权利要求1所述的方法 (20), 其中, 包括: 在通信设备 (6、6A、6B、6C、6D) 从媒体服务断开时, 或者在通信设备 (6、6A、6B、6C、6D) 连接到媒体服务时, 或者在收到涉及链路故障或数据中心故障的信息时, 或者周期性地, 针对参与媒体服务的所有通信设备 (6、6A、6B、6C、6D) 评估所选择的媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的配置; 并且, 当评估证实针对参与媒体服务的一个或多个通信设备 (6、6A、6B、6C、6D) 的所述媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的重选提高了性能时, 执行这样的重选。

9. 一种用于选择分布式云 (2) 的媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的控制器节点 (5), 所述分布式云 (2) 包括能够被配置为处理相应媒体服务实例中的媒体服务所需要的媒体处理的两个或更多个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 每个媒体处理单元包括数据中心, 并且所述控制器节点 (5) 包括处理器 (30) 和存储器 (32), 所述存储器 (32) 包含能够由所述处理器 (30) 执行的指令, 由此所述控制器节点 (5) 能够用于:

- 从通信设备 (6、6A、6B、6C、6D) 接收对媒体服务的请求,

- 针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 获得与媒体服务的处理有关的参数的至少一个可配置参数值, 以及

- 基于所述至少一个参数值来选择用于处理所述通信设备 (6、6A、6B、6C、6D) 所请求的媒体服务的媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 包括以下步骤:

- i. 针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12), 通过对相应媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的所有参数值求和来确定 (24) 分数, 相应的和构成每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的相应分数,

- ii. 选择 (25) 具有最高分数的第一数据中心 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12),

- iii. 确定在所选择的第一数据中心 (4、4A、4B、4D; 10、11、12) 中创建了最大数量的媒体服务器实例以及确定所有的媒体服务器实例满负荷,

- iv. 选择具有第二高分数的第二数据中心 (4、4A、4B、4D; 10、11、12), 以及

- v. 将所述通信设备 (6、6A、6B、6C、6D) 连接到所选择的第二数据中心 (4、4A、4B、4D; 10、11、12) 的媒体服务器实例, 其中, 所述媒体服务器实例被配置为处理媒体处理需要。

10. 根据权利要求9所述的控制器节点 (5), 被配置为: 通过将具体媒体处理单元的相应参数值除以所有媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 的所有对应参数值之和来将每个参数值归一化, 针对每个处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 确定分数。

11. 根据权利要求9所述的控制器节点 (5), 被配置为: 通过还针对每个参数获得与该参数的重要性相对应的权重, 来针对每个媒体处理单元 (3、3A、3B、3D; 4、4A、4B、4D; 10、11、12) 获得至少一个可配置参数值, 以及其中, 所述控制器节点 (5) 被配置为还基于权重来进行确定。

12. 根据权利要求9所述的控制器节点 (5), 包括数据中心 (4A、4B、4D; 10、11、12), 所述

数据中心被配置为:

-选择所选择的数据中心(4、4A、4B、4D;10、11、12)的现有媒体服务器实例,或者在所选择的数据中心(4、4A、4B、4D;10、11、12)中创建新的媒体服务器实例,以及

-将所述通信设备(6、6A、6B、6C、6D)连接到所选择的现有媒体服务器实例或连接到创建的新媒体服务器实例。

13. 根据权利要求9所述的控制器节点(5),被配置为基于以下各项中的一项或多项来选择现有媒体服务器实例:媒体服务器实例的负荷、媒体服务器实例的能力、媒体服务器群中的媒体服务器实例的数量、媒体服务器实例所服务的现有会话、其上正在运行媒体服务器实例的硬件的能力、连接性、带宽和操作状态。

14. 根据权利要求9所述的控制器节点(5),被配置为:通过在确定现有媒体服务器实例的负荷高于第一阈值时创建新的媒体服务器实例,并且在确定现有媒体服务器实例的负荷低于第二阈值时移除媒体服务器实例,将所述通信设备(6、6A、6B、6C、6D)连接到所选择的数据中心(4、4A、4B、4D;10、11、12)的媒体服务实例,并且基于业务负荷来适配媒体服务器实例的数量。

15. 根据权利要求9所述的控制器节点(5),被配置为:基于从以下各项中选择的至少一个参数的参数值执行对媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)的选择:传输质量、媒体服务处理、传输链路成本、传输链路容量、拓扑约束、当前系统负荷、媒体服务器能力、末端用户通信设备能力、安全性、数据中心中的媒体服务器的可用性、使用数据中心的成本、位置。

16. 根据权利要求9所述的控制器节点(5),其中,包括:在通信设备(6、6A、6B、6C、6D)从媒体服务断开时,或者在通信设备(6、6A、6B、6C、6D)连接到媒体服务时,或者在收到涉及链路故障或数据中心故障的信息时,或者周期性地,针对参与媒体服务的所有通信设备(6、6A、6B、6C、6D)评估所选择的媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)的配置;并且,当评估证实针对参与媒体服务的一个或多个通信设备(6、6A、6B、6C、6D)的媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)的重选提高了性能时,执行这样的重选。

17. 一种存储器(32),包括计算机程序(31)和其上存储有所述计算机程序(31)的计算机可读装置,其中,所述计算机程序(31)由控制器节点(5)用来选择分布式云(2)的媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12),所述分布式云(2)包括能够被配置为处理媒体服务所需要的媒体处理的两个或更多个媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12),所述计算机程序(31)包括计算机程序代码,当在所述控制器节点(5)上运行所述计算机程序代码时,导致所述控制器节点(5):

-从通信设备(6、6A、6B、6C、6D)接收对媒体服务的请求,

-针对每个媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)获得与所述媒体服务的处理有关的参数的至少一个可配置参数值,以及

-基于所述至少一个参数值来选择用于处理所述通信设备(6、6A、6B、6C、6D)所请求的媒体服务的媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12),包括以下步骤:

i. 针对每个媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12),通过对相应媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)的所有参数值求和来确定(24)分数,相应的和构成每个媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)的相应分数,

- ii. 选择 (25) 具有最高分数的第一数据中心 (3、3A、3B、3D;4、4A、4B、4D;10、11、12) ,
- iii. 确定在所选择的第一数据中心 (4、4A、4B、4D;10、11、12) 中创建了最大数量的媒体服务器实例以及确定所有的媒体服务器实例满负荷,
- iv. 选择具有第二高分数的第二数据中心 (4、4A、4B、4D;10、11、12) ,以及
- v. 将所述通信设备 (6、6A、6B、6C、6D) 连接到所选择的第二数据中心 (4、4A、4B、4D;10、11、12) 的媒体服务器实例,其中,所述媒体服务器实例被配置为处理媒体处理需要。

用于选择媒体处理单元的方法和网络节点

技术领域

[0001] 本文公开的技术大体上涉及分布式云环境的领域,并且具体涉及在这样环境中对用于媒体处理的媒体处理单元的选择。

背景技术

[0002] 在传统的集中式云环境中,所有的计算在一个集中式大型数据中心内执行。与此不同,在分布式云环境中,不存在单个中央数据中心。相反,分布式云由潜在地大量的地理上分散的数据中心构成。这些数据中心具有各种各样的能力;一些数据中心可能相对较小并且可能位于包括分布式云环境的网络的边缘,而其它数据中心可能位于网络的核心并且可能设置有很高的容量。

[0003] 传统上,使用专用服务器硬件和数字信号处理器(DSP)提供统一通信(UC)服务(例如多方音频和视频会议)。目前,存在从基于硬件的UC解决方案向完全基于软件的虚拟化云环境移植的增长趋势。移植中的第一个步骤是在集中式云环境中提供基于软件的UC服务。接下来的预见步骤是在分布式云环境中提供该服务。

[0004] 图1示出了分布式云环境(在下文中也将其称作网络1)中的媒体处理的简单示例。在附图中,分布式云2正在为4个用户(用户A、B、C和D)提供视频会议服务。媒体处理按照以下方式在云2中分布:存在接近位于网络1的边缘处的用户的本地媒体服务(MS)3A、3B、3D实例。会议的音频混合和切换正由网络1的核心处的大型数据中心中的媒体服务器3处理。每个媒体服务实例运行在数据中心4A、4B、4D、4中的一个虚拟机(VM)中。

[0005] 媒体处理需要分布到若干个虚拟机的原因在于,单个虚拟机的容量通常不足以处理会议中所有用户的媒体处理。这正是例如用户可以使用不同的编解码并因此需要代码转换以在不同的媒体格式之间进行解译的高清晰视频会议的情况。将媒体处理分布到不同数据中心中的虚拟机的原因在于,当媒体处理发生在与会议参与者尽可能近的位置时,对于位于接近本地媒体服务器的用户来说,延时可以最小化。此外,可以使响应性最大化。需要使延时最小化以提高服务的用户体验的质量。最大化的响应性的示例是使用来自本地无线接入网络(RAN)的反馈来动态地适配向用户发送的视频流的比特率的能力。将用户连接到在网络方面最接近的数据中心的另一个优点在于,数据中心之间的连接性通常是受到管理的,而用户的连接性可能是尽力而为的。因此,使媒体流在尽力而为的连接上传送的距离最小化是有意义的。

