

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 3 月 31 日 (31.03.2022)



(10) 国际公布号
WO 2022/062795 A1

(51) 国际专利分类号:
G06F 9/50 (2006.01) **H04L 29/08** (2006.01)

(21) 国际申请号: PCT/CN2021/114001

(22) 国际申请日: 2021 年 8 月 23 日 (23.08.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202011030599.5 2020 年 9 月 27 日 (27.09.2020) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 张诚 (ZHANG, Cheng); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房 (部位: 自编 01-03 和 08-12 单元) (仅限办公用途), Guangdong 510623 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,

(54) **Title:** SERVICE REQUEST ALLOCATION METHOD, APPARATUS AND COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 业务请求分配方法、装置、计算机设备和存储介质

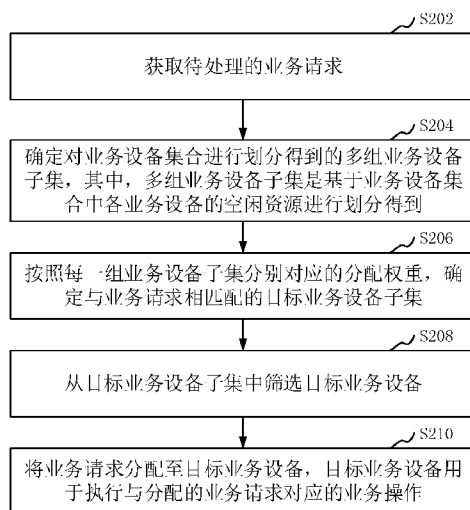


图 2

S202 Acquire a service request to be processed

S204 Determine multiple groups of service device subsets obtained by dividing a service device set, the multiple groups of service device subsets being obtained by means of dividing idle resources of each service device in the service device set

S206 According to the allocation weight corresponding to each group of service device subsets, determine a target service device subset that matches the service request

S208 Screen a target service device from a target service device subset

S210 Allocate the service request to the target service device, the target service device being used for executing a service operation corresponding to the allocated service request

(57) **Abstract:** A service request allocation method, that is executed by a computer device. The method comprises: acquiring a service request to be processed; determining multiple groups of service device subsets obtained by dividing a service device set, the multiple groups of service device subsets being obtained by means of dividing idle resources of each service device in the service device set; according to the allocation weight corresponding to each group of service device subsets, determining a target service device subset that matches the service request; screening a target service device from the target service device subset; and allocating the service request to the target service device, the target service device being used for executing a service operation corresponding to the allocated service request.

AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种业务请求分配方法，由计算机设备执行，所述方法包括：获取待处理的业务请求；确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；从目标业务设备子集中筛选目标业务设备；将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

发明名称：业务请求分配方法、装置、计算机设备和存储介质

本申请要求于 2020 年 09 月 27 日提交中国专利局，申请号为 202011030599.5、发明名称为“业务请求分配方法、装置、计算机设备和存储介质”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。

技术领域

本申请涉及计算机技术领域，特别是涉及一种业务请求分配方法、装置、计算机设备和存储介质。

背景技术

随着计算机技术的发展，提供服务的业务设备集合经常需要处理大量的业务请求。为了避免由于业务设备集合中的业务设备（如服务器）负载较大出现无法响应的情况，往往需要将业务请求合理分配至不同的业务设备中，以实现各业务设备之间能够基本负载均衡。

然而，现有技术中存在多种业务请求的分配方式，比如将业务请求轮询分配至不同的业务设备中，或者将业务请求分配至负载最小的业务设备中等。

但在实际使用过程中，现有的这些业务请求分配方式每执行一次业务请求的分配，就需要实时统计业务设备集合中各业务设备的负载信息，并进行一次的负载信息更新，以保障业务请求接入实现负载均衡。在这种情况下往往需要实时进行大量的计算，存在业务请求分配效率低的问题。

发明内容

本申请各实施例提供了一种业务请求分配方法、装置、计算机设备和存储介质。

一种业务请求分配方法，由计算机设备执行，所述方法包括：

获取待处理的业务请求；

确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；

按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；

从目标业务设备子集中筛选目标业务设备；及

将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

一种业务请求分配装置，装置包括：

获取模块，用于获取待处理的业务请求；

第一确定模块，用于确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；

第二确定模块，用于按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；

筛选模块，用于从目标业务设备子集中筛选目标业务设备；及

分配模块，用于将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理的业务请求；

确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；

按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；

从目标业务设备子集中筛选目标业务设备；及

将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理的业务请求；

确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；

按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；

从目标业务设备子集中筛选目标业务设备；及

将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中，计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上以下步骤：获取待处理的业务请求；确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到；按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集；从目标业务设备子集中筛选目标业务设备；及将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

附图说明

图 1 为一个实施例中业务请求分配方法的应用环境图；

图 2 为一个实施例中业务请求分配方法的流程示意图；

图 3 为一个实施例中确定对业务设备集合进行划分得到的多组业务设备子集步骤的流程示意图；

图 4 为另一个实施例中业务请求分配方法的流程示意图；

图 5 为一个实施例中提供的一种按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集步骤的流程示意图；

图 6 为一个实施例中提供的一种确定目标分配码所属的目标分配码子集步骤的流程示意图；

图 7 为一个实施例中提供的一种将业务请求分配至目标业务设备步骤的流程示意图；

图 8 为一个实施例中提供的一种根据 IP 地址分配业务请求的示意图；

图 9 为一个实施例中根据业务请求分配方法执行业务请求分配时各业务设备对应的资源使用率示意图；

图 10 为一个实施例中业务请求分配装置的结构框图；

图 11 为一个实施例中计算机设备的内部结构图。

具体实施方式

本申请提供的业务请求分配方法，可以应用于如图 1 所示的应用环境中。其中，计算机设备 110

通过网络分别与终端 130 以及业务设备集合 120 进行通信。其中，业务设备集合 120 中包括多于一组的业务设备子集，并且每一个业务设备子集中包括一个或者多个业务设备。具体地，计算机设备 110 获取终端 130 发送的业务请求，并在业务设备集合 120 中确定与业务请求匹配的目标业务设备，然后通过网络将业务请求分配至对应的目标业务设备中，以使得目标业务设备对分配到的业务请求执行业务处理。

需要说明的是，上述计算机设备具体可以是实现路由功能的终端或者服务器，比如路由设备、分配设备、或中心服务器等。业务设备具体可以用独立的服务器或者是多个服务器组成的服务器集群来实现。需要说明的是，上述所提及的服务器具体可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。终端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等，但并不局限于此。终端 130 和计算机设备 110 可以通过有线或无线通信方式进行直接或间接地连接、以及计算机设备 110 和业务设备集合可以通过有线或无线通信方式进行直接或间接地连接，本申请在此不做限制。

在其中一个实施例中，如图 2 所示，提供了一种业务请求分配方法的流程示意图，以该方法应用于图 1 中的计算机设备 110 为例进行说明，该业务请求分配方法包括以下步骤：

步骤 S202，获取待处理的业务请求。

其中，业务请求可以是通信连接请求，也可以业务处理请求等，在不同的业务场景中业务请求的类型可以不同或者相同，在本申请中不做限制。业务请求的发送方可以是用户端也可以是计算机设备端，即业务请求可以是用户端向计算机设备发送的请求，也可以是在计算机设备中自动产生的请求。需要说明的是，计算机设备可以是实体服务器也可以是虚拟服务器，具体可以是网关路由服务器，在本申请中不做限制。

具体地，当业务请求是用户端发送的请求时，具有业务处理请求的用户端可以实时向计算机设备发送待处理的业务请求，或者也可以预先在用户端设置定时任务，当达到预先设定的时间时，通过用户端定时向计算机设备发送业务请求。当业务请求是在计算机设备中产生的请求时，可以预先在计算机设备中设定定时任务，当到达定时任务对应的时间时，在计算机设备中自动产生待处理的业务请求。当计算机设备接收到待处理的业务请求时，根据业务请求对应的预设处理逻辑对业务请求进行处理。

在其中一个实施例中，以用户端向计算机设备发送业务请求，并且业务请求对应为通信连接请求为例进行说明。用户端发出通信连接请求，当计算机设备接收到用户端发送的通信连接请求时，计算机设备在由多个业务设备组成的业务设备集群中为该通信连接请求匹配目标业务设备，并将该通信连接请求分配至对应的目标业务设备中，以指示目标业务设备对分配到的通信连接请求执行业务处理。

步骤 S204，确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到。

其中，业务设备集合中包括多于一个的业务设备，并且在具体的业务场景中多个业务设备协同完成业务请求对应的工作任务。业务设备可以为不同类型的设备，如可以为 FTP 服务器、Web 服务器、企业核心应用服务器和其它主要任务服务器等，在此不做限定。并且，在具体应用中，本申请所提供的业务请求分配方法可以支撑调度不同类型和处理能力的业务设备同时服务，并且能将各业务设备进行分布式部署，根据具体的业务需求增加或者删减业务设备，可扩展性强，提高了对业务场景的适用性，以及对业务处理的效率。

业务设备子集是对业务设备集合中的多个业务设备按照一定的划分规则进行组别划分得到的。业务设备子集的数量为多组，多组可以理解为多于一组，以及每个业务设备子集中包括一个或者多个的业务设备，需要说明的是，本申请各实施例所提及的“多组”、“多个”，均表示“多于一组”、“多于一个”。空闲资源是用于衡量业务设备还可以负载业务的性能指标，具体的，当业务设备对应的空闲资源越多时，可以认为该业务设备还可以负载更多的业务。其中，用于衡量业务设备的空闲资源的性能指标可以包括网络带宽、虚拟房间数量以及可接入用户数量等。

可以理解，在本申请中，可以通过一个性能指标单独来衡量业务设备的空闲资源，也可以利用多个性能指标综合来衡量业务设备的空闲资源。在其中一个实施例中，可以单独利用网络带宽、单独利用虚拟房间数量或者单独利用可接入用户数量，来衡量业务设备的空闲资源。在其中一个实施例中，也可以利用网络带宽、虚拟房间数量以及可接入用户数量的综合值，来衡量业务设备的空闲资源等，在此不作限制。可以理解，业务设备可用的（未被占用的）网络带宽越多，或者可连接的虚拟房间数量越多，或者可接入用户数量越多说明业务设备的空闲资源越多。

在其中一个实施例中，对业务设备集合中的业务设备进行组别划分的划分规则包括根据各业务设备的空闲资源进行组别划分。如可以将空闲资源相近的业务设备划分至同一个组中，使得从属于同一个组别的业务设备形成一个业务设备子集，实现了同一个业务设备子集中的业务设备的空闲资源大致相同或者相差在一定的范围内，即从属于同一个业务设备子集中的业务设备具有相近（几乎同等水平）的业务处理能力。需要说明的是，在对业务设备进行组别划分之前，还可以挑选出空闲资源小于预设值的业务设备，并将这些业务设备作为当前负载压力较大的设备，不参与组别的划分以及后续接收业务请求的步骤，防止当前负载压力较大的业务设备由于负载过多而发生宕机的情况发生。在其中一个实施例中，还可以将这些当前负载压力较大的业务设备划分至同一个组别中，并且该组别中的业务设备不接收新的业务请求。

具体的，对业务设备进行组别划分得到多组业务设备子集的步骤可以在步骤 S202 计算机设备获取到业务请求之后确定的，也可以是在步骤 S202 计算机设备获取到业务请求之前确定的，在本申请中不做限定。

在其中一个实施例中，在步骤 S202 计算机设备获取到业务请求之后的具体步骤还包括：对业务设备进行组别划分得到多组业务设备子集。具体的，当计算机设备获取到待处理的业务请求时，计算机设备实时统计业务设备集合中每一个业务设备对应的空闲资源，根据空闲资源的大小对业务设备进行组别划分，得到多组的业务设备子集。

