

服务单元切换方法、系统及设备

本申请要求 2019 年 07 月 09 日递交的申请号为 201910614678.1、发明名称为“服务单元切换方法、系统及设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及互联网技术领域，尤其涉及一种服务单元切换方法、系统及设备。

背景技术

10 SLB（Server Load Balance，服务器负载均衡）产品的多个业务集群分布在全球多个国家，每个国家的多个地域，每个地域的多个机房内，这些集群构成一个复杂而且庞大的服务网络。每个集群是这个网络的一个服务单元。为了保障整个服务网络的稳定性，系统的服务调度及业务容灾能力异常关键。

15 目前，现有技术中，通常通过 DNS（Domain Name System，域名系统）域名解析切换映射成不同的 VIP（Virtual IP Address，虚拟服务地址），以实现业务系统的容灾切换及容量调度。

然而，在 DNS 域名解析切换时受 DNS 缓存影响，导致无法实时生效，进而在一定程度上影响整个服务网络的稳定性。

20 发明内容

鉴于上述问题，提出了本申请以提供一种解决上述问题或至少部分地解决上述问题的服务单元切换方法、系统及设备。

于是，在本申请的一个实施例中，提供了一种服务单元切换方法。该方法包括：
确定目标服务单元；

25 触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

在本申请的一个实施例中，提供了一种服务单元切换方法。该方法包括：
获取第一服务单元针对第一负载均衡任务的配置信息；

30 根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

在本申请的另一实施例中，提供了一种服务单元切换系统。该系统，包括：

第一服务单元、目标服务单元和分别与所述第一服务单元和所述目标服务单元连接的第一管控单元；

所述第一管控单元，用于确定目标服务单元；触发所述目标服务单元根据第一服务
5 单元针对第一负载均衡任务的配置信息发布第一路由信息；

所述目标服务单元，用于在所述第一管控单元的触发下，根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

在本申请的另一实施例中，提供了一种服务单元切换系统。该系统，包括：

第一服务单元、目标服务单元、与所述第一服务单元连接的第一管控单元以及与所述
10 述目标服务单元连接的目标管控单元；

所述第一管控单元，用于获取所述第一服务单元针对第一负载均衡任务的配置信息；将所述配置信息发送至所述目标管控单元；

所述目标管控单元，用于将所述配置信息转发至所述目标服务单元；触发所述目标服务单元根据所述配置信息发布第一路由信息；

所述目标服务单元，用于在所述目标管控单元的触发下，根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

在本申请的另一实施例中，提供了一种电子设备。该设备，包括：

存储器和处理器，其中，

所述存储器，用于存储程序；

所述处理器，与所述存储器耦合，用于执行所述存储器中存储的所述程序，以用于：
20 确定目标服务单元；

触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

在本申请的另一实施例中，提供了一种电子设备。该设备，包括：存储器和处理器，
25 其中，

所述存储器，用于存储程序；

所述处理器，与所述存储器耦合，用于执行所述存储器中存储的所述程序，以用于：
获取第一服务单元针对第一负载均衡任务的配置信息；

根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。
30

在本申请的一个实施例中，提供了一种服务单元切换方法。该方法包括：

确定目标服务单元；

触发所述目标服务单元根据第一服务单元针对第一任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

5 在本申请的另一实施例中，提供了一种电子设备。该设备，包括：存储器和处理器，其中，

所述存储器，用于存储程序；

所述处理器，与所述存储器耦合，用于执行所述存储器中存储的所述程序，以用于：

确定目标服务单元；

10 触发所述目标服务单元根据第一服务单元针对第一任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

在本申请的一个实施例中，提供了一种服务单元切换方法。该方法包括：

获取第一服务单元针对第一任务的配置信息；

15 根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

在本申请的另一实施例中，提供了一种电子设备。该设备，包括：存储器和处理器，其中，

所述存储器，用于存储程序；

所述处理器，与所述存储器耦合，用于执行所述存储器中存储的所述程序，以用于：

20 获取第一服务单元针对第一任务的配置信息；

根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

本申请实施例提供的技术方案中，先由第一服务单元为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

30

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

- 图 1 为本申请一实施例提供的服务单元切换系统的结构框图；
- 图 2 为本申请又一实施例提供的服务单元切换系统的结构框图；
- 图 3 为本申请另一实施例提供的服务单元切换方法的流程示意图；
- 图 4 为本申请实施例提供的服务单元切换方法的流程示意图；
- 10 图 5 为本申请实施例提供的信令图；
- 图 6 为本申请又一实施例提供的服务单元切换系统的结构框图；
- 图 7 为本申请一实施例提供的服务单元切换装置的结构框图；
- 图 8 为本申请又一实施例提供的服务单元切换装置的结构框图；
- 图 9 为本申请又一实施例提供的电子设备的结构框图；
- 15 图 10 为本申请又一实施例提供的网络架构图；
- 图 11 为本申请另一实施例提供的服务单元切换方法的流程示意图；
- 图 12 为本申请实施例提供的服务单元切换方法的流程示意图。

具体实施方式

- 20 现有技术中，对于通过域名提供服务的服务系统，可以通过 DNS 域名解析将同一域名映射成不同的 VIP，以实现业务系统的容灾切换及容量调度。

举例来说：针对服务域名 www.A.com，在 DNS 服务器中默认配置该域名映射到杭州机房 A 的 VIP-A，同时配置杭州机房 B 的 VIP-B，北京机房 C 的 VIP-C 作为映射备份。当 VIP-A 出现故障时，DNS 服务器将配置 www.A.com 映射至 VIP-B，以实现流量的容灾切换。

- 25 在当前 DNS 域名解析系统中通常采用 DNS 缓存机制，即用户端在正常访问某域名后，会将 DNS 服务器根据这个域名解析得到的 VIP 缓存起来。当用户端再次访问该域名时，用户端就会直接把本地的 DNS 缓存提取显示，而无需请求 DNS 服务器再次进行域名解析，这样可加速网址的解析。但是，用户端本地的 DNS 缓存只有在预设缓存时长到来时才会失效；失效后，用户端再次访问该域名时，才会再次请求 DNS 服务器进行域名
- 30

解析。也就是说，就算现有技术中更改了 DNS 服务器中的域名映射配置，也无法实现业务流量的实时切换，只有等到用户端本地的 DNS 缓存失效时，才能完成业务流量的切换。

此外，已建立的连接无法切换，这是因为已建立的连接是基于 TCP/IP 协议的，该连接对应的源地址（即访问请求端 IP 地址）、源端口（即访问请求端端口）、目标地址（即虚拟服务地址）、目标端口（即虚拟服务端口）都是固定的，无法改变的。现有技术中，DNS 域名映射切换时更改了域名对应的虚拟服务地址，但是无法改变已建立连接的目标地址，这也就致使已建立连接的无法继续使用，也就无法实现已建立连接的正常切换，存在已建立连接流量受损情况。