[0006] 将媒体处理分布到位于不同数据中心中的若干个虚拟机的需求可以导致媒体会话(例如多媒体会议会话)的高复杂度的互连拓扑。由于可能涉及到大量虚拟机和数据中心,导致的拓扑通常明显比在基于硬件的媒体服务的情况下或在非分布式云环境中运行媒体服务器的情况下使用的媒体处理拓扑更复杂。因此,为多媒体会话选择最佳媒体处理拓扑是不平常的问题。

[0007] 一种方法可以是在多媒体会话中实现对仅一个或几个拓扑的支持并使用同一最初选择的拓扑,例如通过将每个新用户连接到地理上最接近的媒体服务器并将每个媒体服

务器连接到位于星型拓扑的中间的同一中央服务器,将新用户和媒体服务器连接到网络。该方法的缺点在于,在多媒体会话的有效期内,最初的拓扑选择可能不能保持最佳。此外,这可能并不总是导致最佳质量的用户体验。对复杂地理分布的云环境而言,该方法缩放性(scale)非常差。作为示例,随着参与到拓扑中的媒体服务器的数量的增长,全网络拓扑可能很快遇到可缩放性的问题。作为另一个示例,随着连接到中央服务器的其它媒体服务器的数量的增长,位于星型拓扑的中间的中央服务器还可能变为瓶颈。

[0008] 使用低效或不合适的媒体处理拓扑可以导致延时、抖动和包丢失增加,以及因此恶化媒体会话质量。由于延时、抖动和包丢失倾向于随着复杂性以及通信端点之间的网络距离的增加而增加,这些问题在全球地理分布的多媒体会话中更加突出。

发明内容

[0009] 本公开的目的在于解决或至少减轻上述问题中的至少一个。

[0010] 根据第一方案,该目的通过一种在分布式云的网络节点中执行的用于选择媒体处理单元的方法来实现。所述分布式云包括能够被配置为处理媒体服务所需要的媒体处理的两个或更多的媒体处理单元。方法包括:从通信设备接收对媒体服务的请求针对每个媒体处理单元来获得与所述媒体服务的处理有关的参数的至少一个可配置参数值,以及基于所述至少一个参数值来选择用于处理所述通信设备的所请求的媒体服务的媒体处理单元。

[0011] 该用于进行选择的方法能够实现针对在分布式云环境中运行的媒体服务(例如多媒体会议)构建改进的媒体处理拓扑。服务提供者可以基于构建哪种拓扑来定义策略。可以通过可配置的一个或多个参数值灵活地定义策略。策略可以是复杂的,其中同时应用多个策略,或者是简单的,其中只应用一个策略,例如“总选择最低成本的备选”。由此,可考虑到服务提供者设置的一个或多个可配置参数来对媒体处理拓扑进行优化。此外,虽然存在媒体处理发生在高动态地理分布的云环境中的虚拟媒体服务器上这一事实,例如考虑到低延迟、低抖动和低包丢失,该方法能够实现高体验质量等级。该方法还可以用于动态调节已经存在的媒体处理拓扑。

[0012] 根据第二方案,该目的通过用于选择分布式云的媒体处理单元的网络节点来实现。所述分布式云包括能够被配置为处理媒体服务所需要的媒体处理的两个或更多的媒体处理单元。所述网络节点包括处理器和存储器,所述存储器包含能够由所述处理器执行的指令,由此所述网络节点能够用于:从通信设备接收对媒体服务的请求;针对每个媒体处理单元,获得与媒体服务的处理有关的参数的至少一个可配置参数值,以及基于至少一个参数值来选择用于处理所述通信设备的所请求的媒体服务的媒体处理单元。

[0013] 根据第三方案,该目的通过网络节点用来选择分布式云的媒体处理单元的计算机程序来实现,其中,所述分布式云包括能够被配置为处理媒体服务所需要的媒体处理的两个或更多的媒体处理单元。所述计算机程序包括计算机程序代码,当在所述网络节点上运行所述计算机程序代码时,导致所述网络节点:从通信设备接收对媒体服务的请求;针对每个媒体处理单元,获得与媒体服务的处理有关的参数的至少一个可配置参数值,以及基于至少一个参数值来选择用于处理所述通信设备的所请求的媒体服务的媒体处理单元。

[0014] 根据第四方案,该目标是由包括上述计算机程序在内的计算机程序产品以及其上存储有所述计算机程序的计算机可读装置来实现的。

[0015] 此外,当阅读以下描述和附图时,本公开的其它特征和优点将变得清楚。

附图说明

- [0016] 图1示意性示出了可以实现本公开的实施例的分布式云环境。
[0017] 图2示出了全网格拓扑。
[0018] 图3示出了链状拓扑。
[0019] 图4示出了星型拓扑。
[0020] 图5示出了树形拓扑。
[0021] 图6示出了地理分布的会议会话的示例。
[0022] 图7是示出了根据本公开的在网络节点中执行的操作的流程图。
[0023] 图8示出了根据本公开的在网络节点中的方法步骤的流程图。
[0024] 图9示意性示出了用于实现本公开的方法的网络节点和装置。
[0025] 图10示出了包括用于实现本公开的方法的功能模块/软件模块的网络节点。

具体实施方式

[0026] 在以下描述中,出于说明而非限制的目的,阐述具体细节,例如特定架构、接口、技术等,以提供完全的理解。在其它实例中省略了对公知设备、电路和方法的详细描述,以避免以因为不必要的细节使描述不清楚。在本说明书全文中,同样的附图标记指代相同或相似的元素。

[0027] 如背景技术部分中描述的,将媒体处理分布到不同数据中心中的若干个虚拟机的需求产生了与例如使用单个数据中心相比更复杂的媒体会话互连建立。在下文中,用于媒体处理的网络元件(例如图1的媒体服务器3、3A、3B、3C)之间的互连被称作媒体处理拓扑。在图2至图5中给出了这样的媒体处理拓扑的一些示例。在这些附图中,每个节点表示媒体服务器。末端用户连接到用于发送和接收实时媒体流的媒体服务器。

[0028] 图2示出了全网格拓扑的示例。在全网格拓扑中,每个单个媒体服务器实例连接到每个其它媒体服务器实例。

[0029] 图3示出了链状拓扑。在该拓扑中,媒体服务器构成链,在链中,第一媒体服务器连接到第二媒体服务器,第二媒体服务器连接到第三媒体服务器,第三媒体服务器连接到第四媒体服务器,以此类推,直到达到在链的末端处的最后一个媒体服务器。

[0030] 图4示出了星型拓扑。在星型拓扑中,每个媒体服务器实例连接到同一中央媒体服务器实例,因此该中央媒体服务器实例变为星型拓扑的中心。

[0031] 图5示出了树形拓扑。在树形拓扑中,存在具有多个子节点的根节点。子节点可以具有其本身的子节点。树形拓扑中不具有任何进一步子节点的最后一级子节点被称作叶节点。

[0032] 暂时回到图1,注意,网络1以及特别是分布式云2还包括控制器节点5(在附图中还表示为CTRL),控制器节点5连接到所有的媒体服务器3、3A、3B、3D以便对其进行控制。这样的控制器节点5还包括在图6中示出的示例情形中,并将在下文中描述,控制器节点5连接到媒体服务器10、11、12。

[0033] 虽然当同一分布式媒体服务器后面的用户正在相互通信时的延时较短,由媒体处

理拓扑表示的媒体服务器链的不同末端处的用户之间的媒体流可能经历较长的延迟。这仅仅是由于以下事实造成的：当从用户A向用户B的媒体流经过多个媒体服务器时，由每个单独的媒体服务器完成的处理增加了携带媒体的实时传输协议(RTP)包所经历的端到端延迟。作为示例，如果媒体通过具有三个媒体服务器的链，在最差的情况下，与使用单个大容量中央媒体服务器的情形相比，媒体服务器上的处理所引入的延迟可能是三倍。根据现有的推荐，为了使用户非常满意，用户之间的单向(即嘴到耳朵的)媒体延迟应当不高于225ms。如果延迟超过300ms，一部分用户将开始变得不满意。对于执行媒体解码和编码的单个基于软件的媒体服务器来说，以100ms的量级增加RTP包经历的端到端延迟是不常见的。因此，对媒体进行编码和解码的三个媒体服务器的存在已经足以使一部分用户不满意了。因此，明显的是，媒体处理拓扑在任意两个用户之间不应该包含太多的媒体服务器层。

[0034] 除了端到端延时增加，低效的媒体处理拓扑还可导致抖动和包丢失二者增加。这是因为，较差地选择的媒体处理拓扑可能包含数据中心之间的质量较低且堵塞的传输链路、任意两个用户之间的过长的媒体服务器链和跨数据中心链路、以及过载的虚拟机和物理服务器。

[0035] 考虑到媒体处理拓扑的不合适的或静态的选择可能在端到端媒体流质量上具有的至少上述缺点，在各种方案中，本公开提供了一种解决方案，该解决方案考虑到分布式云环境的当前状态，可以使用灵活的策略集合来动态构建最佳媒体处理拓扑，并且还能够根据需要适配拓扑。