在其中一个实施例中，在步骤 S202 计算机设备获取到业务请求之前的具体步骤还包括：对业务设备进行组别划分得到多组业务设备子集。具体的，计算机设备统计业务设备集合中每一个业务设备对应的空闲资源，根据空闲资源的大小对业务设备进行组别划分，得到多组的业务设备子集，并进行存储。当计算机设备获取到待处理的业务请求时，计算机设备直接获取预先存储的多组的业务设备子集。可以理解，计算机设备可以在固定的时间点执行对业务设备的子集划分，如固定在每天凌晨 0 点以及凌晨 3 点分别执行一次业务设备子集的划分步骤。计算机设备还可以预先确定一个时间周期，然后在每个时间周期内都执行一次对业务设备的组别划分步骤，如计算机设备可以是每间隔 5 秒钟就执行一次业务设备的组别划分步骤。在本申请中对此不做限制。

步骤 S206，按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集。

其中，分配权重是一个相对值，用于衡量业务设备集合中每一个业务设备子集还可以负载任务的

能力。具体的，分配权重可以对应为一个具体的数值，一般来说，分配权重对应的数值越大，代表该业务设备子集对应的空闲资源越多，即该业务设备子集当前的负载率较小，业务设备在当前时间的可执行能力越强。

具体的，计算机设备根据业务设备集合中每一个业务设备对应的空闲资源，计算总空闲资源，然后计算每一组业务设备子集中的业务设备对应的子集空闲资源，并根据各子集空闲资源与总空闲资源的相对值，确定每一组的业务设备子集对应的分配权重。其中相对值可以是各子集空闲资源与总空闲资源的比值，相对值还可以是各子集空闲资源与总空闲资源的差值等，在此不做限制。可以理解，相对值越大，说明业务设备子集在业务设备集合中所占的比重越大，该业务设备子集的可负载能力越强。

具体的，计算机设备根据各业务设备子集对应的分配权重，从各业务设备子集中提取一个子集，并将该子集确定为目标业务设备子集，并把当前获取到的待处理的业务请求分配至目标业务设备子集中。需要说明的是，在一个时间周期内（也就是各业务设备子集的分配权重固定不变的周期内），分配至各业务设备子集中的业务请求的数量比值与各业务设备对应的分配权重的比值一致。实现了在一个时间周期内均衡地向各业务设备子集分配业务请求目的，实现了业务设备子集的负载均衡。

步骤 S208，从目标业务设备子集中筛选目标业务设备。

由于目标业务设备子集中包括多于一个的业务设备，故而当计算机设备为业务请求匹配到对应的目标业务设备子集后，还包括从目标业务设备子集中再次筛选一个业务设备，并将筛选得到的业务设备作为目标业务设备，用于处理本次获取到的待处理的业务请求。

其中，计算机设备从目标业务设备子集中筛选目标业务设备的方式可包括多种，如可以是轮询筛选算法、空闲资源原则筛选算法、随机筛选算法等，在本申请中不做具体限制。

在其中一个实施例中，计算机设备根据轮询筛选算法从目标业务设备子集中筛选目标业务设备的步骤包括：计算机设备首先根据目标业务设备子集中各业务设备对应的空闲资源对各业务设备进行排队，并确定每一个业务设备的排队顺序，然后以排队顺序作为轮询顺序，将业务请求在本次轮询到的业务设备提取为目标业务设备，并利用目标业务设备执行对业务请求的业务处理。

在其中一个实施例中，计算机设备根据空闲资源原则筛选算法从目标业务设备子集中筛选目标业务设备的步骤包括：计算机设备确定目标业务设备子集中各业务设备对应的空闲资源，然后将空闲资源最大的业务设备筛选为目标业务设备。

在其中一个实施例中，计算机设备根据随机筛选算法从目标业务设备子集中筛选目标业务设备的步骤包括：计算机设备随机从目标业务设备子集中选取一个业务设备，并将随机选取的业务设备确定为目标业务设备。需要说明的是，在本申请中的其他实施例中，还可以根据其他的算法从目标业务设备中筛选目标业务设备，在此不作限制。

步骤 S210，将业务请求分配至目标业务设备，目标业务设备用于执行与分配的业务请求对应的业务操作。

具体的，计算机设备匹配到与待处理的业务请求对应的目标业务设备后，将待处理的业务请求分配至对应的目标业务设备，以指示目标业务设备对分配到的业务请求执行对应的业务操作。其中，业务操作与业务请求的类型相对应，当业务请求的类型为通信连接请求时，那么对应的业务操作为通信连接操作，当业务请求的类型为业务处理请求时，那么对应的业务操作为业务处理操作。

在一个具体的实施例中，计算机设备在对应的时间周期内获取业务设备集合中的所有业务设备的空闲资源，然后按照空闲资源将各业务设备划分组别，得到多个业务设备子集，并根据各业务设备子集中包括的业务设备对应的空闲资源确定各业务设备子集的在当前的时间周期内对应的分配权重。计

计算机设备在当前的时间周期内（当前 5 秒内）根据各业务设备子集对应的分配权重向对应的业务设备子集中分配对应数量的业务请求，以保证在该时间周期内各业务设备子集的负载均衡。然后再根据预设的规则从目标业务子集中筛选一个目标业务设备，再次将业务请求分配至目标业务设备中，并利用目标业务设备对业务请求进行处理。在本实施例中，由于各业务设备子集的分配权重表征了该业务设备子集的负载能力，故而根据各业务设备的分配权重将对应数量的业务请求分配至对应的业务设备子集中，实现了业务设备集合中各业务设备子集之间资源分配的均衡性，避免了由于业务请求分配不均造成业务设备子集的超负荷，进而造成业务设备子集中的业务设备的宕机等弊端。可以理解，在下一个的时间周期内，计算机设备需要再次重新统计业务设备集合中各业务设备的最新的空闲资源，并根据最新的空闲资源重新进行业务设备子集的划分，以及各业务设备子集的分配权重的计算，然后当业务设备接收到业务请求时，执行根据最新的分配权重再次重新确定目标业务设备子集以及目标业务设备的步骤。

这样，将空闲资源相近的业务设备放入相同的业务设备子集，并确定各业务设备子集对应的分配权重，根据分配权重将业务请求分配至目标业务设备子集中，然后再从目标业务设备子集中筛选出目标业务设备，将业务请求分配至目标业务设备中。通过将空闲资源相近的业务设备放入相同的业务设备子集，在不同的业务设备子集之间按分配权重进行分配，在同一个业务设备子集内部进行轮询分配，降低了算法的计算量，同时兼顾了不同分配权重的业务设备子集的分配概率不同，做到了业务请求的合理分配以及业务设备的负载均衡。

上述业务请求分配方法，确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集基于业务设备集合中的多于一个的业务设备的空闲资源进行划分得到，使得每一组的业务设备子集中包括一个或者多个的负载情况相近的业务设备。当获取到待处理的业务请求时，就可按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集。再将业务请求分配至目标业务设备子集中的目标业务设备，以进行相应的业务操作。由于目标业务子集中包括多个的业务设备，故而业务设备子集的数量大大小于业务设备的数量，通过先将业务请求分配至目标业务设备子集中而不是直接分配至某一个具体的业务设备中，减少了分配过程中的计算量，提高了业务请求的分配效率。并且根据各业务设备子集对应的分配权重将业务请求分配至对应的目标业务设备子集中，实现了对业务请求的合理分配，使得各个业务设备子集间满足负载均衡。这样，先按照负载均衡的方向，从粗粒度上将业务请求分配至目标业务设备子集中，然后再将业务请求分配至目标业务设备子集中的某个目标业务设备，既能满足对业务请求进行分配时的负载均衡，又无需实时进行大量的负载更新计算，极大地减少了计算量，大大提高了业务请求的分配效率。

参考图 3，在其中一个实施例中，步骤 S204，也就是确定对业务设备集合进行划分得到的多组业务设备子集的步骤具体包括：

步骤 S302，在当前时间周期内，确定业务设备集合中的多于一个的业务设备分别对应的可用资源信息。

其中，当前时间周期是根据预先定义的时间区间确定的周期，如可以预先定义时间区间是 5 秒，那么当前时间周期就对应为当前 5 秒的时间区间。也就是说将每 5 秒的时间范围作为一个时间周期，并且在一个时间周期内对业务设备集合中的各业务设备的可用资源信息统计并进行更新一次，并根据最新更新的可用资源信息对各业务设备进行排序。可用资源信息是衡量当前负载程度或当前资源空闲程度的信息，具体可以是负载率、资源空闲率或空闲资源等。

在其中一个实施例中，计算机设备可根据当前空闲资源信息与有效资源信息的对比值，来计算各业务设备所对应的可用资源信息。其中，当前空闲资源信息是对业务设备当前的空闲资源进行量化得到的量化信息，具体可以是一个量化数值。有效资源信息是对有效资源进行量化得到的量化信息，具体可以是一个量化数值。计算机设备可将当前空闲资源信息与有效资源信息的对比值，作为业务设备所对应的可用资源信息。对比值是两个数值的差异值，具体可以是当前空闲资源信息与有效资源信息的商。

在另一个实施例中，计算机设备可直接将业务设备的当前空闲资源信息作为可用资源信息。

步骤 S304，根据可用资源信息对每一个业务设备进行排序。

具体的，计算机设备根据业务设备集合中各业务设备对应的可用资源信息的大小，对业务设备进行从大至小，或者从小至大的排序。

步骤 S306，按照预先确定的组别划分比例，将排序后的每一个业务设备划分至对应的组别。其中，从属于同一个组别的业务设备共同构成一组业务设备子集。

其中，组别划分比例是预先确定的、以将业务设备集合划分为不同组别的比例值。根据每个业务设备对应的可用资源信息对业务设备进行排序，并根据各业务设备对应的排序值确定组别划分比例。需要说明的是，在不同的业务场景中，组别划分比例并不是一个固定的数值，可以根据具体的业务需求对组别划分比例做适应性调整。

具体的，计算机设备预先确定各业务设备子集中包括的业务设备的数量，根据各业务设备子集对应的数量确定组别划分比例。在其中一个实施例中，业务设备集合中的业务设备的总数为 24 个，预先确定业务设备子集的数量为 5 组，并且预先确定各业务设备子集中的业务设备的数量分别为 2、4、6、4、8，那么对应的组别划分比例为 1: 2: 3: 2: 4。在具体实施中，计算机设备根据每个业务设备对应的可用资源信息对业务设备进行排序，并从排序后的业务设备中根据 1: 2: 3: 2: 4 的比例划分不同的业务设备子集，得到多组的业务设备子集。

具体的，组别划分比例还可以是根据分档级别比例确定的。在其中一个实施例中，计算机设备根据各业务设备对应的可用资源信息对业务设备进行排序，然后根据预先确定分档级别比例进行组别的划分。如将排序在 0-10% 的业务设备划分至第一档 G1 中生成第一档的业务设备子集，将排序在 10%-25% 的业务设备划分至第二档 G2 中生成第二档的业务设备子集，将排序在 25%-50% 的业务设备划分至第三档 G3 中生成第三档的业务设备子集，将排序在 50%-85% 的业务设备划分至第四档 G4 中生成第四档的业务设备子集。进一步地还可以将 85%-100% 的业务设备划分至第五档 G5 中生成第五档的业务设备子集。需要说明的是，在其他实施例中，由于排序在 85%-100% 之间的业务设备的可用资源信息比较小，即这些业务设备当前的负载率很高，还能负载的业务能力较弱，故而在执行组别划分的过程中可以舍弃这些业务设备，或者将这些业务设备划分之不参与接收业务请求的组别中，以保证负载较高的业务设备中不再接收新的业务请求，保证负载较高的业务设备不会因为接收到过多的业务请求而导致负载过大而宕机的情况发生。

具体的，在当前时间周期内，业务设备根据各业务设备的可用资源信息对业务设备进行组别的划分，得到多个业务设备子集，并同时确定各业务设备子集对应的分配权重。然后在该时间周期内接收到的所有的业务请求都根据在当前时间周期内确定的分配权重进行业务请求的分配。当完成在当前时间周期内的业务请求的分配后，计算机设备开启在下一个时间周期内重新统计各业务设备的可用资源信息，并重新根据重新统计到的可用资源信息执行新一轮周期内的业务设备子集的划分，并重新确定新一轮业务设备子集对应的分配权重，以及根据重新确定的分配权重执行对新一轮时间周期内接收到