为了使本技术领域的人员更好地理解本申请方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述。显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

此外，在本申请的说明书、权利要求书及上述附图中描述的一些流程中，包含了按照特定顺序出现的多个操作，这些操作可以不按照其在本文中出现的顺序来执行或并行执行。操作的序号如 101、102 等，仅仅是用于区分各个不同的操作，序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。需要说明的是，本文中的“第一”、“第二”等描述，是用于区分不同的消息、设备、模块等，不代表先后顺序，也不限定“第一”和“第二”是不同的类型。

图 1 示出了本申请一实施例提供的服务单元切换系统的结构框图。如图 1 所示，该系统包括：第一服务单元 100、目标服务单元 200 和分别与所述第一服务单元 100 和所述目标服务单元 200 连接的第一管控单元 300。其中，

所述第一管控单元 300，用于确定目标服务单元 200；触发所述目标服务单元 200 根据第一服务单元 100 针对第一负载均衡任务的配置信息发布第一路由信息；

所述目标服务单元 200，用于在所述第一管控单元 300 的触发下，根据所述配置信息发布第一路由信息，以接替所述第一服务单元 100 为所述第一负载均衡任务提供流量分发服务。

在本实施例中，第一管控单元 300 可通过发送控制指令给所述第一服务单元 100 和所述目标服务单元 200 实现对所述第一服务单元 100 和所述目标服务单元 200 的管控。

本申请实施例提供的技术方案中，先由第一服务单元为第一负载均衡任务提供流量

分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，

5 采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

本申请实施例提供的服务单元切换系统中各组成单元，如第一服务单元、目标服务单元和第一管控单元的具体工作流程及之间的信令交互将在以下各实施例中作进一步的说明。

10

图 2 示出了本申请又一实施例提供的服务单元切换系统的结构框图。如图 2 所示，该系统包括：第一服务单元 100、目标服务单元 200、与所述第一服务单元 100 连接的第一管控单元 300 以及与所述目标服务单元 200 连接的目标管控单元 400；

所述第一管控单元 300，用于获取所述第一服务单元 100 针对第一负载均衡任务的配置信息；将所述配置信息发送至所述目标管控单元 400；

15

所述目标管控单元 400，用于将所述配置信息转发至所述目标服务单元 200；触发所述目标服务单元 200 根据所述配置信息发布第一路由信息；

所述目标服务单元 200，用于在所述目标管控单元 400 的触发下，根据所述配置信息发布第一路由信息，以接替所述第一服务单元 100 为所述第一负载均衡任务提供流量分发服务。

20

在本实施例中，第一服务单元 100 和目标服务单元 200 分别由第一管控单元 300 和目标管控单元 400 来管控。

本申请实施例提供的技术方案中，先由第一服务单元为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，

25 采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

进一步的，如图 6 所示，上述系统，还包括：分别与多个管控单元连接的全局调度

30

单元 10；所述多个管控单元中包括所述第一管控单元 300 和所述目标管控单元 400；

所述全局调度单元 10，用于发送待迁移信息至所述第一管控单元 300，以由所述第一管控单元 300 根据所述待迁移信息，确定出所述第一服务单元 100、所述第一负载均衡任务以及所述目标服务单元 200；

- 5 所述待迁移信息为所述全局调度单元 10 根据多个服务单元的运行情况确定的；所述多个服务单元中包括所述第一服务单元 100 和所述目标服务单元 200；

所述待迁移信息中包括所述第一服务单元为所述第一负载均衡任务配置的虚拟服务地址以及所述目标服务单元 200 的服务地址。

- 10 本申请实施例提供的服务单元切换系统中各组成单元，如第一服务单元、目标服务单元、第一管控单元和目标管控单元、全局调度单元的具体工作流程及之间的信令交互将在以下各实施例中作进一步的说明。

- 图 3 示出了本申请一实施例提供的通信方法的流程示意图。本方法的执行主体可以为图 1 中服务单元切换系统的第一管控单元 300，还可以为图 2 中的第一管控单元 300 或目标管控单元 400。如图 3 所示，该方法包括：
- 15

101、确定目标服务单元。

102、触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

- 20 上述 101 中，在整个负载均衡网络中，可能会存在多个服务单元，可从多个服务单元中除第一服务单元以外的其他服务单元中确定出目标服务单元，例如：随机选取一服务单元或选取当前所承担的总流量低于流量阈值的服务单元作为目标服务单元。

其中，目标服务单元可以为一个第二服务节点，或由多个第二服务节点组成的集群。需要说明的是，集群中的每一个第二服务节点能够为第一负载均衡任务提供相同的流量分发服务功能。其中，一个第二服务节点可理解为一台真实服务器。

- 25 上述 102 中，第一服务单元针对第一负载均衡任务的配置信息中可包括：所述第一服务单元针对所述第一负载均衡任务配置的虚拟服务地址 VIP。第一负载均衡任务指的是将针对该虚拟服务地址的业务流量分发至该虚拟服务地址对应的多个后端服务器进行处理的任务。其中，流量分发服务指的即是将针对该虚拟服务地址的业务流量分发至该虚拟服务地址对应的多个后端服务器上。该后端服务器指的是用来处理针对该虚拟服务地址的服务请求得到最终处理结果的真实服务器。
- 30

在一种可实现的方案中，上述 102 中“触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息”，具体为：

1021、触发所述目标服务单元根据所述虚拟服务地址发布第一路由信息，使得针对所述虚拟服务地址的服务请求寻址到所述目标服务单元。

5 目标服务单元根据虚拟服务地址发布第一路由信息，也即是针对虚拟服务地址发布第一路由信息。

其中，第一路由信息与所述虚拟服务地址相对应，该第一路由信息可使得针对所述虚拟服务地址的服务请求寻址到所述目标服务单元，以实现目标服务单元接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

10 具体应用时，上述虚拟服务地址与第一域名对应；当用户端设备需要访问第一域名时，用户端设备会先请求 DNS 服务器对第一域名进行解析，DNS 服务器将第一域名解析成上述虚拟服务地址，并将第一域名和上述虚拟服务地址的对应关系返回给用户端设备进行 DNS 缓存。同时，用户端设备将发送针对上述虚拟服务地址的服务请求，该服务请求可通过目标服务单元发布的第一路由信息寻址到目标服务单元。后续，若用户端还
15 想继续访问第一域名，用户端设备将根据 DNS 缓存信息，再次发送针对上述虚拟服务地址的服务请求。本实施例中，在由目标服务单元接替第一服务单元提供第一负载均衡任务提供流量分发服务后，由于目标服务单元也是针对上述虚拟服务地址发布的第一路由信息，用户端设备根据 DNS 缓存内容继续发送针对上述虚拟服务地址的服务请求，该服务请求是能够寻址到目标服务单元。