[0036] 简言之，本公开提供了可以被应用以选择最佳媒体处理单元(例如数据中心、数据中心中的最佳媒体服务器群、以及媒体服务器群中的最佳媒体服务器)的机制，以服务于加入多媒体会议会话的新用户。媒体服务器群指代在单个数据中心或若干个数据中心中运行的媒体服务器实例的集合。也就是说，媒体服务器群可以只贯穿单个数据中心，或者贯穿若干个数据中心。媒体服务器实例是其上运行媒体服务器的虚拟机。使用根据本公开的选择机制，能够实现现在分布式云环境中的不同数据中心中运行的媒体服务器之间建立最佳的媒体处理拓扑。该机制还可以用于响应于改变中的操作情况来周期性地调节媒体处理拓扑。通过可以由服务提供者定义的灵活的策略集合来引导选择机制的操作。

[0037] 图6示出了分布式实时多媒体会话的示例，并且具体地，示出了包括媒体流的会议会话。在所示出的情况下，使用的媒体服务器(附图中未示出)可以部署在3个数据中心10、11、12中，数据中心10、11、12分别位于美国、欧洲和亚洲。根据现有技术，为了尽可能靠近用户执行媒体处理，亚洲的末端用户将被引导与亚洲数据中心12中的媒体服务器连接，美国的末端用户将连接到美国数据中心10，并且欧洲的末端用户将连接到欧洲数据中心11。然而，由于各种参数(例如，诸如传输链路成本、质量需求、带宽、处理容量、使用的媒体解码等)，可能发生的是，在某些情况下，亚洲的用户最好直接连接到在例如欧洲的媒体服务器。例如，这可能是由于每个媒体服务器向媒体流添加延时这一事实造成的，尤其是如果媒体流需要被解码和/或编码。因此，为了保持延时较低，把处理媒体流的媒体服务器的数量保持尽可能低可能是有利的。

[0038] 当在(例如图6中示出的)地理分布的数据中心环境中发生媒体处理时，可以基于会议参与者所在的位置来在构成分布式云的不同数据中心10、11、12中动态地布置媒体服务实例。随着现有媒体服务器实例的负荷提高，可以创建媒体服务器的新实例来处理业务。

当业务量较低时,可以减少媒体服务器实例的数量。当在数据中心中已经创建了最大数量的媒体服务器实例并且所有的媒体服务器实例都满负荷时,需要将新的媒体会话重新引导到备选数据中心,即使媒体质量将不是最佳的。

[0039] 针对例如一个大规模分布式会议中的一个媒体会话,可以存在很多媒体服务器。无论新用户在什么时候加入会议,都需要做出以下决定:将用户连接哪个媒体服务器,或者是否需要将用户连接到没有之前的媒体服务器实例的新数据中心。所导致的媒体服务器的互连拓扑可能是高度复杂的。因此,当将用户和为其提供服务的新媒体服务器连接到会议的媒体处理拓扑时,需要遵循仔细的选择程序。在下文中,描述了根据本公开的方案的媒体处理单元选择程序。例如,媒体处理单元可以是数据中心或媒体服务器。在下文中,数据中心被用作这样的媒体处理单元的示例,以便描述选择机制。

[0040] 选择数据中心

[0041] 如上所述,当新的末端用户希望加入会议时,应当确定该用户连接到的最佳数据中心。在分布式云环境中,通常存在可用于为用户提供服务的多个备选的数据中心。作为示例,使用蜂窝因特网连接从她的移动电话加入会议的用户可能连接到在无线接入网络中运行的若干个小的附近的数据中心中的一个,或者连接到能够运行媒体服务器实例的虚拟化边缘路由器,或者连接到在分布式云环境中的不同位置处运行的任意数量的较大数据中心。

[0042] 无论多个数据中心在什么时候可用于为用户提供服务,都可以应用选择过程来从候选数据中心的集合中选择最佳数据中心。存在多个策略,特别是多个参数,其可以影响新用户应当连接到的数据中心的策略。这样的策略可以但不限于包括以下各项中的一个或多个:

[0043] • 传输质量,例如包括:用户与候选数据中心之间的点到点延时;如果将用户连接到该候选数据中心,该用户与会议中的其他用户之间的平均端到端延时;如果连接到该候选数据中心,用户将经历的抖动;用户与候选数据中心之间的网络吞吐量。

[0044] • 处理例如现有媒体会话中的现有媒体编码方法是什么,以及新的末端用户偏好什么。

[0045] • 传输链路的成本;传输链路可能具有与其相关联的不同货币成本。作为示例,在尽力而为的因特网上运行的传输链路可以免费使用但具有较低的质量,而数据中心之间的专用的虚拟专用网络(VPN)的使用昂贵但提供较高的吞吐量和质量。

[0046] • 传输链路容量;作为示例,在两个数据中心之间可用的跨洋传输链路可能具有非常有限的容量。因此,将位于传输链路的一端处的用户连接到在链路的同一端的数据中心,该数据中心可以在将业务传递给传输链路之前从多个用户聚集业务,这可能是有利的。这样做的优点在于提高链路利用率。

[0047] • 拓扑约束,例如在媒体服务器之间创建的树形拓扑中的层的最大数量。

[0048] • 当前系统负荷;候选数据中心中的可用媒体服务器实例的负荷和容量。

[0049] • 能力,包括媒体服务器能力,例如,诸如可用编解码和执行转码或视频比特率适配的能力,以及末端用户通信设备能力。

[0050] • 安全,例如数据中心间业务的安全(加密)链路的可用性。

[0051] • 数据中心中的媒体服务器的可用性。作为示例,如果数据中心还没有任何媒体

服务器虚拟机,并且如果创建新的媒体服务器群的成本较高(例如就延时而言),则可能优选将用户连接到已经运行媒体服务器群的数据中心。

[0052] • 使用数据中心的成本;作为示例,使用位于虚拟化RAN站点中的较小数据中心可能比使用大型中央数据中心更昂贵。作为另一个示例,由于时差,在欧洲的办公时间内,使用位于美国的数据中心可能比使用欧洲的数据中心更便宜。具体地,假设数据中心提供者采用动态定价,由于在美国是晚间,美国数据中心可能具有大量的空余容量,因此使用可能更便宜。

[0053] 当选择要为用户提供服务的数据中心时需要考虑的不同策略(特别是参数)的数量可发生变化,并且可以取决于具体的部署。

[0054] 在数据中心选择过程中,上述策略(参数)中的每一个可以带有不同的权重。被看做是非常重要的策略具有较高的权重,而被视为不太重要的策略具有较低的权重。因此,可以考虑其对操作者的重要性来对策略/参数进行排序。权重的值可以取决于服务提供者、应用和/或末端客户的偏好。通过这种方式,策略定义了媒体会话的某些方面的重要性。作为示例,在一些情况下,传输链路的低成本可能比传输质量更重要,即,在低优先级的企业内会议中,与VPN链路相比,优选尽力而为为传输链路,而在其它的情况下,情况相反(对于企业的客户也参与的高优先级企业会议来说,传输的成本可能不是问题)。在前一示例中,参数“传输链路成本”因此被赋予较高权重,而在后一示例中,该参数被赋予较低权重。

[0055] 如上所述,每个策略(例如数据中心的系统负荷)具有权重和数值(例如,如果策略与数据中心的负荷有关,值是“80%”,即数据中心加载到其容量的80%)。使用策略的权重和归一化值,可以针对新用户潜在可连接到的每个数据中心计算总的分数。此后,可以选择具有最高分数的数据中心来为用户提供服务。如果选择的数据中心还没有任何媒体服务器实例在其中运行,则在数据中心内创建新的媒体服务器群。通过遵循该选择机制,创建了在不同数据中心中运行的媒体服务器实例的最佳拓扑。

[0056] 在下文中,参照图7更详细地描述上述媒体处理单元(以数据中心为例)选择程序。图7是示出了根据本教导的方案的数据中心选择程序100中的步骤的流程图。程序100可以在网络节点(例如,诸如图1和图6的控制器节点5或专用会议控制器)中实现并由网络节点执行。

[0057] 程序100在方框102中开始,其中发起拓扑定义或选择过程。

[0058] 在方框104中,针对所期望的媒体服务(例如具有多个参与者的所计划的媒体会议),分布式云环境2的操作者定义要在数据中心选择程序中使用的策略(即参数)。在下文中,会议被用作媒体服务的示例。如上文解释的,策略的定义可以考虑各种优化方案来进行,并且对于不同的操作者可以极大不同。

[0059] 在方框106中(其属于可选步骤),操作者可以针对在之前的步骤中定义的每个策略来定义权重值。

[0060] 在方框108中,新用户希望加入会议。

[0061] 在方框110中,例如基于在之前的步骤中(方框108)从用户接收到的输入,网络节点确定用户应当连接到哪个会议,或者是否应当创建新会议。传统的呼叫控制可被用于实现该步骤,包括例如获得媒体会话和媒体描述输入。

[0062] 在方框112中,网络节点针对所确定的会议从数据库或类似的系统取得服务提供

者已定义的策略以及策略的权重(如果存在的话)。这可被实现为网络节点中的拓扑控制器功能。

[0063] 在方框112中,网络节点(例如作为拓扑控制器功能的一部分)确定加入的用户可以连接到的候选数据中心的完整集合。

[0064] 然后,针对每个候选数据中心执行下列步骤(其与方框114、116、118、120相对应):