的业务请求进行分配。如此往复循环，每到一个时间周期就重新执行一次的业务设备子集的更新，保证了业务设备子集划分的准确性以及实时性，进而实现了将业务请求合理地分配至业务设备子集中的目的。

在其中一个实施例中，以一个时间周期为 5 秒为例进行说明。计算机设备预先统计各业务设备实时对应的可用资源信息，并根据可用资源信息的大小将各业务设备进行组别划分，得到多个级别的业务设备子集，以在该 5 秒时间段不重新执行业务设备子集的划分，而是都以该业务设备子集为基准进行业务请求的分配。只有当时间超出当前 5 秒时，业务设备才会重新统计业务设备集合中各业务设备的最新可用资源信息，并重新进行业务设备子集的划分，以使得在下一个 5 秒内接收到的所有业务请求都根据最新统计到的业务设备子集进行分配。

上述实施例中，对业务设备集合中的所有的业务设备的负载进行量化得到可用资源信息，并且每 N 秒收集一次业务设备的可用资源信息，并根据收集到的可用资源信息对业务设备进行排序，并按照排序对应的顺序对业务设备进行分档，得到多组的业务设备子集。通过每 N 秒执行一次组别划分，对业务设备集合中的可用资源信息的收集时延要求相对较低，同时减少了计算量。以及计算每一个业务设备子集的分配权重，进而可以根据分配权重将业务请求分配到目标业务设备子集中。进一步地，还可以在目标业务设备子集中进行轮询分配，实现对业务请求的合理分配，实现了各业务设备的负载均衡。

如图 4 所示，图 4 为另一个实施例中业务请求分配方法的流程示意图，在其中一个实施例中，业务请求分配方法还包括：

步骤 S402，获取当前时间周期内多于一个的业务设备分别对应的可用资源信息。

步骤 S404，根据可用资源信息对业务设备进行组别划分，并确定各业务设备分别所属的业务设备子集。

步骤 S406，当业务设备在当前时间周期内接收到业务请求时，按照各组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集。

步骤 S408，将业务请求分配至目标业务设备子集中的目标业务设备中。

对以上步骤 S402、步骤 S404、步骤 S406 以及步骤 S408 的限定参照上文中对步骤 S302、步骤 S204、步骤 S206、步骤 S210 以及步骤 S210 的相关描述，本申请实施例在此不作赘述。

上述实施例中，通过在当前时间周期内统计一次各业务设备的可用资源信息，并根据统计到的可用资源信息执行对业务设备的分组步骤，得到当前时间周期对应的多个业务设备子集。进而在当前时间周期内获取到的所有的业务请求都能根据当前时间周期确定的业务设备子集执行对业务请求的分配。当到达下一个时间周期时，在重新统计各业务设备的可用资源信息，并重新进行业务设备子集的划分。在整个过程中，以一个时间周期为单位执行一次负载计算，并执行一次业务设备子集的更新，而不需要实时地对业务设备子集进行更新，不仅不会存在负载信息统计延时的问题，并且也减少了计算量，提高了业务请求分配的效率。

在其中一个实施例中，在当前时间周期内，确定业务设备集合中的多于一个的业务设备分别对应的可用资源信息，包括：确定业务设备集合中的各业务设备分别对应的有效资源信息。在当前时间周期内，收集业务设备集合中的各业务设备分别对应的当前空闲资源信息，根据当前空闲资源信息与有效资源信息的对比值，确定每一个业务设备分别对应的可用资源信息。

其中，有效资源信息是对有效资源进行量化得到的量化信息，具体可用一个具体数值表示。并且有效资源信息对应的数值越大说明该业务设备的有效资源越多，也可以理解为有效资源信息为该业务

设备可负载的最大资源数值。有效资源信息是每种类型的业务设备在上线前压测时就获取到的数据，并且有效资源可通过各业务设备的cpu、内存、流量、磁盘等指标进行衡量。其中，流量可以为内网流量，cpu是一个动态的指标，波动较大。在其中一个实施例中，可以使用可接入用户数量和虚拟房间数量进行有效资源的衡量，如可以通过业务设备可支撑的最大可接入用户数量(max_user_cnt)、最大虚拟房间数量(max_room_cnt)、最大上下行带宽(max_in_flow,max_out_flow)来确定有效资源信息。需要说明的是，在其他实施例中，考虑到待处理的业务请求对磁盘的影响很小，可以忽略不计，故而可以不用磁盘占用的资源对有效资源信息进行衡量。

空闲资源信息是对空闲资源进行量化得到的量化信息，当前空闲资源信息是业务设备在当前时间周期内对应的空闲资源信息。空闲资源是业务设备当前未被占用的、且可供使用的设备资源。

具体地，在当前时间周期内基于心跳算法，收集业务设备集合中所有的业务设备分别对应的有效资源信息以及当前空闲资源信息，根据当前空闲资源信息与有效资源信息的比值，确定每一个业务设备分别对应的可用资源信息。比如，计算机设备可根据与当前业务设备相对应的有效资源信息、以及与当前业务设备相对应的当前空闲资源信息，确定与当前业务设备对应的可用资源信息。

在其中一个实施例中，计算机设备在当前时间周期内，确定各业务设备在当前时间周期对应的有效资源信息。需要说明的是，由于有效资源信息是业务设备固有的属性信息，故而计算机设备可以直接获取到各业务设备的有效资源信息，并且在不同的时间周期内，各业务设备的有效资源信息可以是固定不变的。其中，有效资源信息可包括可接入用户最大数量(max_user_cnt)、可申请虚拟房间最大数量(max_room_cnt)、入流量最大值(max_in_flow)和出流量最大值(max_out_flow)。其中，入流量最大值和出流量最大值还可称作有效上下行带宽(max_in_flow,max_out_flow)。计算机设备统计各业务设备的当前空闲资源信息，如统计当前服务的当前可接入用户数量(curr_user_cnt)、当前虚拟房间数量(curr_room_cnt)和当前上下行带宽(curr_in_flow,curr_out_flow)。然后计算机设备根据统计到的当前空闲资源信息与有效资源信息的对比值，确定每一个业务设备分别对应的可用资源信息。其中，对比值是两个数值的差异值，具体可以通过数学计算来确定两个数值的差异值。数学计算，比如将两个数直接相除、取对数后相除、相减，或进行其他运算后再相除等等。

在其中一个实施例中，对比值可以为求比值计算得到的比值，此时各比值分别为：可接入用户数量比例= $\text{curr_user_cnt}/\text{max_user_cnt}$ 、可用虚拟房间数量比例= $\text{curr_room_cnt}/\text{max_room_cnt}$ 、入流量空闲比例= $\text{curr_in_flow}/\text{max_in_flow}$ 、出流量空闲比例= $\text{curr_out_flow}/\text{max_out_flow}$ 。在另外的实施例中，对比值还可以是求差计算得到的差值等，在此不作限制，只要得到的对比值可以衡量一个数值相对于另一个数值的差异状态即能够表达当前空闲资源信息与有效资源信息的相对大小关系都在本申请的保护范围之内。在其中一个实施例中，计算机设备可综合可接入用户数量比例、可用虚拟房间数量比例、入流量空闲比例以及出流量空闲比例，得到当前空闲资源信息与有效资源信息的对比值。

在其中一个实施例中，当前空闲资源信息包括当前可接入用户数量、当前虚拟房间数量、当前入流量空闲值和当前出流量空闲值。其中，当前可接入用户数量具体可以是，可接入用户最大数量与当前已接入用户数量的差值；当前虚拟房间数量具体可以是，可申请虚拟房间最大数量与当前已申请虚拟房间数量的差值；当前入流量空闲值具体可以是，入流量最大值与当前入流量占用值的差值；当前出流量空闲值具体可以是，出流量最大值与当前出流量占用值的差值。

需要说明的是，若当前已接入用户数量大于可接入用户最大数量时，将当前已接入用户数量赋值为可接入用户最大数量，防止当前可接入用户数量在计算时为负值。若当前已申请虚拟房间数量大于可申请虚拟房间最大数量时，将当前已申请虚拟房间数量赋值为可申请虚拟房间最大数量，防止当前

虚拟房间数量在计算时为负值。若当前入流量占用值大于入流量最大值时，将当前入流量占用值赋值为入流量最大值，防止当前入流量空闲值在计算时为负值。若当前出流量占用值大于出流量最大值时，将当前出流量占用值赋值为出流量最大值，防止当前出流量空闲值在计算时为负值。

上述实施例中，通过对业务设备的有效资源以及空闲资源进行量化得到有效资源信息以及空闲资源信息。然后在当前时间周期内，通过计算各业务设备当前空闲资源信息与有效资源信息的对比值，确定每一个业务设备分别对应的可用资源信息，并根据可用资源信息对业务设备进行组别划分，得到多个的业务设备子集。实现了在不同时间周期内实时对业务设备子集进行更新，提高了对业务请求分配的准确性。

在其中一个实施例中，有效资源信息包括可接入用户最大数量、可申请虚拟房间最大数量、入流量最大值、以及出流量最大值。当前空闲资源信息包括当前可接入用户数量、当前可接入虚拟房间数量、当前入流量空闲值、以及当前出流量空闲值。根据当前空闲资源信息与有效资源信息的对比值，确定每一个业务设备分别对应的可用资源信息，包括：根据各业务设备分别对应的当前可接入用户数量和可接入用户最大数量，确定各业务设备各自对应的第一对比值；根据各业务设备分别对应的当前可申请虚拟房间数量和可申请虚拟房间最大数量，确定各业务设备各自对应的第二对比值；根据各业务设备分别对应的当前入流量空闲值和入流量最大值，确定各业务设备各自对应的第三对比值；根据各业务设备分别对应的当前出流量空闲值和出流量最大值，确定各业务设备各自对应的第四对比值；及对于业务设备集中的每个业务设备，均将相应业务设备对应的第一对比值、第二对比值、第三对比值和第四对比值中的最小值，作为与相应业务设备对应的可用资源信息。

其中，用于评估业务设备有效资源的性能指标可以包括但不限于可接入用户数量、虚拟房间数量、流量数值。具体的，可接入用户最大数量是指业务设备可以接入的用户的最大数量值，可申请虚拟房间最大数量是指业务设备提供的可申请的虚拟房间的最大数量值，入流量最大值以及出流量最大值是指业务设备可提供的进、出流量的最大数值。需要说明的是，有效资源信息是业务设备本身固有的属性信息，即对于一个具体的业务设备而言，其能负担的有效资源信息是一个确定的数值，当达到该有效资源信息的最大值时，在该业务设备的可执行能力范围内，该业务设备是不能再接收更多的负载任务，否则该业务设备会出现负载过多，出现宕机等故障。

当前空闲资源信息是指业务设备在当前时间还具有的空闲资源，并且可以利用该空闲资源执行具体的业务请求任务。相对于有效资源对应得指标性能，该空闲资源可以包括当前可接入用户数量、当前可接入虚拟房间数量、当前入流量空闲值、以及当前出流量空闲值。可以理解，业务设备的当前空闲资源与业务设备当前占用资源之和确定业务设备的有效资源，一般来说，对于同一个业务设备而言，其有效资源信息是一个确定的数值，当其当前占用资源越少时，其对应的当前空闲资源信息越大，即接收业务请求的概率越大。

具体的，计算机设备可以分别根据各性能指标对应的有效资源信息以及空闲资源信息计算每一个性能指标对应的对比值。在其中一个实施例中，计算机设备根据各业务设备分别对应的当前可接入用户数量和可接入用户最大数量，确定相应的第一对比值，也即确定每个业务设备各自对应的第一对比值。根据各业务设备分别对应的当前可申请虚拟房间数量和可申请虚拟房间最大数量，确定相应的第二对比值，也即确定每个业务设备各自对应的第二对比值。根据各业务设备分别对应的当前入流量空闲值和入流量最大值，确定相应的第三对比值，也即确定每个业务设备各自对应的第三对比值。根据各业务设备分别对应的当前出流量空闲值和出流量最大值，确定相应的第四对比值，也即确定每个业务设备各自对应的第四对比值。