20 在实际应用时，目标服务单元可根据第一服务单元针对第一负载均衡任务的配置信息在本地完成所述第一负载均衡任务的配置，即目标服务单元针对第一负载均衡任务的配置信息与第一服务单元针对第一负载均衡任务的配置信息是一致的，也即是目标服务单元针对第一负载均衡任务配置的虚拟服务地址与第一服务单元针对第一负载均衡任务配置的虚拟服务地址是一样的。

25 通常，该配置信息中还可包括：虚拟服务端口、调度算法，与所述虚拟服务地址对应的多个后端服务器的服务地址，所述多个后端服务器端口以及有关所述后端服务器的健康检测配置信息中的一项或多项。目标服务单元可根据该健康检测配置信息执行对多个后端服务器的健康检测。目标服务单元还可根据调度算法将与所述第一负载均衡任务相关的业务流量分发至与所述虚拟服务地址对应的多个后端服务器上，即根据调度算法
30 将针对上述虚拟服务地址的服务请求分发至与所述虚拟服务地址对应的多个后端服务器

上。

其中，第一服务单元可以为一个第一服务节点，或由多个第一服务节点组成的集群。需要说明的是，集群中的每一个第一服务节点能够为第一负载均衡任务提供相同的流量分发服务功能。其中，一个第一服务节点可理解为一台真实服务器。

- 5 在本实施例中，将第一负载均衡任务从第一服务单元转移至目标服务单元，并没有改变第一负载均衡任务对应的虚拟服务地址。这样，就算用户端存在 DNS 缓存，也不影响业务流量的实时切换。

在实际应用中，在目标服务单元发布第一路由信息之前，第一服务单元就已经针对所述虚拟服务地址发布了第二路由信息。为了确保针对所述虚拟服务地址的服务请求寻址到目标服务单元，而不是寻址到第一服务单元，可采用如下两种方式来实现：

方式一：上述 1021 中“触发所述目标服务单元根据所述虚拟服务地址发布第一路由信息”，具体可采用如下步骤来实现：

S11、获取所述第一服务单元针对所述虚拟服务地址已发布的第二路由信息；

- 15 S12、将所述第二路由信息发送至所述目标服务单元，以触发所述目标服务单元根据所述虚拟服务地址及所述第二路由信息，发布优先级高于所述第二路由信息的第一路由信息。

即第一路由信息的优先级需高于所述第一服务单元针对所述虚拟服务地址已发布的第二路由信息。

- 在一实例中，上述第一路由信息可以为 BGP（Border Gateway Protocol，边界网关协议）路由信息，可通过设置第一路由信息中的 as-path 属性，例如：在第一路由信息中设置短于第二路由信息中的第一 as-path 属性的第二 as-path 属性，实现第一路由信息的优先级高于第二路由信息的优先级。

- 25 在另一实例中，第一路由信息中包括有关所述虚拟服务地址的网段信息，第一路由信息中的网段信息的粒度细于第二路由信息中的网段信息的粒度。例如：第二路由信息中的网段信息为 1.1.1.0/24，第一路由信息中的网段信息为 1.1.1.0/25 和 1.1.1.128/25。其中，网段中符号“/”后的数字越大，粒度越细，路由信息的优先级越高。

其中，目标服务单元内部所有第二服务节点均发布相同优先级的第一路由信息，使业务流量在第二服务节点内部通过 ECMP（Equal-Cost Multipath Routing，等价多路径）等价路由实现负载均衡。

- 30 在实际应用时，还可采用其他方式来实现第一路由信息的优先级高于第二路由信息

的优先级，本申请实施例对此不做具体限定。

其中，目标服务单元发布第一路由信息，可理解为目标服务单元向前一级网络设备发送第一路由信息。

方式二：在目标服务单元发布第一路由信息之前，触发所述第一服务单元撤销第二
5 路由信息。

在方式二中，容易在第二路由信息撤销后以及第一路由信息发布前这一段时间内出现相关服务请求包的丢失或遗漏问题。故在实际应用中，具体可采用方式一，在方式一中第二路由信息不用撤销或可在第一路由信息发布后再撤销，这样可实现无缝切换，避免了服务请求包的丢失或遗漏等问题。本申请实施例提供的技术方案中，先由第一服务
10 单元为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够确保与第一负载
15 均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

由于在目标服务单元接替第一服务单元提供有关第一负载均衡任务的流量分发服务之前，第一服务单元就已提供过有关第一负载均衡任务的流量分发服务，也就是说，通过第一服务单元已建立了一些连接。这些连接对应的源地址（即访问请求端 IP 地址）、
20 源端口（即访问请求端端口）、目标地址（即虚拟服务地址）、目标端口（即虚拟服务端口）等属性，在上述各实施例中由目标服务单元接替第一服务单元前后，并没有发生过变化，因此，这些已建立连接可正常使用。

这些已建立连接可理解为访问请求端（例如用户客户端）通过第一服务单元与后端服务器已建立的连接。通常，访问请求端需要针对一项功能需求（例如：计算需求）连续发送多个访问请求，这多个访问请求应该发送给同一个后端服务器，由同一个后端服务器进行相关处理才能够保证完成该项功能需求。若这多个访问请求分别发送给不同的
25 后端服务器，会致使该功能需求无法实现。为了确保那些已建立连接的成功切换，需要将第一服务单元上记录的有关第一负载均衡任务的会话信息发送给目标服务单元，这样目标服务单元在对有关已建立连接的服务请求进行分发时就可以进行参考，以将有关已
30 建立连接的服务请求分发至正确的后端服务器上，以避免已建立连接的带宽资源受损。

具体地，上述方法，还可包括：

103、向所述第一服务单元发送会话发送指令，以触发所述第一服务单元将其记录的
有关所述第一负载均衡任务的会话信息发送至所述目标服务单元，以供所述目标服务单
元在对有关已建立连接的服务请求进行分发时参考。

- 5 其中，所述会话信息中包括所述已建立连接的连接信息；所述已建立连接是通过所
述第一服务单元建立的。其中，会话发送指令具体可以为会话同步指令，以触发第一服
务单元将会话信息同步至目标服务单元。

在一实例中，已建立连接的连接信息中可包括：有关已建立连接的访问请求端 IP 地
址、访问请求端口、虚拟服务地址、虚拟服务端口、后端服务器 IP 地址以及后端服务
10 器服务端口等。

这样，后续目标服务单元接收到访问请求时，若该访问请求中包括的访问请求端 IP
地址位于会话信息中，则将该访问请求分发至会话信息中与该访问请求端 IP 地址对应记
录的后端服务器 IP 地址对应的后端服务器。这样，就可将有关已建立连接的服务请求分
发至正确的后端服务器上，以避免已建立连接的带宽资源受损。

- 15 具体实施时，在图 1 所示的服务单元切换系统架构中，目标服务单元完成第一负载
均衡任务的相关配置后，可向第一管控单元发送配置完成信息；第一管控单元接收到配
置完成信息后，向第一服务单元发送会话发送指令。