[0065] 在方框114中,网络节点5获得每个策略(即每个参数)的数值。作为示例,网络节点5可以检查当前候选数据中心或在其中运行的现有媒体服务器群的系统负荷。参数中的一些可以是所谓的开/关参数。开/关参数的示例是转码支持:数据中心支持转码(在这种情况下,该值为“开启”),或者不支持转码(该值为“关闭”)。“开启”值的数值是1,并且“关闭”值的数值是0。网络节点5还可以将当前候选数据中心的所有参数的值归一化。下文描述归一化程序的示例。

[0066] 在方框116中,网络节点5将权重(如果存在的话)应用到数据中心的所有参数的数值。

[0067] 在方框118中,网络节点5确定参数的“质量”前缀。“质量”是指对于参数来说,较高值是否比较低值更好这一事实,或反之亦然。如果优选较高值,则质量前缀为正(+)。如果优选较低值,则质量前缀为负(-)。作为示例,对于系统负荷来说,在选择过程中,通常优选较低值。相比之下,对于吞吐量来说,优选较高值。下文将描述“质量”前缀的使用示例。

[0068] 在方框120中,网络节点5通过将归一化的值与有关的权重相乘并将质量前缀(正号或负号)应用到该结果,计算每个参数的分数。最后,网络节点5将每个单独的参数的分数求和,以形成当前候选数据中心的分数。

[0069] 方框114、116、118、120的步骤执行一次(操作者定义的唯一策略),或者针对操作者已经定义的每个策略来重复。在方框122中,如果确定已经处理了所有所需要的策略,则流程继续进行到方框124。

[0070] 在方框124中,如果已经计算出了每个候选数据中心的分数,网络节点5选择具有最高分数的数据中心。如果选择的数据中心还没有任何现有的媒体服务器,则网络节点5在选择的数据中心中创建新的媒体服务器群。

[0071] 在方框126中,网络节点5将用户连接到选择的数据中心,即具有最高分数的数据中心。

[0072] 在方框128中,网络节点5将用户连接到在所选择的数据中心中的媒体服务器群。下文介绍媒体服务器群选择的程序。网络节点5将用户连接到媒体服务器群中的一个媒体服务器。下文描述媒体服务器选择的程序。

[0073] 在方框130中,网络节点5更新对应的媒体性能数据库(与步骤112相比)以及媒体会话。这可以实现动态改变的媒体服务器,其中例如当新用户进入会议或现有用户退出会议时,数据中心可以针对参与用户发生变化。

[0074] 下面给出媒体处理单元选择过程的具体示例,同样使用数据中心作为将要选择的媒体处理单元,其中图6中示出的分布式云正在运行分布式会议服务。三个数据中心10、11、12位于美国的堪萨斯城(Kansas City)、欧洲的法兰克福(Frankfurt)和中国北京(Beijing)。我们还假设存在希望加入正在进行的会议的位于北京附近的用户,名字叫Ning。在本示例中,会议服务的服务提供者已经针对数据中心选择程序定义的策略(即参

数)如下:

- [0075] • 加入的用户和候选数据中心 (DC) 之间的点到点延时, 例如以毫秒 (ms) 测量
- [0076] • 如果使用感兴趣的候选数据中心, 加入的用户与会议中的现有用户之间的平均端到端 (E2E) 延时
- [0077] • 候选数据中心中的媒体服务器群的负荷
- [0078] • 使用候选数据中心的有关成本 (在金钱方面)
- [0079] • 如果使用候选数据中心, 加入用户与会议中的现有用户之间的平均抖动
- [0080] • 如果使用候选数据中心, 树形拓扑中的层的最大数量
- [0081] • 如果使用候选数据中心, 所估计的加入用户的平均包丢失
- [0082] • 在候选数据中心中速率转换 (transrating) 的视频流的支持的可用性 (所谓的开/关参数)
- [0083] • 加入的用户与候选数据中心之间的吞吐量
- [0084] 表格1示出了在假设上述策略 (即参数) 的情况下三个数据中心的得分过程的示例。

[0085]

	到该DC的延时 [ms]	与其他用户的平均E2E延时 [ms]	MS群的负荷 [%]	相对DC成本	与其他用户的平均抖动 [ms]	如果使用该DC, 树中的层	对其他用户的平均包丢失 [%]	DC所支持的速率转换	对第一跳DC的吞吐量 [Mbps]
权重	1	5	4	3	4	2	2	4	5
质量	-	-	-	-	-	-	-	+	+
相对权重	-0,0333	-0,1667	-0,133	-0,1	-0,13333	-0,06667	-0,067	0,1333333	0,166667
分数									
Beijing	50	138	90	0,5	20	4	0,03	0	5
Frankfurt	150	238	50	1	30	3	0,02	1	2
Kansas City	100	188	25	0,4	40	3	0,01	1	3
总和	300	564	165	1,9	90	10	0,06	2	10
归一化分数									
Beijing	0,1667	0,24468	0,5455	0,26316	0,222222	0,4	0,5	0	0,5
Frankfurt	0,5	0,42199	0,303	0,52632	0,333333	0,3	0,3333	0,5	0,2
Kansas City	0,3333	0,33333	0,1515	0,21053	0,444444	0,3	0,1667	0,5	0,3
加权归一化分数									
Beijing	-0,0056	-0,0408	-0,073	-0,0263	-0,02963	-0,02667	-0,033	0	0,083333
Frankfurt	-0,0167	-0,0703	-0,04	-0,0526	-0,04444	-0,02	-0,022	0,0666667	0,033333
Kansas City	-0,0111	-0,0556	-0,02	-0,0211	-0,05926	-0,02	-0,011	0,0666667	0,05
结果	加权分数总和								
Beijing	-0,15								
Frankfurt	-0,17								
Kansas City	-0,08								

[0086] 表格1

[0087] 如表格1中记录的计算数据中心的分数的过程如下。首先, 向 (如最上面一行中定义的) 每个参数分配1到5之间的权重值 (该权重值的范围仅用作示例)。还向每个参数分配质量 (正或负)。通过将权重除以所有权重之和并且将质量 (正号或负号) 应用到参数, 计算每个参数的相对权重。例如, 对参数“到该DC的延时 [ms]”执行该处理, 得到: 1 / (1+5+4+3+4+

$2+2+4+5) * (-1) = -1/30 = -0.0333$ 。

[0088] 接下来,通过将每个单独的值除以所有数据中心10、11、12的单独的值之和来将每个参数(例如延时)的每个值归一化。例如,对参数“到该DC的延时[ms]”执行该处理,针对北京得到: $50 / (50+150+100) = 0.1667$,即归一化分数0.1667。

[0089] 此后,通过将参数的归一化值乘以相对权重来计算每个参数的分数。例如,对参数“到该DC的延时[ms]”执行该处理,针对北京的数据中心12得到: $0.1667 * (-0.0333) = -0.0056$ 。

[0090] 最终,对针对给定数据中心的每个参数的分数求和,得到该数据中心的分数。例如,对北京的数据中心12执行该处理,得到: $-0.0056 + (-0.0480) + (-0.073) + (-0.0263) + (-0.02963) + (-0.02667) + (-0.033) + 0 + 0.083333 = -0.15$ 。在这种情况下,堪萨斯城的数据中心10具有最高分(-0.08),并且因此选择它来对Ning提供服务。

[0091] 媒体服务器群/媒体服务器选择

[0092] 还可以应用上述媒体处理单元选择机制(其中数据中心被用作媒体处理单元的示例),以从在例如单个数据中心中运行的多个候选媒体服务器群中选择一个媒体服务器群。当将媒体处理单元选择机制用于媒体服务器群选择时,可能需要应用不同的策略。用于数据中心选择的策略中的一些可能不是相关的媒体服务器群选择(例如与数据中心的成本有关的策略)。此外,附加的相关策略可以应用到媒体服务器群选择的问题(例如群中的轻度加载或未加载的媒体服务器的数量)。

[0093] 此外,可以针对各种目的来应用选择媒体处理单元的方法,包括但不限于数据中心、媒体服务器群、媒体服务器和传输链路选择。

[0094] 还可以应用媒体处理单元选择机制以从构成媒体服务器群的所有媒体服务器虚拟机中选择一个媒体服务器虚拟机。

[0095] 传输链路选择

[0096] 在一对数据中心之间存在多个备选传输链路(例如尽力而为因特网上的传输链路、多协议标签交换虚拟专用网(MPLS VPN)链路、专用光纤连接、或者甚至是天线链路)的情况下,上述选择程序可以用来选择用于媒体流或媒体流集合的最佳链路。