在其中一个实施例中，对于每个业务设备，计算机设备从相对应的第一对比值、第二对比值、第三对比值和第四对比值中提取最小值，并将提取到的最小值作为相应业务设备所对应的可用资源信息。如可用资源信息（量化值）= Min （第一对比值、第二对比值、第三对比值、第四对比值）。并且，在实际场景中，还可以将得到的可用资源信息扩大一定的倍数，如可以扩大 100000 倍，以将可用资源信息进行归一化处理，防止在计算过程中处理的数据都是小数，导致计算机设备的处理精度丢失。在其中一个实施例中，对于业务设备集中的每个业务设备，计算机设备均可根据与当前业务设备相对应第一对比值、第二对比值、第三对比值和第四对比值计算综合值，根据综合值与当前业务设备对应的可用资源信息。

在其中一个实施例中，对于多个业务设备中的每个业务设备，计算机设备均将与当前业务设备相对应的当前可接入用户数量与可接入用户最大数量之间的商值，作为与当前业务设备相对应的第一对比值；将与当前接入设备相对应的当前可申请虚拟房间数量与可申请虚拟房间最大数量之间的商值，作为与当前业务设备相对应的第二对比值；将与当前接入设备相对应的当前入流量空闲值和入流量最大值之间的商值，作为与当前接入设备相对应的第三对比值；将与当前接入设备相对应的当前出流量空闲值和出流量最大值之间的商值，作为与当前接入设备相对应的第四对比值。

上述实施例中，通过计算业务设备的多个性能指标分别对应的对比值，并将多个对比值的最小值确定为业务设备的可用资源信息。利用最短边效应（木桶效应）确定业务设备的可用资源信息，保证了业务设备在可用资源信息内接收业务请求时不会出现由于负载过大而出现宕机的情况，使得可用资源信息的确定更加合理。以及，将业务设备集合中不同型号的业务设备都统一抽象化为一个可用资源信息，以利用可用资源信息来表征不同类型的业务设备的最大服务能力，进而可以实现对不同类型的业务设备的统一分配处理。

在其中一个实施例中，该业务请求分配方法还包括：分别根据每组业务设备子集中各业务设备的空闲资源，确定每组业务设备子集各自对应的子集权重。根据每组业务设备子集各自对应的子集权重，确定与业务设备集合对应的集合总权重。分别根据每组业务设备子集各自对应的子集权重与集合总权重的对比值，确定每组业务设备子集分别对应的分配权重。

其中，子集权重是指业务设备子集在业务设备集合中所占得比重。子集权重越大对应为业务设备子集在业务设备中所占的比重越大，即该业务设备子集当前的空闲资源越多，还可以接收更多的业务负载。集合总权重是由业务设备集合中所有的业务设备对应的空闲资源确定的，如集合总权重可以为业务设备集合中所有业务设备对应的空闲资源的和值。

具体的，每一组业务设备子集分别对应的分配权重的确定步骤具体包括：对每一组业务设备子集中的每一个业务设备对应的空闲资源信息求和计算，确定每一组业务设备子集对应的子集权重。对各业务设备子集分别对应的子集权重求和计算确定集合总权重。根据每一个子集权重与集合总权重的比值确定每一组业务设备子集分别对应的分配权重。

在其中一个实施例中，计算机设备统计业务设备集合中所有业务设备的空闲资源信息，并根据空闲资源信息对应的数值大小对各业务设备进行排序。根据预先确定的组别划分比例（分档比例），将业务设备分成 N 档，例如当分档比例 $N=5$ 时，将排名在 0-10%范围内的业务设备划分至 G1 档，将排名在 10-25%范围内的业务设备划分至 G2 档，将排名在 25-50%范围内的业务设备划分至 G3 档，将排名在 50%-85%范围内的业务设备划分至 G4 档，以及将排名在 85-95%范围内的业务设备划分至 G5 档。然后，业务设备根据每档的业务设备子集中包括的业务设备对应的空闲资源信息，求得每组的业务设备子集对应的子集权重分别为： w_1 、 w_2 、 w_3 、 w_4 以及 w_5 ，计算所有业务设备对应的集合总权重为：

$W_{all} = w_1 + w_2 + w_3 + w_4 + w_5$ ，计算每一组业务设备子集对应的分配权重分别为： $W_1 = w_1 / W_{all} * 100$ 、 $W_2 = w_2 / W_{all} * 100$ 、 $W_3 = w_3 / W_{all} * 100$ 、 $W_4 = w_4 / W_{all} * 100$ 以及 $W_5 = w_5 / W_{all} * 100$ 。可以理解，当业务设备子集对应为 N 档时，此时集合总权重为 $W_{all} = (w_1 + \dots + w_n) * 100$ ，第 Gm 档的业务设备子集对应的分配权重为： $W_m = w_m / w_{all} * 100$ 。需要说明的是，在本实施例中考虑到排名在 95%-100% 范围内的业务设备的空闲资源信息很小，即这些业务设备当前的负载率已经很高，甚至已经濒临超负荷状态，故而实际情况下可以考虑不向排名在 95%-100% 范围内的业务设备继续分配业务请求。

上述实施例中，通过计算每一个业务设备子集的分配权重，实现了对每一个业务设备子集的空闲资源信息的量化处理，并根据空闲资源的大小为各业务设备子集设置对应大小的分配权重。进而根据各业务设备子集对应的分配权重为其分配对应数量的业务请求，实现对业务请求的合理分配，以及业务设备子集中的业务设备的负载均衡。

如图 5 所示，图 5 为一个实施例中提供的一种按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集步骤的流程示意图。在其中一个实施例中，步骤 S206，按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集的步骤具体还包括：

步骤 S502，从分配码集合中随机确定与业务请求对应的目标分配码。分配码集合包括与多组业务设备子集一一对应的分配码子集。其中，分配码子集的分配码数量与分配码集合的分配码总数的商，与相对应的业务设备子集所对应的分配权重相匹配。

其中，分配码集合中包括多于一个的分配码，并且分配码的数值大小在预设的范围区间内，以及每一个分配码可以对应为相同或者不同的数值，在本申请中不做限制。具体的，分配码可以为随机码，随机码是在计算机设备中随机产生的并且是在预设范围内的一个数值。

分配码子集是对分配码集合中的分配码进行组别划分得到的子集，并且每一个分配码子集中包括多于一个的分配码。具体的，可以对分配码集合中的各分配码按照数值大小进行排序，然后按照预设的组别划分比例将分配码进行组别划分得到多于一个的分配码子集。并且需要说明的是，分配码子集的数量与业务设备子集的数量相同，且每一个分配码子集对应一个业务设备子集。

在其中一个实施例中，计算机设备按照分配码的排序顺序，依次将预设数量的分配码划分至不同组别中，得到多组的分配码子集。如分配码集合中的分配码的取值区间为 0-100，可以将取值区间在 0-10 的分配码划分至 g1 档，将取值区间在 10-25 的分配码划分至 g2 档，将取值区间在 25-50 的分配码划分至 g3 档，将取值区间在 50-85 的分配码划分至 g4 档，将取值区间在 85-95 的分配码划分至 g5 档，并根据不同档的分配码得到多组的分配码子集。

在其中一个实施例中，计算机设备还可以按照排序将分配码划分至不同的组别得到多组的分配码子集，如可以将排序在 0-10% 范围内的分配码划分至 g1 档，将排名在 10-25% 范围内的分配码划分至 g2 档，将排名在 25-50% 范围内的分配码划分至 g3 档，将排名在 50%-85% 范围内的分配码划分至 g4 档，以及将排名在 85-95% 范围内的分配码划分至 g5 档，并根据不同档的分配码得到多组的分配码子集。

在其中一个实施例中，计算机设备将预设数量的分配码划分至不同组别中，得到多组的分配码子集。如分配码集合中的分配码的取值区间为 0-100，可以将不同的 10 个分配码划分至 g1 档，将不同的 15 个分配码划分至 g2 档，将不同的 25 个分配码划分至 g3 档，将不同的 35 个分配码划分至 g4 档，将剩下的 10 个分配码划分至 g5 档，并根据不同档的分配码得到多组的分配码子集。

需要说明的是，每一个业务分配码子集中包括的分配码数量与分配码集合的分配码总数的商，

与相对应的业务设备子集所对应的分配权重相匹配。也就是说分配码子集中包括的分配码数量越多，该分配码子集对应的分配码子集权重也就越大，此时该分配码子集与分配权重较大的业务设备子集相对应，反之，分配码子集中包括的分配码的数量越少，说明该分配码子集的分配码子集权重也越小，此时该分配码子集与分配权重较小的业务设备子集对应。

在其中一个实施例中，计算机设备每接收到一个业务请求，就在业务设备中根据 $\text{rand}() \% 100$ 函数产生一个随机数，并且该随机数的范围在 0-100 之间，并将该随机数作为目标分配码。然后根据随机产生的目标分配码确定匹配的目标业务设备。在其中一个实施例中，也可以理解为分配码集合中存储着 0-100 范围内的数值，每当计算机设备接收到一个业务请求就从分配码集合中随机获取一个数值，然后将该数值作为目标分配码，并根据目标分配码匹配目标业务设备。

步骤 S504，确定目标分配码所属的目标分配码子集。

计算机设备根据目标分配码对应的数值大小，确定目标分配码所属的目标分配码子集。

步骤 S506，将多组业务设备子集中与目标分配码子集对应的业务设备子集，作为与业务请求相匹配的目标业务设备子集。

具体的，计算机设备根据各分配码子集中的分配码数量与分配码集合中的分配码总数量的商确定每一个分配码子集对应的分配码子集权重。并且每一个分配码子集权重都与一个业务设备子集的分配权重相对应。

在其中一个实施例中，确定分配码子集与业务设备子集之间的对应关系包括：确定各分配码子集中的分配码数量、和分配码集合中的分配码总数量，并确定各分配码数量与分配码总数量之间的商；根据各分配码数量与分配码总数量之间的商，确定每一个分配码子集的分配码子集权重；按照分配码子集权重的大小对分配码子集进行排序，并按照分配权重的大小对业务设备子集进行排序；将具有相同排序顺序的分配码子集和业务设备子集进行关联存储，得到分配码子集与业务设备子集之间的对应关系。

具体的，按照分配码子集权重的大小对分配码子集进行排序，以及按照分配权重的大小对业务设备子集进行排序，并根据对应的排序顺序依次匹配对应的分配码子集与业务设备子集，进一步地可以将对应的分配码子集与业务设备子集进行关联对应存储。具体的，较大的分配码子集权重与较大的分配权重相对应，也就是说较大的分配码子集权重对应的分配码子集与较大的分配权重对应的业务设备子集相对应。

计算机设备根据确定的目标分配码子集，在预先对应存储的数据中查找与目标分配码子集匹配的业务设备子集，并将查找到的业务设备子集作为与业务请求相匹配的目标业务设备子集。

上述实施例中，由于每一个分配码子集对应得分配码子集权重是根据各分配码子集中包括得分配码数量确定的，也就是说包含分配码数量越多的分配码子集对应的分配码子集权重越大。在具体的业务场景中，每一个业务请求都对应一个随机确定的目标分配码，由于目标分配码是随机确定的，也就是说目标分配码为分配码集合中任何一个分配码的概率是一样的，故而目标分配码落在每一个分配码子集中的概率与各分配码子集对应的分配码子集权重是一致的。即分配码子集权重越大，随机确定的目标分配码落在其范围内的概率就越大，将业务请求分配至与之对应的目标业务设备子集的概率也就越大。在上述实施例中，巧妙地利用随机事件中的确定性原理，为每一个业务请求匹配对应的目标业务设备子集，实现了分配权重较大的业务设备子集被分配到业务请求的概率就越大，分配权重较小的业务设备子集被分配到业务请求的概率就越小，进而实现了合理分配业务请求，实现业务设备负载均衡的效果。