在图 2 所示的服务单元切换系统架构中，目标服务单元完成第一负载均衡任务的相关
配置后，向目标管控单元发送配置完成信息；目标管控单元向第一管控单元转发配置
20 完成信息；第一管控单元接收到配置完成信息后，向第一服务单元发送会话发送指令。

- 当所述目标服务单元为由多个第二服务节点组成的集群时，上述 103 中“向所述第
一服务单元发送会话发送指令，以触发所述第一服务单元将其记录的有关所述第一负载
均衡任务的会话信息发送至所述目标服务单元”，具体为：向所述第一服务单元发送会
话发送指令，以触发所述第一服务单元将其记录的有关所述第一负载均衡任务的会话信
25 息发送至所述多个第二服务节点中的目标服务节点，并由所述目标服务节点将所述会话
信息同步至所述多个第二服务节点中的其他第二服务节点。

通常第一服务单元和目标服务单元不在一个组播域中，需要通过单播的方式来发送。
目标服务单元中的多个第二服务节点位于同一个组播域中，故可通过组播的方式进行同
步。具体地：第一服务单元可通过三层单播网络将其记录的有关所述第一负载均衡任务
30 的会话信息发送至所述多个第二服务节点中的目标服务节点。所述目标服务节点可通过

三层组播网络将所述会话信息同步至所述多个第二服务节点中的其他第二服务节点。

进一步的，上述 102 中“触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息”，具体可采用如下步骤来实现：

S21、获取第一服务单元针对第一负载均衡任务的配置信息。

- 5 S22、将所述配置信息发送至目标服务单元，以由所述目标服务单元根据所述配置信息完成配置，并根据所述配置信息发布所述第一路由信息。

在本实施例中，无需事先在多个服务单元上针对同一域名重复配置负载均衡任务，可避免现有技术中针对同一域名需要事先在多个不同的机房上以多个不同 VIP 进行重复配置负载均衡任务，占用配置空间，浪费资源等技术问题。

- 10 当服务单元切换系统的结构如图 1 所示的那样，则第一管控单元可直接将该配置信息发送给目标服务单元。

当服务单元切换系统的结构如图 2 所示的那样，则上述 S22 中“将所述配置信息发送给所述目标服务单元”，具体为：将所述配置信息发送至所述目标服务单元对应的目标管控单元，以由所述目标管控单元将所述配置信息转发至所述目标服务单元。

- 15 需要补充的是，当目标服务单元为由多个第二服务节点组成的集群时，第一管控单元或目标管控单元需要将该配置信息发送给多个第二服务节点中每一个第二服务节点。

在实际应用中，可通过人工的方式来触发服务单元之间的切换，例如：工作人员在查看到某个服务单元超负荷时，可向该服务单元对应的管控单元发送切换指令，以实现由另一个服务单元来接替该服务单元为该服务单元上的某一负载均衡任务提供流量分发
20 服务。在一种自动化的应用场景中，可根据整个负载均衡网络中的所有服务单元的运行情况来确定是否需要执行切换操作。具体地，上述方法，还可包括：

104、根据待迁移信息，确定出所述第一服务单元、所述第一负载均衡任务以及所述目标服务单元。

- 25 其中，所述待迁移信息为根据检测到的多个服务单元的运行情况确定的；所述多个服务单元中包括所述第一服务单元和所述目标服务单元；所述待迁移信息中包括所述第一服务单元针对所述第一负载均衡任务配置的虚拟服务地址以及所述目标服务单元的服务地址（即目标服务单元的 IP 地址）。

- 30 第一管控单元根据待迁移信息中的虚拟服务地址，即可确定出配置有该虚拟服务地址的第一服务单元以及第一负载均衡任务，根据待迁移信息中的服务地址即可确定出目标服务单元。

待迁移信息可以由第一管控单元自己检测到的，或者由全局调度单元根据检测到的所述多个服务单元的运行情况确定的。具体地，上述方法，还可包括：

105、接收全局调度单元发送的其根据检测到的所述多个服务单元的运行情况确定的所述待迁移信息。

5 其中，所述运行情况中包括：当前所承担的总流量、总连接成功率和/或异常报警信息。

例如：全局调度单元可将当前所承担的总流量超过预设阈值的服务单元确定为第一服务单元；并将第一服务单元上流量占比最大的负载均衡任务确定为第一负载均衡任务；将当前所承担的总流量低于预设阈值的服务单元确定为目标服务单元。再例如：全局调度单元可将当前总连接成功率低于第一预设成功率的服务单元确定为第一服务单元，并将第一服务单元上连接成功率最低的负载均衡任务确定为第一负载均衡任务；将当前所承担的总流量低于预设阈值且总连接成功率高于第二预设成功率的服务单元确定为目标服务单元。又例如：全局调度单元可将发出异常报警信息的服务单元确定为第一服务单元。

15 全局调度单元可将所述第一服务单元针对确定出的第一负载均衡任务配置的虚拟服务地址以及目标服务单元的服务地址作为待迁移信息。

在目标服务单元接替第一服务单元提供第一负载均衡任务的流量分发服务后，第一服务单元即可撤销其针对第一负载均衡任务已发布的第二路由信息。具体地，上述方法，还可包括：

20 106、接收到有关所述第一路由信息的已发布通知后，触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

其中，已发布通知中可携带有上述虚拟服务地址。

当上述方法的执行主体为图 1 中的第一管控单元时，第一管控单元接收到目标服务单元发送来的已发布通知后，向所述第一服务单元发送路由撤销指令，以触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

当上述方法的执行主体为图 2 中的第一管控单元时，第一管控单元接收到目标管控单元转发来的已发布通知后，向所述第一服务单元发送路由撤销指令，以触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

当上述方法的执行主体为图 2 中的目标管控单元时，目标管控单元接收到目标服务单元发送来的已发布通知后，向第一管控单元转发所述已发布通知，以使所述第一管控

单元接收到所述已发布通知后，触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

此外，考虑到每一个服务单元所能承载的配置数量是有限制的，为了不影响第一服务单元对其他新的负载均衡任务的配置，可触发第一服务单元删除有关第一负载均衡任务的配置。具体地，上述方法，还可包括：

107、接收到有关所述第一路由的已发布通知信息后，触发所述第一服务单元删除有关所述第一负载均衡任务的配置。

当上述方法的执行主体为图 1 中的第一管控单元时，第一管控单元接收到目标服务单元发送来的已发布通知后，向所述第一服务单元发送配置删除指令，以触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

当上述方法的执行主体为图 2 中的第一管控单元时，第一管控单元接收到目标管控单元转发来的已发布通知后，向所述第一服务单元发送配置删除指令，以触发所述第一服务单元删除有关所述第一负载均衡任务的配置。

当上述方法的执行主体为图 2 中的目标管控单元时，目标管控单元接收到目标服务单元发送来的已发布通知后，向第一管控单元转发所述已发布通知，以使所述第一管控单元接收到所述已发布通知后，触发所述第一服务单元删除有关所述第一负载均衡任务的配置。