[0097] 到目前为止,已经描述了当新用户加入多媒体会议时如何执行选择过程。选择程序的使用情况的另一个示例是周期地重新执行该程序,以响应于改变中的操作情况来调节媒体处理拓扑。在这种情况下,选择程序用于基于最新的性能参数(例如系统负荷、端到端延时、吞吐量、包丢失、抖动等等)来重新计算媒体处理单元(例如数据中心(和/或媒体服务器群、媒体服务器、传输链路))的新的分数。如果重新执行选择机制的结果是一些旧的拓扑决定不再是最优的,则可以采取措施来调节拓扑。作为示例,数据中心之间的业务可以从一个传输链路切换到另一个传输链路,在两个数据中心之间使用的媒体格式可以被改变,或者可以通过将处理转移到树中不同层处的另一媒体服务器来在树形拓扑中绕开媒体服务器。因此,选择方法可以是动态运行时调节的媒体处理拓扑选择方法。

[0098] 图8示出了根据本公开的在网络节点5中的方法20步骤的流程图。因此,用于选择媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的方法20可以在分布式云2的网络节点5中执行。分布式云2包括两个或更多个能够被配置为处理媒体服务所需要的媒体处理的处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0099] 方法20包括从通信设备6、6A、6B、6C、6D接收21对媒体服务的请求。例如,通信设备6、6A、6B、6C、6D可以包括智能电话、手机、膝上型计算机、个人计算机、视频会议设备或任何其它的末端用户设备。例如,媒体服务可以是多方会议呼叫或视频呼叫。这样的请求的接收可以按照传统的方式来处理,例如涉及以下步骤:诸如验证用户、确定用户应当连接哪个会议、生成对用户的响应、向下一跳目的地路由该请求,等等。

[0100] 方法20包括:针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12来获得22与媒体服务的处理有关的参数的至少一个可配置参数值。已经给出了这样的可配置参数值的若干个示例,并且可配置参数值的若干个示例可以例如与媒体服务的媒体处理有关。

[0101] 方法20包括:基于所述至少一个参数值来选择23用于处理通信设备6、6A、6B、6C、6D的所请求的媒体服务的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0102] 在实施例中,选择23包括:

[0103] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12,通过将相应媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有参数值求和来确定24分数,相应的和构成每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的相应分数,以及

[0104] -选择25具有最高分数的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0105] 在上述实施例的变体中,针对每个处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12确定24分数包括:通过将具体媒体处理单元的相应参数值除以所有媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有对应参数值之和来将每个参数值归一化。

[0106] 在其它变体中,针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12来获得22至少一个可配置参数值还包括:针对每个参数获得与该参数的重要性相对应的权重,以及所述确定24还基于所述权重。

[0107] 在实施例中,选择23包括:

[0108] -通过以下方式,针对所述两个或更多个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12中的每一个来确定24分数:

[0109] -针对每个参数获得与该参数的重要性相对应的数字权重值和包括值1或-1的质量指示符,

[0110] -针对每个参数,通过将该参数的所获得的数字权重值除以该参数的所有数字权重值之和,且将结果乘以所述质量指示符,确定相对数字权重值,

[0111] -通过将具体媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的相应参数值除以所有媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有对应参数值之和,来将每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的每个所获得的参数值归一化,提供每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的归一化参数值,

[0112] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的每个所获得的参数值,通过将相应的归一化参数值乘以相应的相对数字权重值来确定加权归一化分数,

[0113] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12,将所有参数的加权归一化分数求和,提供每个媒体处理单元的分数的分数,以及

[0114] -选择25具有最高分数的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0115] 在实施例中,媒体处理单元包括数据中心4、4A、4B、4D;10、11、12,并且方法20包括:

[0116] -确定在所选择的数据中心4、4A、4B、4D;10、11、12中创建了最大数量的媒体服务器实例以及确定所有的媒体服务器实例满负荷,

[0117] -选择具有第二高分数的数据中心4、4A、4B、4D;10、11、12,以及

[0118] 将所述通信设备6、6A、6B、6C、6D连接到所选择的数据中心4、4A、4B、4D;10、11、12的媒体服务器实例。

[0119] 在实施例中,媒体处理单元包括数据中心4、4A、4B、4D;10、11、12,并且方法20包括:

[0120] -选择所选择的数据中心4、4A、4B、4D;10、11、12的现有媒体服务器实例,或者在所选择的数据中心4、4A、4B、4D;10、11、12中创建新的媒体服务器实例,以及

[0121] -将所述通信设备6、6A、6B、6C、6D连接到所选择的现有媒体服务器实例或连接到创建的新媒体服务器实例。

[0122] 在上述实施例的变体中,对现有媒体服务器实例的所述选择基于以下各项中的一项或多项:媒体服务器实例的负荷、媒体服务器实例的能力、媒体服务器群中的媒体服务器实例的数量、媒体服务器实例所服务的现有会话、其上运行媒体服务器实例的硬件(例如虚拟化硬件)的能力、连接性、带宽和操作状态。

[0123] 在实施例中,媒体处理单元包括数据中心4、4A、4B、4D;10、11、12,并且方法20包括:通过在确定现有媒体服务器实例的负荷高于第一阈值时创建新的媒体服务器实例,并且在确定现有媒体服务器实例的负荷低于第二阈值时移除媒体服务器实例,将所述通信设备6、6A、6B、6C、6D连接到所述数据中心4、4A、4B、4D;10、11、12的媒体服务实例,并且基于业务负荷来适配媒体服务器实例的数量。

[0124] 在实施例中,方法20包括:在通信设备6、6A、6B、6C、6D从媒体服务断开时,或者在通信设备6、6A、6B、6C、6D连接到媒体服务时,或者在接收涉及链路故障或数据中心故障的信息时,或者周期地(例如,在定时器到期时),针对参与媒体服务的所有通信设备6、6A、6B、6C、6D评估所选择的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的配置,并且,当评估证实针对参与媒体服务的一个或多个通信设备6、6A、6B、6C、6D的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12(或数据中心4、4A、4B、4D;10、11、12)重选提高了性能时,执行这样的重选。

[0125] 在实施例中,对媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的选择23基于从以下各项中选择的至少一个参数的参数值:传输质量、媒体服务处理、传输链路成本、传输链路容量、拓扑约束、当前系统负荷、媒体服务器能力、末端用户通信设备能力、安全性、数据中心中的媒体服务器的可用性、使用数据中心的成本、位置。

[0126] 注意,媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12可以包括例如数据中心、媒体服务器或媒体服务器实例,并且所描述的方法的各种实施例可以用于例如选择数据中心。

[0127] 图9示意性示出了用于实现本公开的方法的网络节点和装置。具体地,通过上述控制器节点5来对网络节点进行示出和举例说明。网络节点5可被布置为从媒体服务器3、3A、3B、3D;10、11、12接收控制信令,例如与媒体服务器的互连有关的控制信令,例如对应当按照某个格式来向另一媒体服务器转发媒体流的一个媒体服务器进行指示的控制信令。这样的控制信令的其它示例包括对媒体服务器进行排序,以对具体媒体流执行媒体处理操作

(例如转码),或者命令媒体服务器将从用户接收到的若干个音频流混合成一个混合的音频流。网络节点5可被布置为执行各种任务(例如在各种数据中心中创建或查找现有媒体服务器虚拟机),被配置为执行所描述的方法和程序。网络节点5可以被配置为提供在某种意义上是最佳的互连拓扑,例如考虑到关于一个或多个参数(例如延时、链路容量等)对媒体流进行优化,在分布式云中分布媒体服务器。控制节点5还可以被配置为在云中分布媒体处理。

[0128] 网络节点5包括对分布式云2的其它网络节点(例如数据中心4、4A、4B、4D;10、11、12)的接口设备34。在附图中,这样的接口34由输入/输出(I/O)设备表示。因此,网络节点5可以通过接口设备34来与分布式云2的其它网络节点通信。通信可能牵涉从其它网络节点接收数据和向其它网络节点发送数据。接口设备34被配置为向处理器30传输这样的通信。

[0129] 网络节点5还可以包括数据库7,或者被配置为从这样的数据库7检索数据。也就是说,数据库7可以是网络节点5的一部分,或者连接到网络节点5。数据库7可以包括基于其来做出选择的全部各种策略(参数)。因此,操作者可以访问数据库7来输入相关数据。

[0130] 因此,网络节点5还包括处理器30,处理器30包括能够执行存储在存储器32(因此可以是计算机程序产品32)中的软件指令的以下各项中的一项或多项的任意组合:中央处理单元(CPU)、多处理器、微控制器、数字信号处理器(DSP)、应用专用集成电路,等等。处理器30可以被配置执行如上文描述的(例如参照图8描述的)方法的各种实施例中的任意一个。

[0131] 具体地,提供了用于选择分布式云2的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的网络节点5。分布式云2包括两个或更多个能够被配置为处理媒体服务所需要的媒体处理的处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。网络节点5包括处理器30和存储器32,所述存储器32包含能够由所述处理器30执行的指令,由此所述网络节点5能够用于:

[0132] -从通信设备6、6A、6B、6C、6D接收对媒体服务的请求,

[0133] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12,获得与媒体服务的处理有关的参数的至少一个可配置参数值,以及

[0134] -基于所述至少一个参数值来选择用于处理所述通信设备6、6A、6B、6C、6D的所请求的媒体服务的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0135] 在实施例中,网络节点5被配置为通过以下步骤来进行选择:

[0136] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12,通过对相应媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有参数值求和来确定分数,相应的和构成每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的相应分数,以及