如图 6 所示，图 6 为一个实施例中提供的一种确定目标分配码所属的目标分配码子集步骤的流程示意图。在其中一个实施例中，确定目标分配码所属的目标分配码子集的步骤具体还包括：

步骤 S602，将多个分配码子集进行排序，得到各个分配码子集各自对应的排序值。

具体地，计算机设备根据分配码子集中包括的分配码数量，对各分配码子集进行排序，得到各个分配码子集各自对应的排序值。

在其中一个实施例中，计算机设备根据分配码子集中包括的分配码的数量对各分配码子集进行排序，得到各分配码子集各自对应的排序值，如表一所示，表一为一个实施例中提供的一种分配码子集信息。在表一中，若第一分配码子集包括的分配码数量为 35 个，第二分配码子集包括的分配码数量为 30 个，第三分配码子集包括的分配码数量为 20 个，第四分配码子集包括的分配码数量为 10 个，第五分配码子集包括的分配码数量为 5 个。那么将多个分配码子集进行排序，得到各个分配码子集各自对应的排序值分别为：第一分配码子集的排序值为第一、第二分配码子集的排序值为第二、第三分配码子集的排序值为第三、第四分配码子集的排序值为第四以及第五分配码子集的排序值为第五。

表一 分配码子集信息

分配码子集	第一分配码子集	第二分配码子集	第三分配码子集	第四分配码子集	第五分配码子集
分配码数量	35	30	20	10	5
排序值	第一	第二	第三	第四	第五
当前排序值	第一	第二	第三	第四	第五
前序排序值	无	第一	第二、第三	第一、第二、第三	第一、第二、第三、第四

步骤 S604，获取当次轮询的当前排序值。

计算机设备按照各分配码子集的排序值依次轮询对应的分配码子集，当次轮询是指在本次的循环过程。当前排序值是指在当次轮询状态下轮询到的排序值。具体的，计算机设备按照排序值从大到小的顺序依次轮询对应的分配码子集，并确定每次轮询时对应的当前排序值。

在其中一个实施例中，参考表一，计算机设备首先轮询排序值在第一的第一分配码子集，并确定当前排序值为第一，然后计算机设备依次轮询排序值在第二的第二分配码子集，并确定当次当前排序值为第二，以及当计算机设备当次轮询至第三分配码子集时，那么当次轮询对应的当前排序值为第三。

步骤 S606，确定排序在当前排序值之前的前序排序值，并确定前序排序值所对应的各分配码子集的前序分配码总数量。

其中，前序排序值是相对于当次排序值而言的，并且前序排序值为当次排序值之前的所有的排序值。

在其中一个实施例中，继续参考表一，当次轮询至第三分配码子集时，那么对应的当前排序值为第三，那么此时对应的前序排序值为第二以及第一，那前序排序值对应的分配码子集的前序分配码总数量分别为第一分配码子集的总数量 35 以及第二分配码子集的总数量 30。也就是说此时对应的前序排序值所对应的各分配码子集的前序分配码总数量为 65。

步骤 S608，将目标分配码与前序分配码总数量间的差值，作为待比较值。

具体地，计算机设备计算根据业务请求随机确定的目标分配码与前序分配码总数量间的差值，根据差值确定待比较值。在其中一个实施例中，当目标分配码为 70、前序分配码总数量为 65 时，那么计算得到的待比较值为 15。

步骤 S610, 当待比较值大于零, 且小于当前排序值所对应的分配码子集的分配码数量时, 确定当前排序值所对应的分配码子集为目标分配码所属的目标分配码子集。

继续上述实施例, 可知待比较值为 15 大于零, 并且待比较值 15 小于当前排序值所对应的分配码子集的分配码数量 20, 故而确定当前排序值所对应的分配码子集即第三分配码子集为目标分配码所属的目标分配码子集。

步骤 S612, 否则, 将当前排序值之后的下一个排序值作为下一次轮询的当前排序值, 并返回确定排序在当前排序值之前的前序排序值的步骤继续执行。

在其中一个实施例中, 继续参考表一, 当目标分配码为 70、当前排序值为第二时, 那么此时前序分配码总数量为 35, 那么计算得到的待比较值为 35。由于待比较值 35 大于零, 但是待比较值 35 大于当前排序值 30, 故而此时将当前排序值之后的下一个排序值作为下一次轮询的当前排序值, 并返回确定排序在当前排序值之前的前序排序值的步骤继续执行。可知执行至当前排序值为第三时, 满足条件, 确定当前排序值所对应的第三分配码子集为目标分配码所属的目标分配码子集。

在其中一个实施例中, 计算机设备每接收到一个业务请求时, 利用函数 $\text{rand}() \% 100$ 随机生成一个随机数 R 作为目标分配码, 若目标分配码 R 对应的目标分配码子集为 W_m , 那么此时应该满足的条件为: $0 < (R - (W_1 + \dots + W_{m-1})) < W_m$, 其中 W_m 为当前排序值, $W_1..W_{m-1}$ 为当前排序值对应的前序排序值。

在其中一个实施例中, 确定目标分配码所属的目标分配码子集包括: 将多个分配码子集进行排序; 从初始排序值开始依次对各个排序值所对应的分配码子集进行轮询, 直至从多个分配码子集中确定与目标分配码所属的目标分配码子集为止; 其中, 在每次轮询中均执行以下步骤: 确定排序在当前排序值之前的前序排序值, 并确定前序排序值所对应的各分配码子集的前序分配码总数量; 将目标分配码与前序分配码总数量间的差值, 作为待比较值; 当待比较值大于零, 且小于当前排序值所对应的分配码子集的分配码数量时, 确定当前排序值所对应的分配码子集为目标分配码所属的目标分配码子集; 否则, 进行下一轮的轮询。

上述实施例中, 确定目标分配码所属的目标分配码子集的过程中, 通过计算每次轮询过程中对应的待比较值, 并根据待比较值的大小确定目标分配码, 使得目标分配码子集的确定简单易行, 提高了目标分配码子集确定的效率。

下面通过举例的方式, 来详细说明对业务设备集合进行组别划分, 得到多组业务设备子集, 从而在获取到业务请求时, 从多组业务设备子集中确定目标业务设备子集的具体过程:

假设业务设备集合中包括 100 台业务设备, 根据业务经验确定将业务设备划分为 5 档, 并且每一档对应的区间范围分别为 [1%-10%]、[11%-25%]、[26%-50%]、[51%-85%]、[86%-95%]。在具体实施中, 在对业务设备进行分配之前, 按照各业务设备对应的可用资源信息 (w_1 、 $w_2 \dots w_{100}$) 对业务设备进行从大到小的排序, 那么将排序在 1-10 的业务设备划分至第一档得到第一档业务设备子集, 该第一档业务设备子集对应为最空闲的一档, 并且该第一档业务设备子集对应的子集权重为: $W_1 = w_1 + w_2 + \dots + w_{10}$ 。其中, w_1 、 $w_2 \dots w_{10}$ 分别为排序在 1-10 的业务设备各自对应的空闲资源。将排序在 11-25 的业务设备划分至第二档得到第二档业务设备子集, 并且该第二档业务设备子集对应的子集权重为: $W_2 = w_{11} + w_{12} + \dots + w_{25}$ 。将排序在 26-50 的业务设备划分至第三档得到第三档业务设备子集, 并且该第三档业务设备子集对应的子集权重为: $W_3 = w_{26} + w_{27} + \dots + w_{50}$ 。将排序在 51-85 的业务设备划分至第四档, 得到第四档业务设备子集, 并且该第四档业务设备子集对应的子集权重为: $W_4 = w_{51} + w_{52} + \dots + w_{85}$ 。将排序在 86-95 的业务设备划分至第五档得到第五档业务设备子集, 并且该第五档业务设备子集为最

忙的一档，对应的子集权重为： $W5=w86+w52+\dots+w95$ 。以及所有的业务设备对应的总权重为： $Wall=w1+w2+\dots+w95 = W1+W2+W3+W4+W5$ 。进一步地，可以根据每一档业务设备子集的子集权重与所有业务设备的总权重比值确定每一档业务设备子集的分配权重。

当计算机设备接收到业务请求时，随机生成与业务请求对应的目标分配码 R ，具体可通过以下公式得到目标分配码： $R=\text{Rand}[1,Wall]-1$ 。并根据 R 与各档次的分配权重的大小关系确定目标分配码 R 所属的档次。具体地，当 $R < W1$ 时，对应为目标分配码从属于第一档，当 $R < W1+W2$ 时，对应为目标分配码从属于第二档，当 $R < W1+W2+W3$ 时，对应为目标分配码从属于第三档，当 $R < W1+W2+W3+W4$ 时，对应为目标分配码从属于第四档，当 $R < W1+W2+W3+W4+W5$ 时，对应为目标分配码从属于第五档。

参考表二，表二为不同的业务设备子集中业务设备的空闲资源信息表。假设业务设备集合中包括 100 个业务设备，根据各业务设备对应的空闲资源信息对业务设备集合中的业务设备进行排序，并将排序在[1-10%]之内的业务设备划分至同一组，得到第一档业务设备子集，其中第一业务设备子集中包括 10 个业务设备，并且每一个业务设备的空闲资源信息为 $6w$ 。将排序在[11%-25%]之内的业务设备划分至一组，得到第二档业务设备子集，其中第二业务设备子集中包括 15 个业务设备，并且每一个业务设备的空闲资源信息为 $5w$ 。将排序在 [26%-50%]之内的业务设备划分至一组，得到第三档业务设备子集，其中第三档业务设备子集中包括 25 个业务设备，并且每一个业务设备的空闲资源信息为 $2w$ 。将排序在[51%-75%]之内的业务设备划分至一组，得到第四档业务设备子集，其中第四业务设备子集中包括 25 个业务设备，并且每一个业务设备的空闲资源信息为 $1w$ 。以及将排序在[76%-100%]之内的业务设备划分至一组，得到第五档业务设备子集，其中第五业务设备子集中包括 25 个业务设备，并且每一个业务设备的空闲资源信息为 $0w$ 。需要说明的是，在第五档业务设备子集中的各业务设备的空闲资源信息几乎为 0，故而业务请求不再分配至第五档业务设备子集中。并且在具体应用场景中，当业务请求为会议请求时，会议请求对应的会议时长远大于对业务设备子集进行更新的时间周期，故而在一个时间周期内（比如 5s 内）不向第五档业务设备子集（高负载的业务设备子集）中分配业务请求，还可以防止不同档次的业务设备子集中负载任务的差距进一步拉大。

表二 不同的业务设备子集中业务设备的空闲资源信息

随机策略					
业务设备子集	第一档业务设备子集	第二档业务设备子集	第三档业务设备子集	第四档业务设备子集	第五档业务设备子集
区间范围	[1-10%]	[11%-25%]	[26%-50%]	[51%-75%]	[76%-100%]
业务设备1	$6w$	$5w$	$2w$	$1w$	0
业务设备2	$6w$	$5w$	$2w$	$1w$	0
业务设备...	...	...	...	...	...
业务设备m	$6w$	$5w$	$2w$	$1w$	0
子集权重	$6w*10$	$5w*15$	$2w*25$	$1w*25$	0

从表二中看出，每一档业务设备子集中的业务设备的空闲资源信息大体上可以认为是相同的。在其中一个实施例中，可以将分配到目标业务设备子集中的业务请求按照轮询策略分配至目标业务设备子集中的业务设备中。需要说明的是，在其他实施例中，每一个业务设备子集中的各业务设备的空闲资源信息也可以具有较大差异，在本申请中不做限定。

在其中一个实施例中，步骤 S208，从目标业务设备子集中筛选目标业务设备的具体步骤还包括：确定目标业务设备子集中各业务设备的排队顺序，将按照排队顺序进行轮询时当前所轮询至的业务设