进一步的，上述方法，还可包括：

108、接收所述目标服务单元针对所述第一负载均衡任务检测得到的检测信息。

109、根据所述检测信息，判断所述目标服务单元是否满足接替条件。

110、所述目标服务单元满足所述接替条件时，执行所述触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息的步骤。

当上述方法的执行主体为图 1 中的第一管控单元时，第一管控单元可直接接收目标服务单元发送的所述检测信息。第一管控单元可通过向目标服务单元发送路由发布指令，以触发所述目标服务单元根据所述配置信息发布第一路由信息。

当上述方法的执行主体为图 2 中的目标管控单元时，目标管控单元可直接接收目标服务单元发送的所述检测信息。目标管控单元可通过向目标服务单元发送路由发布指令，以触发所述目标服务单元根据所述配置信息发布第一路由信息。

在本实施例中，根据所述检测信息，判断所述目标服务单元是否满足接替条件，可确保目标服务单元接替第一服务单元之后，能够正常提供流量分发服务，确保切换的有

效性。

需要说明的是，不同的服务单元对多个后端服务器的健康检测结果很可能会存在差异，导致差异的原因可能在于服务单元中有关后端服务器的检测接口是否正常。若检测接口存在问题，目标服务单元的检测结果显示后端服务器的健康状态为异常状态。这样，目标服务单元在后续进行流量分发时，会自动忽略异常状态的后端服务器，一旦自动忽略的后端服务器的数量较大，会导致：在目标服务单元接替第一服务单元之后，针对第一负载均衡任务，负载均衡网络基于目标服务单元难以提供可靠的负载均衡服务。

所述检测信息中可包括：所述目标服务单元针对多个后端服务器进行检测得到的第一健康状态信息。其中，所述多个后端服务器用于处理与所述第一负载均衡任务相关的服务请求。与第一负载均衡任务相关的服务请求也即是针对第一服务单元针对第一负载均衡任务配置的虚拟服务地址的服务请求。在上述实施例中介绍过：上述配置信息中包括：后端服务器的健康检测配置信息，故目标服务单元可根据该健康检测配置信息来检测与所述多个后端服务器的健康状态。

在一种可实现的方案中，上述 109 中“根据所述检测信息，判断所述目标服务单元是否满足接替条件”，具体为：

1091、根据所述第一健康状态信息，判断所述目标服务单元是否满足接替条件。

在一实例中，第一健康状态信息显示所述多个后端服务器中超过预设数量个的后端服务器的健康状态为正常状态，则可判定目标服务单元满足接替条件。例如：多个后端服务器为 10 个，预设数量个为 8 个。

在另一实例中，获取所述第一服务单元针对所述多个后端服务器进行检测得到的第二健康状态信息；所述第一健康状态检测信息与所述第二健康状态检测信息之间的差异满足第一预设条件时，判定所述目标服务单元满足所述接替条件。

在上述实施例中介绍过：上述配置信息中包括：后端服务器的健康检测配置信息，故第一服务单元可根据该健康检测配置信息来检测与所述第一负载均衡任务相关的多个后端服务器的健康状态。

举例来说：第一健康状态检测信息显示多个后端服务器中 n 个后端服务器正常，第二健康状态检测信息显示多个后端服务器中 m 个后端服务器正常， m 和 n 的差值的绝对值小于预设差值时，判定所述第一健康状态检测信息与所述第二健康状态检测信息之间的差异满足第一预设条件。

这样，针对第一负载均衡任务，负载均衡网络基于目标服务单元可以提供与基于第

一服务单元相同且可靠的负载均衡服务。

进一步的，所述目标服务单元为由多个第二服务节点组成的集群；所述检测信息中还包括：所述多个第二服务节点之间有关所述第一负载均衡任务的会话信息的同步量。所述会话信息是由所述第一服务单元在为所述第一负载均衡任务提供流量分发服务时记录的。上述 1091 中“根据所述第一健康状态信息，判断所述目标服务单元是否满足接替条件。”，具体为：

S31、根据所述第一健康状态信息和所述同步量，判断所述目标服务单元是否满足所述接替条件。

会话信息的同步量可理解为已建立连接的同步量，例如：已建立连接的总数为 1000 个，当前已同步完成的已建立连接的数量为 900 个，则同步量即为 900。

上述 S31 中“根据所述第一健康状态信息和所述同步量，判断所述目标服务单元是否满足所述接替条件”，具体可采用如下步骤来实现：

a、获取所述第一服务单元针对所述多个后端服务器进行检测得到的第二健康状态信息。

b、判断所述第一健康状态检测信息与所述第二健康状态检测信息之间的差异是否满足第一预设条件。

c、确定所述会话信息的同步量在所述会话信息的总量中的占比。

d、所述差异满足第一预设条件且所述占比大于预设占比时，则判定所述目标服务单元满足所述接替条件。

其中，预设占比可以设为 95%，本申请实施例对此不做具体限定。

上述步骤 a 和 b 的具体实现可参见上述实施例中相应内容，在此不再赘述。占比大于预设占比，以确保大部分已建连接的流量能够正常切换。

综上所述，通过发布更高优先级的路由实现流量在两个服务单元之间的动态切换，通过三层单播网络和三层组播网络的方案实现负载均衡会话信息在负载均衡网络之间的同步。本方案解决了现有技术中 DNS 域名切换方案的两个缺点。首先，可保证流量切换的实时性，目标服务单元的第一路由发布后，核心网络会通过路由协议自动学习到新的路由路径并进行切换。其次，本方案切换时由于虚拟服务地址 VIP 不发生变化，并且通过三层网络同步了所有的会话信息，保证了已建立连接的正常转发，且新建连接也不会受到任何影响，实现了负载均衡系统在机房间业务流量的无感知、无中断迁移。

图 4 示出了本申请又一实施例提供的服务单元切换方法的流程示意图。本申请实施例提供的方法的执行主体可以为上述各实施例中的目标服务单元。该方法，包括：

501、获取第一服务单元针对第一负载均衡任务的配置信息。

502、根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

基于图 1 所示的服务单元切换系统架构，第一管控单元可从本地存储中或从第一服务单元处获取第一服务单元针对第一负载均衡任务的配置信息，并将配置信息发送给目标服务单元。

基于图 2 所示的服务单元切换系统架构，第一管控单元可从本地存储中或从第一服务单元处获取第一服务单元针对第一负载均衡任务的配置信息，并将配置信息发送给目标管控单元，目标管控单元将配置信息发送给目标服务单元。