[0137] -选择具有最高分数的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0138] 在上述实施例的变体中,网络节点5被配置为:通过将具体媒体处理单元的相应参数值除以所有媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有对应参数值之和并将每个参数值归一化,针对每个处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12确定分数。

[0139] 在其它变体中,网络节点5被配置为:通过还针对每个参数获得与该参数的重要性相对应的权重,来针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12获得至少一个可配置参数值,以及其中,所述网络节点5被配置为还基于权重来进行确定。

[0140] 在实施例中,网络节点5被配置为通过以下步骤来进行选择:

[0141] -通过以下方式,针对所述两个或多个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12中的每一个确定分数:

[0142] -针对每个参数获得与该参数的重要性相对应的数字权重值和包括值1或-1的质量指示符,

[0143] -针对每个参数,通过将该参数的所获得的数字权重值除以该参数的所有数字权重值之和,且将结果乘以所述质量指示符,确定相对数字权重值,

[0144] -通过将具体媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的相应参数值除以所有媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的所有对应参数值之和,来将每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的每个所获得的参数值归一化,提供每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、的归一化参数值,

[0145] 针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12的每个所获得的参数值,通过将相应的归一化参数值乘以相应的相对数字权重值来确定加权归一化分数,

[0146] -针对每个媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12,将所有参数的加权归一化分数求和,提供每个媒体处理单元的分数,以及

[0147] -选择具有最高分数的媒体处理单元3、3A、3B、3D;4、4A、4B、4D;10、11、12。

[0148] 在实施例中,媒体处理单元(3、3A、3B、3D;4、4A、4B、4D;10、11、12)包括数据中心(4A、4B、4D;10、11、12),数据中心被配置为:

[0149] -确定在所选择的数据中心4、4A、4B、4D;10、11、12中创建了最大数量的媒体服务器实例以及确定所有的媒体服务器实例满负荷,

[0150] -选择具有第二高分数的数据中心4、4A、4B、4D;10、11、12,以及

[0151] -将所述通信设备6、6A、6B、6C、6D连接到所选择的数据中心4、4A、4B、4D;10、11、12的媒体服务器实例。

[0152] 在实施例中,媒体处理单元包括数据中心4、4A、4B、4D;10、11、12,并且网络节点5被配置为:

[0153] -选择所选择的数据中心4、4A、4B、4D;10、11、12的现有媒体服务器实例,或者在所选择的数据中心4、4A、4B、4D;10、11、12中创建新的媒体服务器实例,以及

[0154] -将所述通信设备6、6A、6B、6C、6D连接到所选择的现有媒体服务器实例或连接到创建的新媒体服务器实例。

[0155] 在上述实施例的变体中,网络节点5被配置为基于以下各项中的一项或多项来选择现有媒体服务器实例:媒体服务器实例的负荷、媒体服务器实例的能力、媒体服务器群中的媒体服务器实例的数量、媒体服务器实例所服务的现有会话、其上运行媒体服务器实例的硬件(例如虚拟化硬件)的能力、连接性、带宽和操作状态。

[0156] 在实施例中,媒体处理单元包括数据中心4、4A、4B、4D;10、11、12,并且网络节点5被配置为:通过在确定现有媒体服务器实例的负荷高于第一阈值时创建新的媒体服务器实例,并且在确定现有媒体服务器实例的负荷低于第二阈值时移除媒体服务器实例,将所述通信设备6、6A、6B、6C、6D连接到所选择的数据中心4、4A、4B、4D;10、11、12的媒体服务实例,并且基于业务负荷来适配媒体服务器实例的数量。

[0157] 在实施例中,网络节点5被配置为:在通信设备6、6A、6B、6C、6D从媒体服务断开时,或者在通信设备6、6A、6B、6C、6D连接到媒体服务时,或者在接收涉及链路故障或媒体处理

单元3、3A、3B、3D；4、4A、4B、4D；10、11、12（例如数据中心）故障的信息时，或者周期地，针对参与媒体服务的所有通信设备6、6A、6B、6C、6D评估所选择的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12的配置，并且，当评估证实针对参与媒体服务的一个或多个通信设备6、6A、6B、6C、6D的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12重选提高了性能时，执行这样的重选。

[0158] 在实施例中，网络节点5被配置为基于从以下各项中选择的至少一个参数的参数值执行对媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12的选择：传输质量、媒体服务处理、传输链路成本、传输链路容量、拓扑约束、当前系统负荷、媒体服务器能力、末端用户通信设备能力、安全性、数据中心中的媒体服务器的可用性、使用数据中心的成本、位置。

[0159] 再次注意，媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12可以包括例如数据中心、媒体服务器或媒体服务器实例。

[0160] 仍然参照图9，存储器32可以是读写存储器（RAM）和只读存储器（ROM）的任意组合。存储器32还包括持久存储设备，其例如可以是磁存储器、光存储器、固态存储器或甚至远程安装存储器中的任意单独一个或组合。

[0161] 数据存储器33还可以被设置用于当在处理器30中执行软件指令期间读取和/或存储数据。数据存储器33可以是读写存储器（RAM）和只读存储器（ROM）的任意组合。

[0162] 本申请的公开还包含计算机程序产品32，包括用于实现如上所述的方法的计算机程序31，以及其上存储了计算机程序31的计算机可读装置。计算机程序产品32可以是读写存储器（RAM）或只读存储器（ROM）的任意组合。计算机程序产品32还包括持久性存储器，持久性存储器例如可以是磁存储器、光存储器或固态存储器中的任意单独一个或组合。

[0163] 因此，本公开包括网络节点5用来选择分布式云2的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12的计算机程序31，其中，所述分布式云2包括能够被配置为处理媒体服务所需要的媒体处理的两个或更多的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12。计算机程序31包括计算机程序代码，当在所述网络节点5上运行所述计算机程序代码时，导致所述网络节点5：从通信设备6、6A、6B、6C、6D接收对媒体服务的请求；针对每个媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12获得与所述媒体服务的处理有关的参数的至少一个可配置参数值；以及，基于所述至少一个参数值来选择用于处理所述通信设备6、6A、6B、6C、6D的所请求的媒体服务的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12。

[0164] 本公开还提供了包括所述计算机程序31在内的计算机程序产品32、以及其上存储有计算机程序31的计算机可读装置。计算机程序产品32或存储器因此包括能够由处理器30执行的指令。这样的指令可以包括在计算机程序中，或者包括在一个或多个软件模块或功能模块中。

[0165] 在图10中示出了使用功能模块/软件模块的实现的示例，具体示出了包括用于实现本公开的方法的功能模块的网络节点5。网络节点5包括装置，具体包括第一功能模块41，用于从通信设备6、6A、6B、6C、6D接收对媒体服务的请求。网络节点5包括装置，具体包括第二功能模块42，用于针对每个媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12来获得与上述媒体服务的处理有关的参数的至少一个可配置参数值。网络节点5包括装置，具体包括第三功能模块43，用于基于所述至少一个参数值来选择用于处理所述通信设备6、6A、6B、6C、6D的所请求的媒体服务的媒体处理单元3、3A、3B、3D；4、4A、4B、4D；10、11、12。

[0166] 功能模块41、42、43可以使用软件指令(例如在处理器中执行的计算机程序)和/或使用硬件(例如应用专用集成电路、现场可编程门阵列、离散逻辑元件等)来实现。

[0167] 受益于以上的描述和相关联的附图中呈现的教导,本领域技术人员将想到对公开的实施例的修改和其它实施例。因此,应当理解本公开不受限于所公开的具体实施例,且修改和其他实施例预期被包括在本公开的范围之内。虽然本文可能使用了特定术语,但是其用于一般性或描述性意义,且不用于限制目的。