备，作为当前的目标业务设备。

具体地，在目标业务设备子集中，可根据各业务设备分别对应的空闲资源信息对各业务设备进行排序，或者各业务设备在业务设备子集中也可以是随机排列的。

在其中一个实施例中，将各业务设备在目标业务设备中的排序，当业务请求分配至目标业务设备子集后，将按照排队顺序进行轮询时当前所轮询至的业务设备，作为当前的目标业务设备。也就是说，按照排队顺序依次在目标业务设备子集中轮询业务设备，并将在当次轮询到的业务设备作为与当次的业务请求对应的目标业务设备。通过轮询算法实现筛选目标业务设备，使得目标业务设备子集中的业务设备被确定为目标业务设备的概率较为一致，实现了目标业务设备子集中的业务设备的负载均衡。

在其中一个实施例中，业务请求分配至目标业务设备子集后，按照各业务设备在目标业务设备中的排序，从中筛选一个空闲资源信息最大的业务设备作为目标设备，然后将业务请求分配至对应的目标业务设备中。通过将业务请求分配至空闲资源信息最大的目标业务设备中，以利用执行能力最强的目标业务设备执行业务请求，提高了业务请求的处理效率，并且也实现了目标业务设备子集中业务设备的负载均衡。

在另一个实施例中，业务请求分配至目标业务设备子集后，从目标业务设备子集中随机选取一个业务设备作为目标业务设备，然后将业务请求分配至对应的目标业务设备中。由于目标业务设备子集中的各业务设备的空闲资源大致相同，故而可以利用目标业务设备子集中的任何一个业务设备作为目标业务设备来处理业务请求，不需要利用复杂的算法筛选目标业务设备，使得算法更加高效，较少计算机资源的利用，提高了对业务请求处理的效率。

如图 7 所示，图 7 为一个实施例中提供的一种将业务请求分配至目标业务设备步骤的流程示意图。在其中一个实施例中，步骤 S210，将业务请求分配至目标业务设备的步骤具体还可以包括：

步骤 S702，获取目标业务设备对应的目标地址，并根据目标地址将业务请求分配至对应的目标业务设备。

其中，目标地址可以为 ip 地址，并且目标地址与目标业务设备唯一对应。具体地，计算机设备根据目标地址将获取到的业务请求发送至对应的目标业务设备。

在其中一个实施例中，业务请求分配方法还包括：

步骤 S704，确定业务请求中携带的请求访问地址。

请求访问地址是业务请求中携带的、并且需要访问的地址。具体地，目标地址与请求访问地址并不是同一个地址。

在其中一个实施例中，计算机设备接收到业务请求时，从业务请求中提取携带的请求访问地址，然后计算机设备为该业务请求分配对应的目标业务设备，并获取目标业务设备对应的目标地址，并把请求访问地址修改为目标地址，以根据目标地址将业务请求发送至对应的目标业务设备。在本步骤中，实现为业务请求匹配对应的目标业务设备，并同时根据目标业务设备对应的目标地址将业务请求分配至对应的目标业务设备中。

步骤 S706，接收由目标业务设备发送的、且在执行业务请求对应的业务操作时所得到的操作数据。

操作数据是目标业务设备接收到对应的业务请求后，并对业务请求执行对应的业务操作后对应的数据。然后计算机设备获取目标业务设备执行业务操作所得到的操作数据，以将操作数据传输至发送业务请求的用户端。

步骤 S708，响应于业务请求，通过请求访问地址反馈操作数据。

计算机设备获取到目标业务设备产生的操作数据后，将目标地址修改为请求访问地址，并根据请

求访问地址将操作数据反馈至发送业务请求的用户端。

如图 8 所示，图 8 为一个实施例中提供的一种根据 IP 地址分配业务请求的示意图。在图 8 中，计算机设备可以是路由服务器（route），业务设备集合可以为媒体代理设备集合，业务设备集合中的业务设备可以为媒体代理设备集合中的多个媒体代理设备，并且媒体代理设备集合中包括 N 组的媒体代理设备，具体为媒体代理设备 1、媒体代理设备 2...媒体代理设备 N。具体的，用户终端的一个或者多个会议用户向路由服务器（route）发送获取媒体代理设备 ip 地址的请求，路由服务器根据本申请提供的一种业务请求分配方法为用户发送的请求匹配对应的媒体代理设备，当路由服务器匹配到的媒体代理设备为媒体代理设备 2 时，那么执行将用户与媒体代理设备 2 建立媒体连接，以实现用户端与媒体代理设备 2 的通信连接。进一步地，还可以在媒体代理设备 2 中再选择一个目标媒体代理设备，以利用目标媒体代理设备实现与用户端的通信连接。在本实施例中，会议用户发送请求接入到媒体代理设备的 ip 地址，考虑到对媒体代理设备的 ip 地址分配直接影响媒体代理设备的使用率，故而通过本申请提供的分配方法为会议用户合理分配媒体代理设备的 ip 地址，实现了对媒体代理设备资源的合理利用，实现了媒体代理设备集合中各媒体代理设备的负载均衡。

上述实施例中，当计算机设备接收到用户发送的业务请求时，从业务请求中提取请求访问地址，并通过业务请求分配方法为业务请求匹配对应的目标业务设备。然后将请求访问地址修改为目标业务设备对应得目标地址，以根据目标地址实现不同设备之间得通信连接，使得利用目标业务设备执行业务设备成为了可能，提高了对业务请求处理的效率。

在其中一个实施例中，待处理的业务请求为会议连接请求，业务请求分配方法还包括：接收会议连接请求，会议连接请求中携带有访问请求地址。根据业务请求分配方法确定与会议连接请求匹配的目标业务设备，并获取目标业务设备对应的目标地址，将访问请求地址修改为目标地址，并将会议连接请求分配至目标地址对应的目标业务设备，以指示目标业务数据执行对会议连接请求的连接操作。

不同的会议连接请求差异比较大，如有的会议连接请求对应的会议时长可能是几个小时，有的会议连接请求对应的会议时长可能只是几分钟。且会议连接请求的毛刺现象严重，如会议连接请求为整点会议的概率较大，如在早上 8 点或者早上 9 点早会期间对应的会议连接请求较多。基于以上问题，若利用传统的会议连接请求分配方法在每接收到一次会议连接请求后，都需要重新统计业务设备集合中各业务设备的空闲资源，并实时对各业务设备的空闲资源进行更新，造成计算量大，且造成系统不稳定。本申请每 N 秒对业务设备集合中的各业务设备的空闲资源收集一次，并根据各业务设备的空闲资源对业务设备集合中所有的业务设备进行排序，并按排序顺序进行业务设备子集的划分，以及计算每一个业务设备子集的分配权重。以将用户发送的会议连接请求分配权重分配到不同的业务设备子集中，并在业务设备子集中进行轮询分配会议连接请求，不仅实现了对会议连接请求的合理分配，同时也实现了业务设备的负载均衡。特别是在会议用户使用情况差异大的场景下，将空闲资源相近的业务设备放入相同的业务设备子集，在不同的业务设备子集之间按分配权重进行分配，在同一个业务设备子集内部进行轮询分配，降低了算法的计算量，同时兼顾了不同分配权重的业务设备子集的分配概率不同，做到了负载均衡。

以及，本申请对业务设备集合中所有业务设备的空闲资源进行量化，可支持不同类型的业务设备，该算法可扩展性强，能进行分布式部署，提高了业务设备的可扩展性。并且本申请对会议场景的负载均衡实现简单、高率（空闲资源每 N 秒计算一次），对现业务设备集合中的空闲资源的收集时延要求相对较低。

结合图 9，图 9 为一个实施例中根据业务请求分配方法执行业务请求分配时各业务设备对应的资

源使用率示意图。在图 9 中，竖坐标为分配权重，横坐标为业务设备的设备 id (1-8000)。其中每个业务设备的负载使用率上限是 10W，从图 9 中可以看出业务设备集中的业务设备的负载使用率分布在 5w-7W 之间，基本上实现了业务设备的负载均衡。

通过本申请提供的业务请求分配方法计算机设备每隔 N 秒执行一次业务设备的分组步骤，在一个时间周期内只收集一次业务设备的当前可用资源信息，并计算一次分配权重即可，可以简单、高效率地统计各业务设备子集的分配权重，不需要每执行分配一次业务请求就重新计算一次分配权重。对收集业务设备集合中各业务设备的当前空闲信息的实时性要求降低了，并且同时提高了系统的可靠性。并且本方法可以支撑不同类型和处理能力的业务设备同时服务，还可以分布式部署各业务设备，特别是当业务请求的数量比较大的情况下，还可以对业务设备进行水平扩展，提高业务设备的处理能力。

在其中一个实施例中，待处理的业务请求包括会议连接请求，业务设备包括媒体代理设备，业务设备的设备资源包括实现用户接入的资源、实现虚拟房间申请的资源、及网络流量资源中的至少一种。

本申请还提供一种应用场景，该应用场景应用上述的业务请求分配方法。具体地，该业务请求方法在该应用场景的应用如下：

以计算机设备为路由服务器、业务请求为会议连接请求、业务设备集合为媒体代理设备集合、业务设备为媒体代理设备为例进行说明。具体的，路由服务器在当前时间周期内获取媒体代理设备集合中的各媒体代理设备的有效资源信息以及当前空闲资源信息，并根据有效资源信息以及当前空闲资源信息的比值确定各媒体代理设备的可用资源信息，并根据可用资源信息对各媒体代理设备进行排序。根据预先确定的组别划分比例，将各媒体代理设备划分至不同的组别得到当前时间周期内对应的多组的媒体代理设备子集。并根据各媒体代理设备子集中包括的媒体代理设备的可用资源信息计算各媒体代理设备子集的分配权重。其中，分配权重的范围为 0-100，并且分配权重越大的媒体代理设备子集被分配会议连接请求的概率越大。

路由服务器在当前时间周期内接收到会议用户在用户端中发送会议连接请求后，随机获取一个范围在 0-100 范围内的目标分配码，然后根据目标分配码落在的分配权重区间确定目标分配码对应的目标媒体代理设备子集，并将会议连接请求分配至目标媒体代理设备子集中，以及在目标媒体代理设备子集中轮询访问媒体代理设备，并将在当次轮询到的媒体代理设备作为处理业务请求的目标媒体代理设备。路由服务器获取目标媒体代理设备对应的目标地址，并根据目标地址将会议连接请求发送至目标媒体代理设备，当目标媒体代理设备完成对会议连接请求的业务操作后，将得到的操作数据通过目标媒体代理设备对应的目标地址传送至路由服务器中。路由服务器将操作数据发送至会议用户，完成会议连接操作。

当完成在本次时间周期内的会议连接请求时，路由服务器再次重新统计媒体代理设备集合中各媒体代理设备的当前空闲资源信息，并重新根据当前空闲资源信息进行媒体代理设备子集的划分，得到多组的媒体代理设备子集。使得在下一个时间周期内接收到会议连接请求时，根据重新确定的媒体代理设备子集执行目标媒体代理设备的确定步骤。通过在不同的媒体代理设备子集中按照分配权重进行会议连接请求的分配，以及在目标媒体代理设备子集中随机轮询确定目标媒体代理设备，简单高效，提高了会议连接请求的分配效率。

应该理解的是，虽然图 2-7 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-7 中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤

或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 10 所示，提供了一种业务请求分配装置 1000，该装置可以采用软件模块或硬件模块，或者是二者的结合成为计算机设备的一部分，该装置具体包括：获取模块 1002、第一确定模块 1004、筛选模块 1008 和分配模块 1010，其中：

获取模块 1002，用于获取待处理的业务请求。

第一确定模块 1004，用于确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集是基于业务设备集合中各业务设备的空闲资源进行划分得到。

第二确定模块 1006，用于按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集。

筛选模块 1008，用于从目标业务设备子集中筛选目标业务设备。

分配模块 1010，用于将业务请求分配至目标业务设备；目标业务设备用于执行与分配的业务请求对应的业务操作。

在其中一个实施例中，第一确定模块 1004 还用于在当前时间周期内，确定业务设备集合中的多于一个的业务设备分别对应的可用资源信息；根据可用资源信息对每一个业务设备进行排序；按照预先确定的组别划分比例，将排序后的每一个业务设备划分至对应的组别；其中，从属于同一个组别的业务设备共同构成一组业务设备子集。