上述 501 和 502 的具体实现过程可参见上述各实施例中相应内容，在此不再赘述。

本申请实施例提供的技术方案中，先由第一服务单元为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

进一步的，上述 502 中“根据所述配置信息发布第一路由信息”，具体可采用如下步骤来实现：

5021、获取所述第一服务单元针对所述第一负载均衡任务已发布的第二路由信息。

5022、根据所述配置信息和所述第二路由信息，发布优先级高于所述第二路由信息的第一路由信息。

基于图 1 所示的服务单元切换系统架构，目标服务单元可从第一管控单元处获取第二路由信息。

基于图 2 所示的服务单元切换系统架构，第一管控单元可将第二路由信息发送给目标管控单元，目标管控单元将第二路由信息转发给目标服务单元。

上述步骤 5021 和 5022 的具体实现可参见上述各实施例中相应内容，在此不再赘述。

进一步的，上述方法，还可包括：

503、接收所述第一服务单元发送来的有关所述第一负载均衡任务的会话信息。

504、将所述会话信息同步至集群内部的其他第二服务节点。

其中，所述集群中每一个第二服务节点均用于为所述第一负载均衡任务提供流量分发服务。

5 在本实施例中，目标服务单元为由多个第二服务节点组成的集群，上述方法的执行主体为所述多个第二服务节点中的一个第二服务节点。

此外，需要说明的是，在上述各实施例中第一服务单元也可以为由多个第一服务节点组成的集群。多个第一服务节点通过会话同步方式，每一服务节点上均存储有有关所述
10 所述第一负载均衡任务的会话信息。多个第一服务节点中的一个第一服务节点将有关所述
第一负载均衡任务的会话信息发送给多个第二服务节点中的一个目标服务节点即可。该
目标服务节点将会话信息同步给多个第二服务节点中的其他第二服务节点，例如：该目
标服务节点可通过三层组播网络将会话信息同步给多个第二服务节点中的其他第二服
务节点。

进一步的，上述方法，还可包括：

15 505、向目标管控单元发送针对所述第一负载均衡任务检测得到的检测信息，以由所述目标管控单元根据所述检测信息确定是否满足接替条件。

506、接收到所述目标管控单元在确定出满足接替条件时发送的接替指令后，执行所述根据所述配置信息发布第一路由信息的步骤。

上述步骤 505 和 506 的具体实现可参见上述各实施例中的相应内容，在此不再赘述。

20 这里需要说明的是：本申请实施例提供的所述方法中各步骤未尽详述的内容可参见上述实施例中的相应内容，此处不再赘述。此外，本申请实施例提供的所述方法中除了上述各步骤以外，还可包括上述各实施例中其他部分或全部步骤，具体可参见上述各实施例相应内容，在此不再赘述。

25 下面将结合图 5 示出的基于图 2 示出的服务单元切换系统架构的信令图来介绍本申请实施例提供的技术方案：

601、第一管控单元向目标管控单元发送第一服务单元针对第一负载均衡任务的配置信息。

30 602、目标管控单元将该配置信息下发至目标服务单元内所有物理服务器（即第二服务节点）。

603、第一管控单元向第一服务单元下发有关第一负载均衡任务的会话发送指令。

604、第一服务单元通过三层单播网络向目标服务单元发送有关第一负载均衡任务的会话信息。

605、目标服务单元在集群内部通过三层组播网络进行会话同步。

5 其中，目标服务单元由多台物理服务器（即第二服务节点）组成，通过会话同步使每台物理服务器上均具有有关第一负载均衡任务的全量连接信息，即有关第一负载均衡任务对应的VIP的全量连接信息。

606、目标服务单元向目标管控单元反馈有关迁移第一负载均衡任务的健康检查状态及同步连接数量（即同步量）。

10 607、目标管控单元判断是否达到接替条件；判断条件达到时，给目标服务单元下发路由发布指令。

608、目标服务单元通过发布相较第一服务单元更高优先级的BGP路由信息将流量引至该服务节点。

15 其中，目标服务单元内部所有物理服务器发布相同的BGP路由信息，使流量在服务节点内部通过ECMP等价路由实现负载均衡。

609、第一管控单元向第一服务单元下发有关第一负载均衡任务的路由撤销指令。

610、第一服务单元撤销有关第一负载均衡任务的BGP路由信息。

611、第一管控单元向第一服务单元下发有关第一负载均衡任务的配置删除指令。

612、第一服务单元删除有关第一负载均衡任务的配置。

20 在本实施例中，目标服务单元替代第一服务单元提供服务，通过同步的连接信息，保证已建立的连接通过负载均衡系统时实现正确的转换，将流量发送到正确的后端服务器去。

25 图10示出了本申请实施例提供的一种网络架构图。下面将结合图10对本申请实施例提供的切换方法进行详细介绍：

一开始，由第一负载均衡设备1202（即上述第一服务单元）针对VIP向互联网发布第二路由信息。用户端设备1201发送域名解析请求至DNS服务器1210，DNS服务器1210对域名进行解析得到上述VIP，并将VIP返回给用户端设备1201进行DNS缓存。用户端设备1201根据DNS缓存向互联网发送携带有该VIP的服务请求。由于当前第一
30 负载均衡设备1202针对VIP向互联网发布了第二路由信息，故该服务请求能够寻址到第

一负载均衡设备 1202。由第一负载均衡设备 1202 根据调度算法将该服务请求分发至第一后端服务器 1208、第二后端服务器 1207 和第三后端服务器 1206 中的一个进行服务请求的处理。

第一管控设备 1204（即上述第一管控单元）可向全局调度设备 1209（即上述全局调度单元）上报第一负载均衡设备 1202 的运行情况，目标管控设备 1205（即上述目标管控单元）可向全局调度设备 1209 上报第二负载均衡设备 1203（即上述目标服务单元）的运行情况，全局调度设备 1209 可根据第一管控设备 1204 和目标管控设备 1205 的上报信息，确定是否需要进行第一负载均衡设备和第二负载均衡设备的切换。若需要，全局调度设备 1209 可通知第一管控设备 1204，以由第一管控设备 1204 将第一负载均衡设备 1202 针对 VIP 的配置信息发送给目标管控设备 1205，目标管控设备 1205 将该配置信息下发给第二负载均衡设备 1203。第二负载均衡设备 1203 根据该配置信息进行配置，并针对该 VIP 向互联网发布第一路由信息，第一路由信息的优先级高于第二路由信息。

互联网后续再次接收到用户端设备 1201 发送来的携带有该 VIP 的服务请求时，该服务请求会寻址到第二负载均衡设备 1203，第二负载均衡设备 1203 根据调度算法将该服务请求分发至第一后端服务器 1208、第二后端服务器 1207 和第三后端服务器 1206 中的一个进行服务请求的处理。

由此可见，在本申请实施例中，无需 DNS 服务器进行域名解析切换，提供了用户端设备无感知切换服务。

图 11 示出了本申请又一实施例提供的服务单元切换方法的流程示意图。如图 11 所示，该方法，包括：

1301、确定目标服务单元。

1302、触发所述目标服务单元根据第一服务单元针对第一任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

上述 1301 中，在整个服务网络中，可能会存在多个服务单元，可从多个服务单元中除第一服务单元以外的其他服务单元中确定出目标服务单元。例如：选择负载小的服务单元作为目标服务单元。