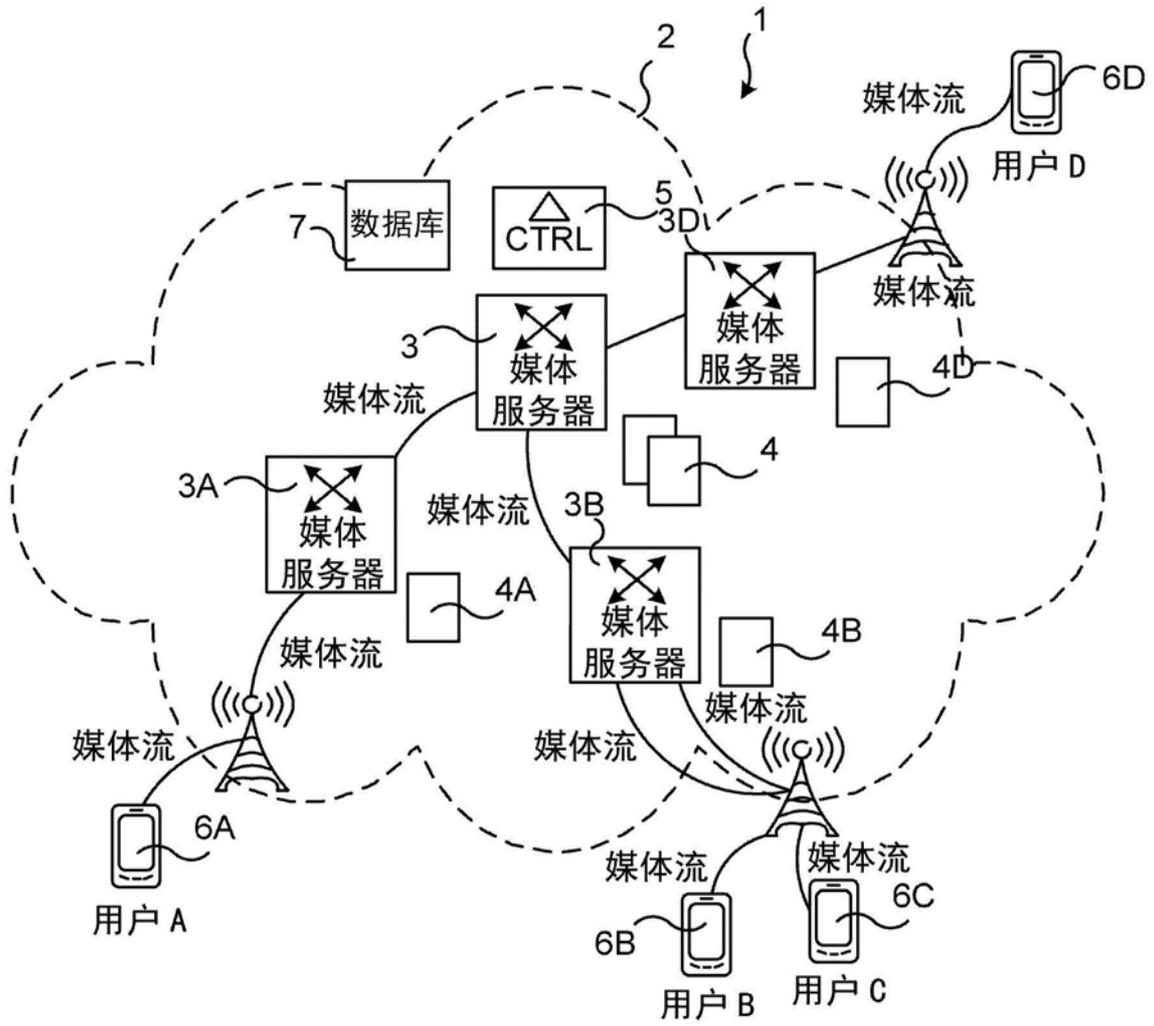


图1

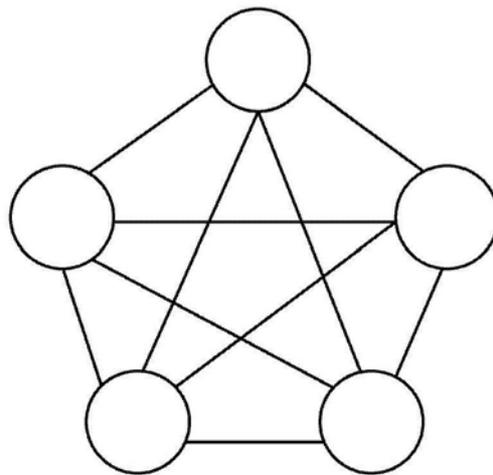


图2

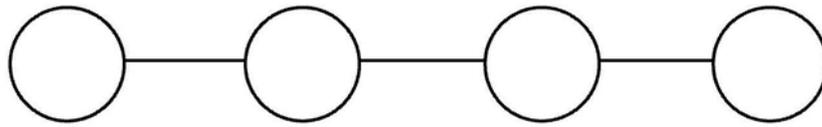


图3

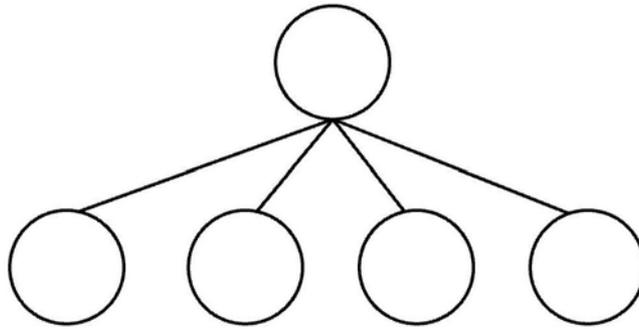


图4

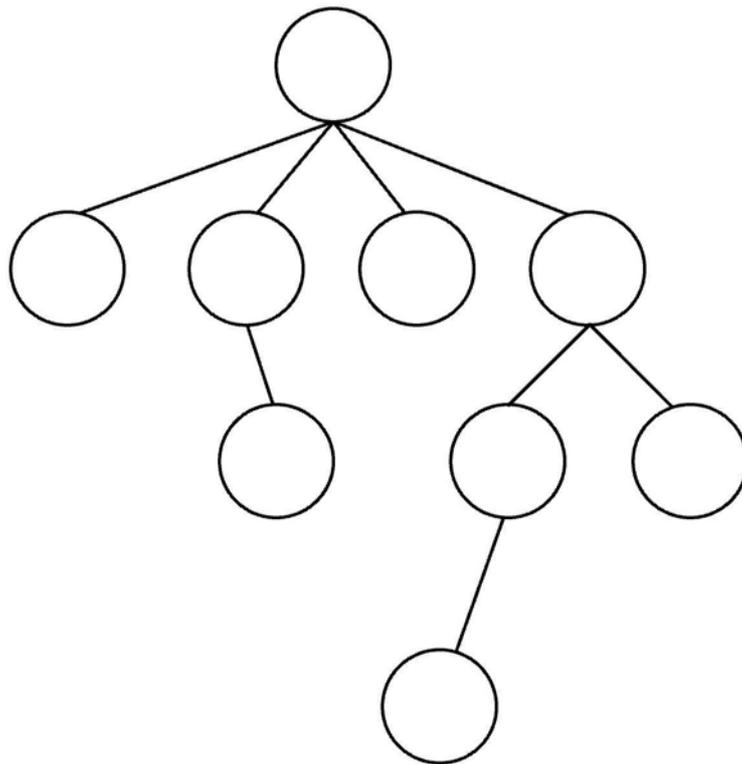


图5