在其中一个实施例中，第一确定模块 1004 还用于确定业务设备集合中的各业务设备分别对应的有效资源信息；在当前时间周期内，收集业务设备集合中的各业务设备分别对应的当前空闲资源信息；根据当前空闲资源信息与有效资源信息的对比值，确定每一个业务设备分别对应的可用资源信息。

在其中一个实施例中，有效资源信息包括可接入用户最大数量、可申请虚拟房间最大数量、入流量最大值、以及出流量最大值。当前空闲资源信息包括当前可接入用户数量、当前可接入虚拟房间数量、当前入流量空闲值、以及当前出流量空闲值。第一确定模块 1004 还用于：根据各业务设备分别对应的当前可接入用户数量和可接入用户最大数量，确定各业务设备各自对应的第一对比值；根据各业务设备分别对应的当前可申请虚拟房间数量和可申请虚拟房间最大数量，确定各业务设备各自对应的第二对比值；根据各业务设备分别对应的当前入流量空闲值和入流量最大值，确定各业务设备各自对应的第三对比值；根据各业务设备分别对应的当前出流量空闲值和出流量最大值，确定各业务设备各自对应的第四对比值；及对于业务设备集中的每个业务设备，均将相应业务设备对应的第一对比值、第二对比值、第三对比值和第四对比值中的最小值，作为与相应业务设备对应的可用资源信息。

在其中一个实施例中，业务请求分配装置 1000 还包括第三确定模块 1012，第三确定模块 1012 用于分别根据每组业务设备子集中各业务设备的空闲资源，确定每组业务设备子集各自对应的子集权重；根据每组业务设备子集各自对应的子集权重，确定与业务设备集合对应的集合总权重；分别根据每组业务设备子集各自对应的子集权重与集合总权重的对比值，确定每组业务设备子集分别对应的分配权重。

在其中一个实施例中，第二确定模块 1006 还用于从分配码集合中随机确定与业务请求对应的目标分配码；分配码集合包括与多组业务设备子集一一对应的分配码子集，其中，分配码子集的分配码数量与分配码集合的分配码总数的商，与相对应的业务设备子集所对应的分配权重相匹配；确定目标分配码所属的目标分配码子集；将多组业务设备子集中与目标分配码子集对应的业务设备子集，作为与业务请求相匹配的目标业务设备子集。

在其中一个实施例中，第二确定模块 1006 还用于将多个分配码子集进行排序，得到各个分配码子集各自对应的排序值；获取当次轮询的当前排序值；确定排序在当前排序值之前的前序排序值，并确定前序排序值所对应的各分配码子集的前序分配码总数量；将目标分配码与前序分配码总数量间的差值，作为待比较值；当待比较值大于零，且小于当前排序值所对应的分配码子集的分配码数量时，确定当前排序值所对应的分配码子集为目标分配码所属的目标分配码子集；否则，将当前排序值之后的下一个排序值作为下一次轮询的当前排序值，并返回确定排序在当前排序值之前的前序排序值的步骤继续执行。

在其中一个实施例中，筛选模块 1008 还用于确定目标业务设备子集中各业务设备的排队顺序；将按照排队顺序进行轮询时当前所轮询至的业务设备，作为当前的目标业务设备。

在其中一个实施例中，分配模块 1010 还用于获取目标业务设备对应的目标地址，并根据目标地址将业务请求分配至对应的目标业务设备；业务请求分配装置反馈模块 1014，反馈模块 1014 用于确定业务请求中携带的请求访问地址；接收由目标业务设备发送的、且在执行业务请求对应的业务操作时所得到的操作数据；响应于业务请求，通过请求访问地址反馈操作数据。

在其中一个实施例中，待处理的业务请求包括会议连接请求，业务设备包括媒体代理设备，业务设备的设备资源包括实现用户接入的资源、实现虚拟房间申请的资源、及网络流量资源中的至少一种。

关于业务请求分配装置的具体限定可以参见上文中对于业务请求分配方法的限定，在此不再赘述。上述业务请求分配装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

上述业务请求分配装置，确定对业务设备集合进行划分得到的多组业务设备子集，其中，多组业务设备子集基于业务设备集合中的多于一个的业务设备的空闲资源进行划分得到，使得每一组业务设备子集中包括一个或者多个的负载情况相近的业务设备。当获取到待处理的业务请求时，就可按照每一组业务设备子集分别对应的分配权重，确定与业务请求相匹配的目标业务设备子集。再将业务请求分配至目标业务设备子集中的目标业务设备，以进行相应的业务操作。由于目标业务子集中包括多个的业务设备，故而业务设备子集的数量大大小于业务设备的数量，通过先将业务请求分配至目标业务设备子集中而不是直接分配至某一个具体的业务设备中，减少了分配过程中的计算量，提高了业务请求的分配效率。并且根据各业务设备子集对应的分配权重将业务请求分配至对应的目标业务设备子集中，实现了对业务请求的合理分配，使得各个业务设备子集间满足负载均衡。这样，先按照负载均衡的方向，从粗粒度上将业务请求分配至目标业务设备子集中，然后再将业务请求分配至目标业务设备子集中的某个目标业务设备，既能满足对业务请求进行分配时的负载均衡，又无需实时进行大量的负载更新计算，极大地减少了计算量，大大提高了业务请求的分配效率。

在其中一个实施例中，提供了一种计算机设备，该计算机设备可以是路由设备，其内部结构图可以如图 11 所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储与各业务设备相关的数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种业务请求分配方法。

本领域技术人员可以理解，图 11 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在其中一个实施例中，还提供了一种计算机设备，包括存储器和一个或多个处理器，存储器中存储有计算机可读指令，该计算机可读指令被处理器执行时，使得一个或多个处理器执行上述各方法实施例中的步骤。

在其中一个实施例中，提供了一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行上述各方法实施例中的步骤。

在其中一个实施例中，提供了一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上述各方法实施例中的步骤。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，计算机程序可存储于一非易失性计算机可读存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器（Read-Only Memory, ROM）、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器（Random Access Memory, RAM）或外部高速缓冲存储器。作为说明而非局限，RAM 可以是多种形式，比如静态随机存取存储器（Static Random Access Memory, SRAM）或动态随机存取存储器（Dynamic Random Access Memory, DRAM）等。

权利要求书

1、一种业务请求分配方法，其特征在于，由计算机设备执行，所述方法包括：

获取待处理的业务请求；

确定对业务设备集合进行划分得到的多组业务设备子集，其中，所述多组业务设备子集是基于所述业务设备集合中各业务设备的空闲资源进行划分得到；

按照每一组业务设备子集分别对应的分配权重，确定与所述业务请求相匹配的目标业务设备子集；从所述目标业务设备子集中筛选目标业务设备；及

将所述业务请求分配至所述目标业务设备；所述目标业务设备用于执行与分配的所述业务请求对应的业务操作。

2、根据权利要求1所述的方法，其特征在于，所述确定对业务设备集合进行划分得到的多组业务设备子集，包括：

在当前时间周期内，确定所述业务设备集合中的多于一个的业务设备分别对应的可用资源信息；

根据所述可用资源信息对每一个所述业务设备进行排序；及

按照预先确定的组别划分比例，将排序后的每一个业务设备划分至对应的组别；其中，从属于同一个组别的业务设备共同构成一组业务设备子集。

3、根据权利要求2所述的方法，其特征在于，所述在当前时间周期内，确定所述业务设备集合中的多于一个的业务设备分别对应的可用资源信息，包括：

确定所述业务设备集合中的各业务设备分别对应的有效资源信息；

在当前时间周期内，收集所述业务设备集合中的各业务设备分别对应的当前空闲资源信息；及

根据所述当前空闲资源信息与所述有效资源信息的对比值，确定每一个所述业务设备分别对应的可用资源信息。

4、根据权利要求3所述的方法，其特征在于，所述有效资源信息包括可接入用户最大数量、可申请虚拟房间最大数量、入流量最大值、以及出流量最大值；所述当前空闲资源信息包括当前可接入用户数量、当前可接入虚拟房间数量、当前入流量空闲值、以及当前出流量空闲值；

所述根据所述当前空闲资源信息与所述有效资源信息的对比值，确定每一个所述业务设备分别对应的可用资源信息，包括：

根据各业务设备分别对应的所述当前可接入用户数量和所述可接入用户最大数量，确定各业务设备各自对应的第一对比值；

根据各业务设备分别对应的所述当前可申请虚拟房间数量和所述可申请虚拟房间最大数量，确定各业务设备各自对应的第二对比值；

根据各业务设备分别对应的所述当前入流量空闲值和所述入流量最大值，确定各业务设备各自对应的第三对比值；

根据各业务设备分别对应的所述当前出流量空闲值和所述出流量最大值，确定各业务设备各自对应的第四对比值；及

对于所述业务设备集中的每个业务设备，均将相应业务设备对应的第一对比值、第二对比值、第三对比值和第四对比值中的最小值，作为与相应业务设备对应的可用资源信息。

5、根据权利要求1所述的方法，其特征在于，所述方法还包括：

分别根据每组所述业务设备子集中各业务设备的空闲资源，确定每组所述业务设备子集各自对应的子集权重；

根据每组所述业务设备子集各自对应的子集权重，确定与所述业务设备集合对应的集合总权重；
及

分别根据每组所述业务设备子集各自对应的子集权重与所述集合总权重的对比值，确定每组业务设备子集分别对应的分配权重。

6、根据权利要求1所述的方法，其特征在于，所述按照每一组业务设备子集分别对应的分配权重，确定与所述业务请求相匹配的目标业务设备子集，包括：

从分配码集合中随机确定与所述业务请求对应的目标分配码；所述分配码集合包括与多组业务设备子集一一对应的分配码子集，其中，所述分配码子集的分配码数量与所述分配码集合的分配码总数量的商，与相对应的业务设备子集所对应的分配权重相匹配；

确定所述目标分配码所属的目标分配码子集；及

将所述多组业务设备子集中与所述目标分配码子集对应的业务设备子集，作为与所述业务请求相匹配的目标业务设备子集。

7、根据权利要求6所述的方法，其特征在于，所述确定所述目标分配码所属的目标分配码子集包括：

将多个所述分配码子集进行排序，得到各个分配码子集各自对应的排序值；

获取当次轮询的当前排序值；

确定排序在所述当前排序值之前的前序排序值，并确定所述前序排序值所对应的各分配码子集的前序分配码总数量；

将所述目标分配码与所述前序分配码总数量间的差值，作为待比较值；

当所述待比较值大于零，且小于所述当前排序值所对应的分配码子集的分配码数量时，确定所述当前排序值所对应的分配码子集为所述目标分配码所属的目标分配码子集；及

否则，将所述当前排序值之后的下一个排序值作为下一次轮询的当前排序值，并返回所述确定排序在所述当前排序值之前的前序排序值的步骤继续执行。

8、根据权利要求6所述的方法，其特征在于，所述方法还包括：

确定各所述分配码子集中的分配码数量、和所述分配码集合中的分配码总数量，并确定各所述分配码数量与所述分配码总数量之间的商；

根据各所述分配码数量与所述分配码总数量之间的商，确定每一个分配码子集的分配码子集权重；

按照分配码子集权重的大小对所述分配码子集进行排序，并按照分配权重的大小对所述业务设备子集进行排序；及

将具有相同排序顺序的分配码子集和业务设备子集进行关联存储，得到分配码子集与业务设备子集之间的对应关系。

9、根据权利要求1所述的方法，其特征在于，所述从所述目标业务设备子集中筛选目标业务设备，包括：

确定所述目标业务设备子集中各业务设备的排队顺序；及

将按照所述排队顺序进行轮询时当前所轮询至的业务设备，作为当前的目标业务设备。

10、根据权利要求9所述的方法，其特征在于，所述确定所述目标业务设备子集中各业务设备的排队顺序，包括：

根据目标业务设备子集中各业务设备对应的空闲资源，对各业务设备进行排队，得到所述目标业务设备子集中各业务设备的排队顺序。

11、根据权利要求 1 所述的方法，其特征在于，所述将所述业务请求分配至所述目标业务设备，包括：

获取所述目标业务设备对应的目标地址，并根据所述目标地址将所述业务请求分配至对应的目标业务设备；

所述方法还包括：

确定所述业务请求中携带的请求访问地址；

接收由所述目标业务设备发送的、且在执行所述业务请求对应的业务操作时所得到的操作数据；及

响应于所述业务请求，通过所述请求访问地址反馈所述操作数据。

12、根据权利要求 1 至 11 任意一项所述的方法，其特征在于，所述待处理的业务请求包括会议连接请求，所述业务设备包括媒体代理设备，所述业务设备的设备资源包括实现用户接入的资源、实现虚拟房间申请的资源、以及网络流量资源中的至少一种。