上述 1302 中，第一任务可以为转发任务或为负载均衡任务。当第一任务为负载均衡任务时，第一服务单元针对第一任务的配置信息中可包括：所述第一服务单元针对所述第一任务配置的虚拟服务地址 VIP。第一任务指的是将针对该虚拟服务地址的业务流量

分发至该虚拟服务地址对应的多个后端服务器进行处理的任务。其中，流量分发服务指的即是针对该虚拟服务地址的业务流量分发至该虚拟服务地址对应的多个后端服务器上。该后端服务器指的是用来处理针对该虚拟服务地址的服务请求得到最终处理结果的真实服务器。

5 当第一任务为转发任务时，第一服务单元针对第一任务的配置信息中可包括：所述第一服务单元针对第一任务配置的第一 IP 地址（Internet Protocol Address）。第一任务指的是将访问第一 IP 地址的业务流量转发至 IP 地址为第一 IP 地址的处理服务器进行处理。

10 目标服务单元可根据配置信息中的第一 IP 地址发布第一路由信息，使得针对所述第一 IP 地址的服务请求寻址到所述目标服务单元，以由目标服务单元实现对该服务请求的转发至 IP 地址为第一 IP 地址的处理服务器进行处理。

当第一任务为负载均衡任务时，具体实现可参见上述各实施例中相应内容，在此不再赘述。

15 在本实施例中，将第一任务从第一服务单元转移至目标服务单元，并没有改变第一任务对应的目的 IP 地址。这样，就算用户端存在 DNS 缓存，也不影响业务流量的实时切换。

本实施例中针对第一任务的服务单元切换过程中未尽详述的内容可参照上述实施例中针对第一负载均衡任务的服务单元切换过程，在此不再赘述。

20 图 12 示出了本申请又一实施例提供的服务单元切换方法的流程示意图。如图 12 所示，该方法，包括：

1401、获取第一服务单元针对第一任务的配置信息。

1402、根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

25 上述 1401 中，可从第一服务单元处获取其针对第一任务的配置信息，例如：向第一服务单元发送获取请求；接收第一服务单元反馈的其针对第一任务的配置信息。

上述 1402 中，配置信息中可包括第一服务单元针对第一任务配置的第一 IP 地址。具体可针对第一 IP 地址发布第一路由信息，使得针对所述第一 IP 地址的服务请求寻址到本地。

30 本申请实施例提供的方法的执行主体可以为上述实施例中的目标服务单元。

上述 1401 和 1402 的具体实现过程可参见上述各实施例中相应内容，在此不再赘述。

在本实施例中，将第一任务从第一服务单元转移至目标服务单元，并没有改变第一任务对应的目的 IP 地址。这样，就算用户端存在 DNS 缓存，也不影响业务流量的实时切换。

5 本实施例中针对第一任务的服务单元切换过程中未尽详述的内容可参照上述实施例中针对第一负载均衡任务的服务单元切换过程，在此不再赘述。

图 7 示出了本申请一实施例提供的服务单元切换装置的结构框图。如图 7 所示，该装置包括：第一确定模块 801 和第一触发模块 802。其中，

10 第一确定模块 801，用于确定目标服务单元；

第一触发模块 802，用于触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

进一步的，所述配置信息中包括：所述第一服务单元针对所述第一负载均衡任务配置
15 的虚拟服务地址；

第一触发模块 802，具体用于：触发所述目标服务单元根据所述虚拟服务地址发布第一路由信息，使得针对所述虚拟服务地址的服务请求寻址到所述目标服务单元。

进一步的，第一触发模块 802，具体用于：获取所述第一服务单元针对所述虚拟服务地址已发布的第二路由信息；

20 将所述第二路由信息发送至所述目标服务单元，以触发所述目标服务单元根据所述虚拟服务地址及所述第二路由信息，发布优先级高于所述第二路由信息的第一路由信息。

进一步的，第一触发模块 802，还用于：

向所述第一服务单元发送会话发送指令，以触发所述第一服务单元将其记录的有关所述第一负载均衡任务的会话信息发送至所述目标服务单元，以供所述目标服务单元在

25 对有关已建立连接的服务请求进行分发时参考；

其中，所述会话信息中包括所述已建立连接的连接信息。

进一步的，所述目标服务单元为由多个第二服务节点组成的集群；

所述第一触发模块 802，具体用于：

30 向所述第一服务单元发送会话发送指令，以触发所述第一服务单元将其记录的有关所述第一负载均衡任务的会话信息发送至所述多个第二服务节点中的目标服务节点，并

由所述目标服务节点将所述会话信息同步至所述多个第二服务节点中的其他第二服务节点。

进一步的，第一触发模块 802，具体用于：获取第一服务单元针对第一负载均衡任务的配置信息；将所述配置信息发送至目标服务单元，以由所述目标服务单元根据所述配置信息完成配置，并根据所述配置信息发布所述第一路由信息。

进一步的，第一触发模块 802，具体用于：将所述配置信息发送至所述目标服务单元对应的目标管控单元，以由所述目标管控单元将所述配置信息转发至所述目标服务单元。

进一步的，上述装置，还包括：

10 第一确定模块，用于根据待迁移信息，确定出所述第一服务单元、所述第一负载均衡任务以及所述目标服务单元；

所述待迁移信息为根据检测到的多个服务单元的运行情况确定的；所述多个服务单元中包括所述第一服务单元和所述目标服务单元；

15 所述待迁移信息中包括所述第一服务单元针对所述第一负载均衡任务配置的虚拟服务地址以及所述目标服务单元的服务地址。

进一步的，上述装置，还包括：

第一接收模块，用于接收全局调度单元发送的其根据检测到的所述多个服务单元的运行情况确定的所述待迁移信息。

20 进一步的，所述运行情况中包括：当前所承担的总流量、总连接成功率和/或异常报警信息。

进一步的，上述装置，还包括：

第二接收模块，用于接收到有关所述第一路由信息的已发布通知后，触发所述第一服务单元撤销其针对所述第一负载均衡任务已发布的第二路由信息。

进一步的，上述装置，还包括：

25 第三接收模块，用于接收到有关所述第一路由信息的已发布通知信息后，触发所述第一服务单元删除有关所述第一负载均衡任务的配置。

本申请实施例提供的技术方案中，第一服务单元当前在执行第一负载均衡任务。当整个服务网络需要进行容灾切换及容量调度时，将第一服务单元针对第一负载均衡任务的配置信息发送给目标服务单元，目标服务单元根据配置信息进行相关配置，并发布有关第一负载均衡任务的第一路由，即可将与第一负载均衡任务相关的流量引流至目标服

务单元上，以由目标服务单元接替第一服务单元来提供有关第一负载均衡任务的流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够有效提高流量切换的实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

- 5 这里需要说明的是：上述实施例提供的服务单元切换装置可实现上述各方法实施例中描述的技术方案，上述各模块或单元具体实现的原理可参见上述各方法实施例中的相应内容，此处不再赘述。