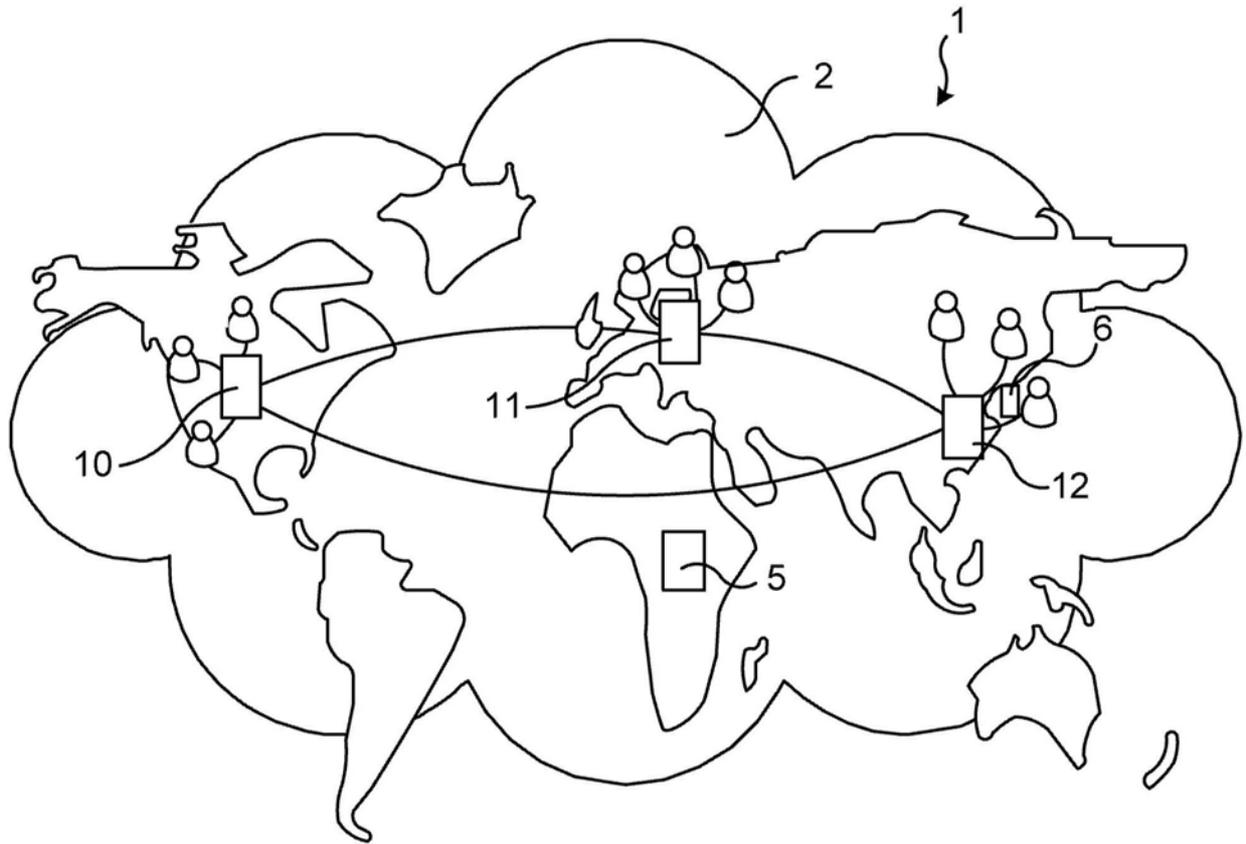


图6

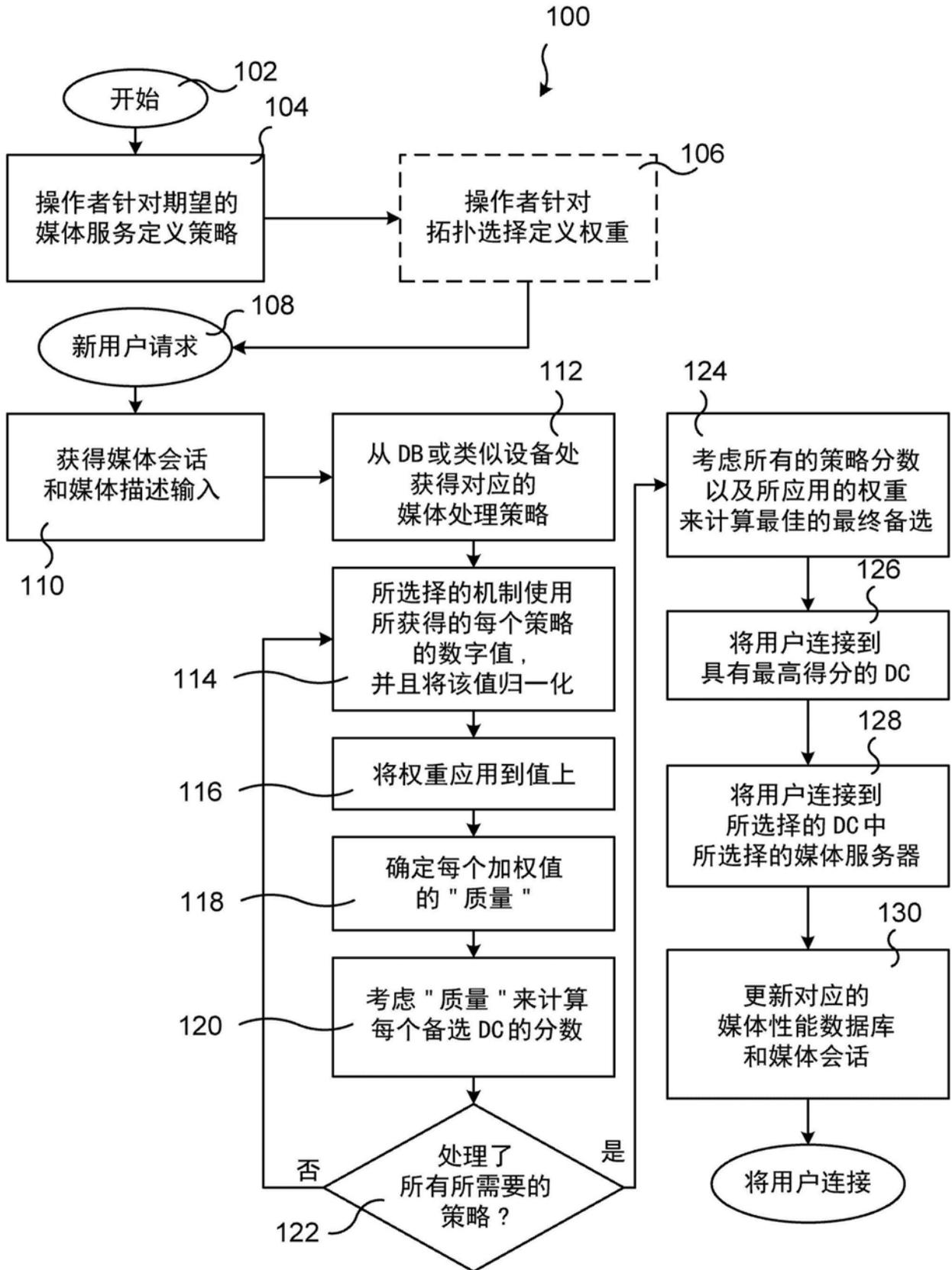


图7

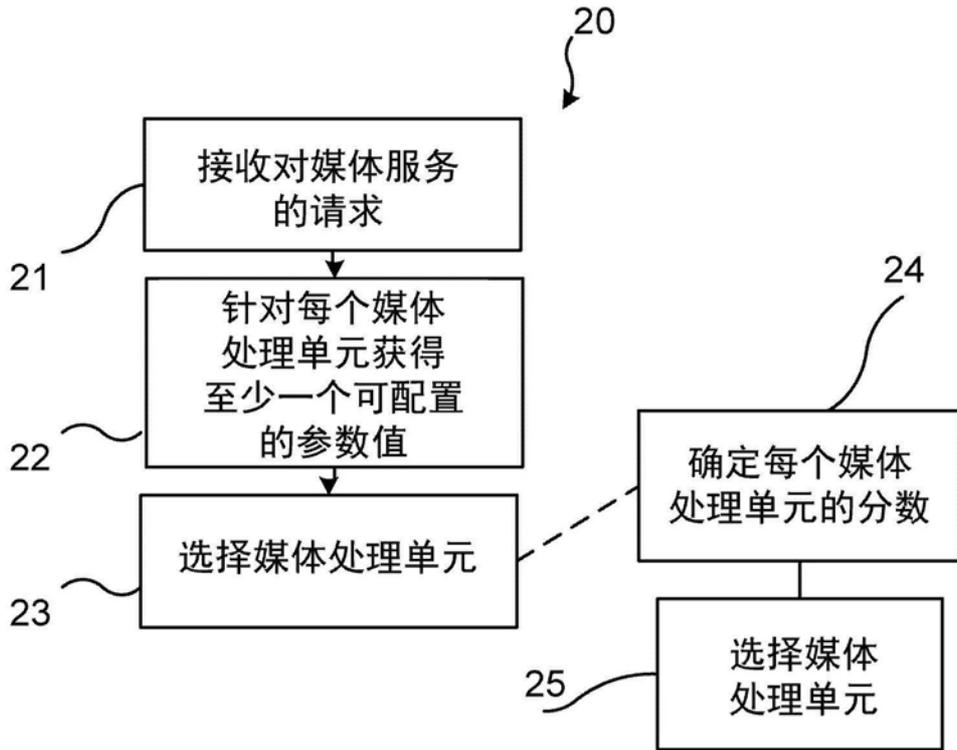


图8

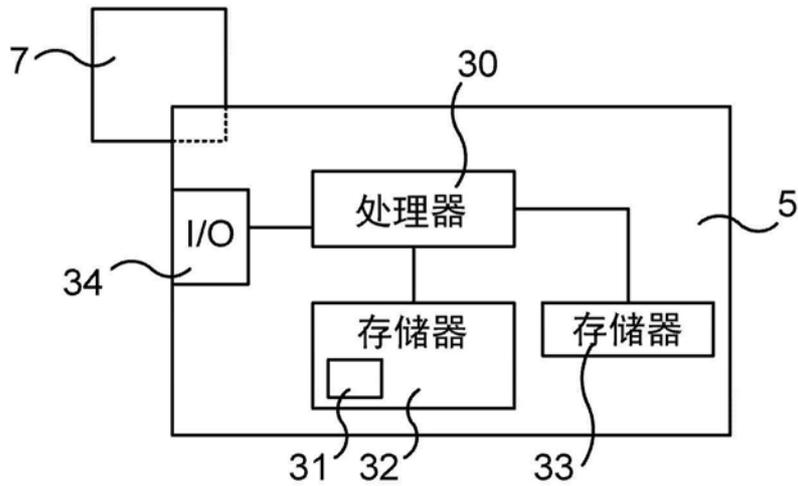


图9

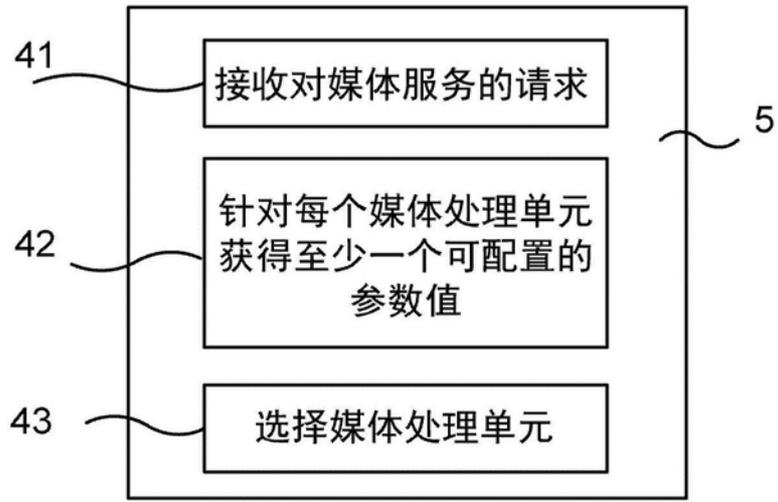


图10