进一步的，上述装置，还包括：

- 10 第四接收模块，用于接收所述目标服务单元针对所述第一负载均衡任务检测得到的检测信息；

第一判断模块，用于根据所述检测信息，判断所述目标服务单元是否满足接替条件；

第一执行模块，用于所述目标服务单元满足所述接替条件时，执行所述触发所述目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息的步骤。

- 15 进一步的，所述检测信息中包括：所述目标服务单元多个后端服务器进行检测得到的第一健康状态信息；其中，所述多个后端服务器用于处理与所述第一负载均衡任务相关的服务请求；以及

第一判断模块，具体用于：

根据所述第一健康状态信息，判断所述目标服务单元是否满足接替条件。

- 20 进一步的，所述目标服务单元为由多个第二服务节点组成的集群；

所述检测信息中还包括：所述多个第二服务节点之间有关所述第一负载均衡任务的会话信息的同步量；以及

第一判断模块，具体用于：

- 25 根据所述第一健康状态信息和所述同步量，判断所述目标服务单元是否满足所述接替条件；

所述会话信息是由所述第一服务单元在为执行所述第一负载均衡任务提供流量分发服务时记录的。

进一步的，第一判断模块，具体用于：

- 30 获取所述第一服务单元针对所述多个后端服务器进行检测得到的第二健康状态信息；

判断所述第一健康状态检测信息与所述第二健康状态检测信息之间的差异是否满足第一预设条件；

确定所述同步量在所述会话信息的总量中的占比；

所述差异满足第一预设条件且所述占比大于预设占比时，则判定所述目标服务单元
5 满足所述接替条件。

本申请实施例提供的技术方案中，第一服务单元当前在为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负
10 载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

这里需要说明的是：上述实施例提供的服务单元切换装置可实现上述各方法实施例中描述的技术方案，上述各模块或单元具体实现的原理可参见上述各方法实施例中的相
15 应内容，此处不再赘述。

图 8 示出了本申请又一实施例提供的服务单元切换装置的结构框图。如图 8 所示，该装置包括：第五获取模块 1001 和第一发布模块 1002。其中，

第五获取模块 1001，用于接收第一服务单元针对第一负载均衡任务的配置信息；
20 第一发布模块 1002，用于根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一负载均衡任务提供流量分发服务。

进一步的，第一发布模块 1002，具体用于：

获取所述第一服务单元针对所述第一负载均衡任务已发布的第二路由信息；

根据所述配置信息和所述第二路由信息，发布优先级高于所述第二路由信息的第一
25 路由信息。

进一步的，上述装置，还包括：

第六接收模块，用于接收所述第一服务单元发送来的有关所述第一负载均衡任务的会话信息；

第一同步模块，用于将所述会话信息同步至集群内部的其他第二服务节点；

30 所述集群中每一个第二服务节点均用于为所述第一负载均衡任务提供流量分发服

务。

进一步的，上述装置，还包括：

第三发送模块，用于向目标管控单元发送针对所述第一负载均衡任务检测得到的检测信息，以由所述目标管控单元根据所述检测信息确定是否满足接替条件；

- 5 第七接收模块，用于接收到所述目标管控单元在确定出满足接替条件时发送的接替指令后，执行所述根据所述配置信息发布第一路由信息的步骤。

本申请实施例提供的技术方案中，第一服务单元当前在为第一负载均衡任务提供流量分发服务。当整个服务网络需要进行容灾切换及容量调度时，目标服务单元根据第一服务单元针对第一负载均衡任务的配置信息发布第一路由信息，即可将与第一负载均衡
10 任务相关的流量引流至目标服务单元上，以由目标服务单元接替第一服务单元为第一负载均衡任务提供流量分发服务。可见，当整个服务网络需要进行容灾切换及容量调度时，采用本申请实施例提供的方法，能够确保与第一负载均衡任务相关的流量切换实时性，避免现有技术中因 DNS 缓存的影响导致的流量切换实时性较差的问题。

这里需要说明的是：上述实施例提供的服务单元切换装置可实现上述各方法实施例中
15 中描述的技术方案，上述各模块或单元具体实现的原理可参见上述各方法实施例中的相应内容，此处不再赘述。

图 7 示出了本申请一实施例提供的服务单元切换装置的结构框图。如图 7 所示，该装置包括：第一确定模块 801 和第一触发模块 802。其中，

- 20 第一确定模块 801，用于确定目标服务单元；

第一触发模块 802，用于触发所述目标服务单元根据第一服务单元针对第一任务的配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

在本实施例中，将第一任务从第一服务单元转移至目标服务单元，并没有改变第一任务对应的目的 IP 地址。这样，就算用户端存在 DNS 缓存，也不影响业务流量的实时
25 切换。

这里需要说明的是：上述实施例提供的服务单元切换装置可实现上述各方法实施例中描述的技术方案，上述各模块或单元具体实现的原理可参见上述各方法实施例中的相应内容，此处不再赘述。

- 30 图 8 示出了本申请又一实施例提供的服务单元切换装置的结构框图。如图 8 所示，

该装置包括：第五获取模块 1001 和第一发布模块 1002。其中，

第五获取模块 1001，用于接收第一服务单元针对第一任务的配置信息；

第一发布模块 1002，用于根据所述配置信息发布第一路由信息，以接替所述第一服务单元为所述第一任务提供服务。

5 在本实施例中，将第一任务从第一服务单元转移至目标服务单元，并没有改变第一任务对应的目的 IP 地址。这样，就算用户端存在 DNS 缓存，也不影响业务流量的实时切换。

 这里需要说明的是：上述实施例提供的服务单元切换装置可实现上述各方法实施例中描述的技术方案，上述各模块或单元具体实现的原理可参见上述各方法实施例中的相应内容，此处不再赘述。

 图 9 示出了本申请一实施例提供的电子设备的结构示意图。如图所示，所述电子设备包括存储器 1101 以及处理器 1102。存储器 1101 可被配置为存储其它各种数据以支持在电子设备上的操作。这些数据的示例包括用于在电子设备上操作的任何应用程序或方法 15 的指令。存储器 1101 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

 所述存储器 1101，用于存储程序；

20 所述处理器 1102，与所述存储器 1101 耦合，用于执行所述存储器 1101 中存储的所述程序，以实现上述各方法实施例提供的服务单元切换方法。

 进一步，如图 9 所示，电子设备还包括：通信组件 1103、显示器 1104、电源组件 1105、音频组件 1106 等其它组件。图 9 中仅示意性给出部分组件，并不意味着电子设备只包括图 9 所示组件。

25 相应地，本申请实施例还提供一种存储有计算机程序的计算机可读存储介质，所述计算机程序被计算机执行时能够实现上述各方法实施例提供的服务单元切换方法步骤或功能。

 以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，

即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下，即可以理解并实施。

5 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件。基于这样的理解，上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行各个实施例或者实施例的某些部分所述的方法。

10 最后应说明的是：以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。