13、一种业务请求分配装置，其特征在于，所述装置包括：

获取模块，用于获取待处理的业务请求；

第一确定模块，用于确定对业务设备集合进行划分得到的多组业务设备子集，其中，所述多组业务设备子集是基于所述业务设备集合中各业务设备的空闲资源进行划分得到；

第二确定模块，用于按照每一组业务设备子集分别对应的分配权重，确定与所述业务请求相匹配的目标业务设备子集；

筛选模块，用于从所述目标业务设备子集中筛选目标业务设备；及

分配模块，用于将所述业务请求分配至所述目标业务设备；所述目标业务设备用于执行与分配的所述业务请求对应的业务操作。

14、根据权利要求 13 所述的装置，其特征在于，所述第一确定模块还用于：在当前时间周期内，确定所述业务设备集合中的多于一个的业务设备分别对应的可用资源信息；根据所述可用资源信息对每一个所述业务设备进行排序；及按照预先确定的组别划分比例，将排序后的每一个业务设备划分至对应的组别；其中，从属于同一个组别的业务设备共同构成一组业务设备子集。

15、根据权利要求 14 所述的装置，其特征在于，所述第一确定模块还用于：确定所述业务设备集合中的各业务设备分别对应的有效资源信息；在当前时间周期内，收集所述业务设备集合中的各业务设备分别对应的当前空闲资源信息；及根据所述当前空闲资源信息与所述有效资源信息的对比值，确定每一个所述业务设备分别对应的可用资源信息。

16、根据权利要求 15 所述的装置，其特征在于，所述有效资源信息包括可接入用户最大数量、可申请虚拟房间最大数量、入流量最大值、以及出流量最大值。当前空闲资源信息包括当前可接入用户数量、当前可接入虚拟房间数量、当前入流量空闲值、以及当前出流量空闲值。所述第一确定模块还用于：根据各业务设备分别对应的所述当前可接入用户数量和所述可接入用户最大数量，确定各业务设备各自对应的第一对比值；根据各业务设备分别对应的所述当前可申请虚拟房间数量和所述可申请虚拟房间最大数量，确定各业务设备各自对应的第二对比值；根据各业务设备分别对应的所述当前入流量空闲值和所述入流量最大值，确定各业务设备各自对应的第三对比值；根据各业务设备分别对应的所述当前出流量空闲值和所述出流量最大值，确定各业务设备各自对应的第四对比值；及对于所述业务设备集中的每个业务设备，均将相应业务设备对应的第一对比值、第二对比值、第三对比值和第四对比值中的最小值，作为与相应业务设备对应的可用资源信息。

17、根据权利要求 13 所述的装置，其特征在于，所述业务请求分配装置还包括第三确定模块，所述第三确定模块用于分别根据每组所述业务设备子集中各业务设备的空闲资源，确定每组所述业务设备子集各自对应的子集权重；根据每组所述业务设备子集各自对应的子集权重，确定与所述业务设备集合对应的集合总权重；及分别根据每组所述业务设备子集各自对应的子集权重与所述集合总权重的对比值，确定每组业务设备子集分别对应的分配权重。

18、根据权利要求 13 所述的装置，其特征在于，所述第二确定模块还用于从分配码集合中随机确定与所述业务请求对应的目标分配码；所述分配码集合包括与多组业务设备子集一一对应的分配码子集，其中，所述分配码子集的分配码数量与所述分配码集合的分配码总数量的商，与相对应的业务设备子集所对应的分配权重相匹配；确定所述目标分配码所属的目标分配码子集；及将所述多组业务设备子集中与所述目标分配码子集对应的业务设备子集，作为与所述业务请求相匹配的目标业务设备子集。

19、一种计算机设备，包括存储器和一个或多个处理器，所述存储器存储有计算机可读指令，其特征在于，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行权利要求 1 至 12 中任一项所述的方法的步骤。

20、一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1 至 12 中任一项所述方法的步骤。

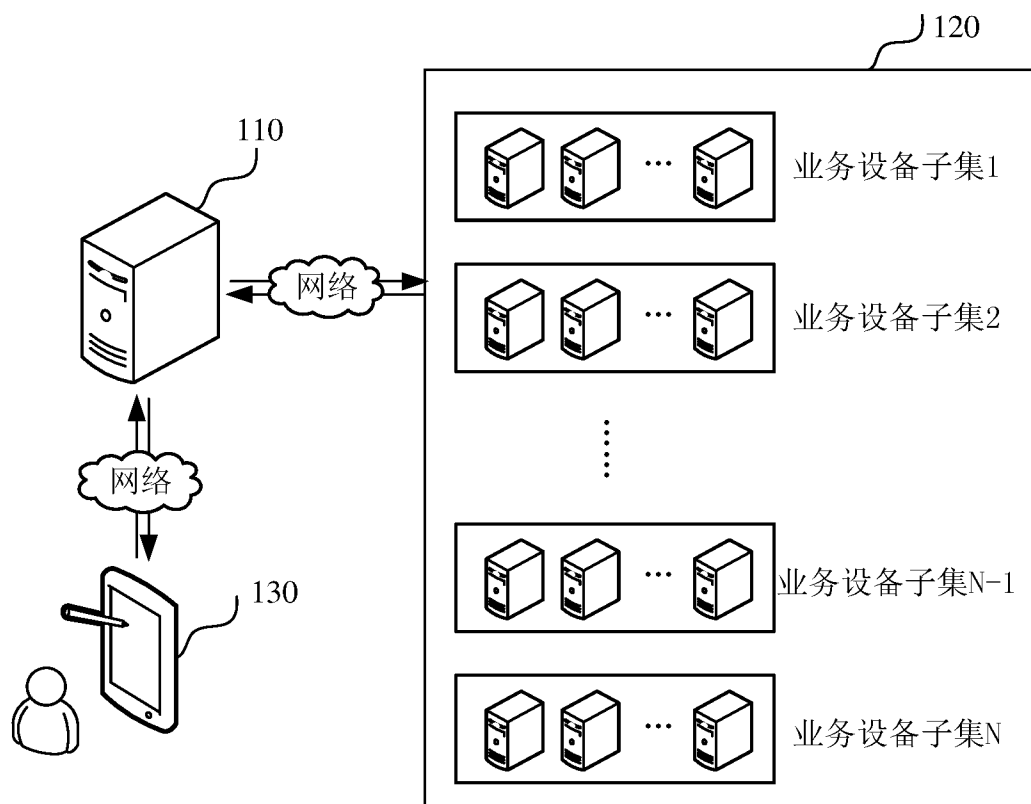


图 1

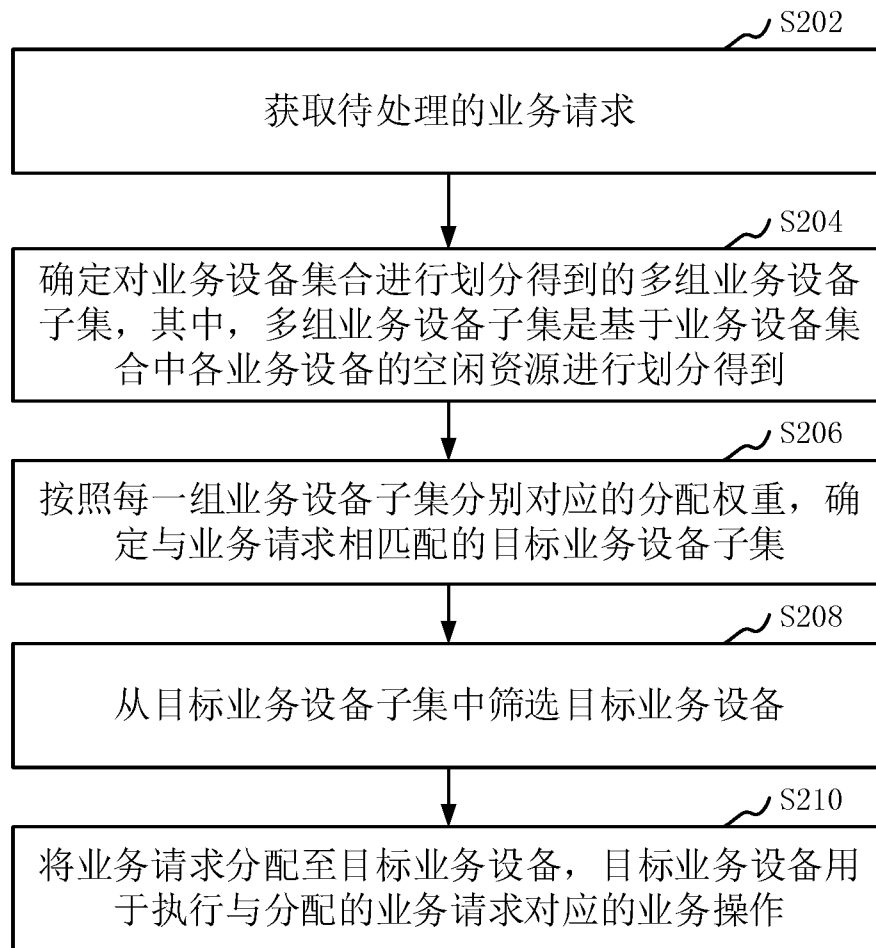


图 2

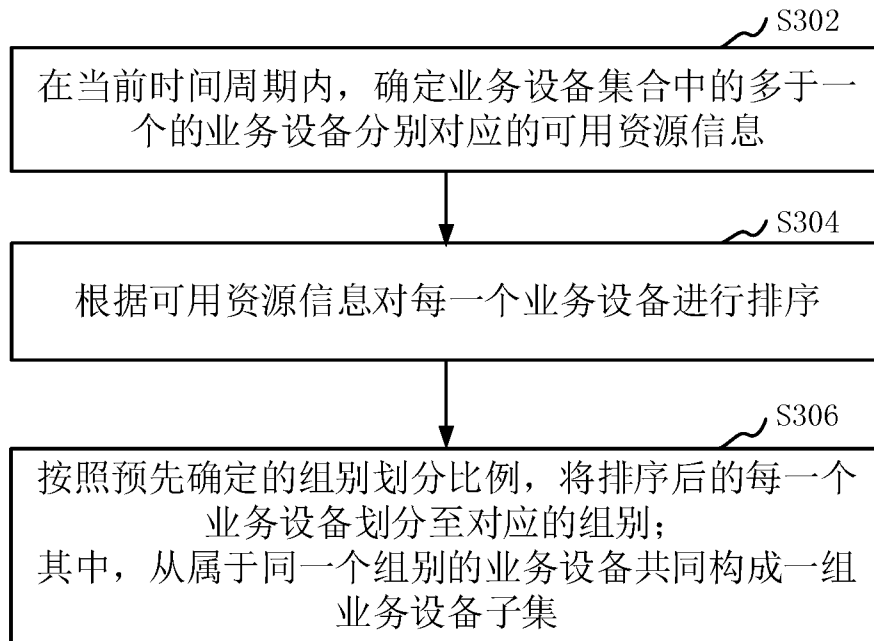


图 3

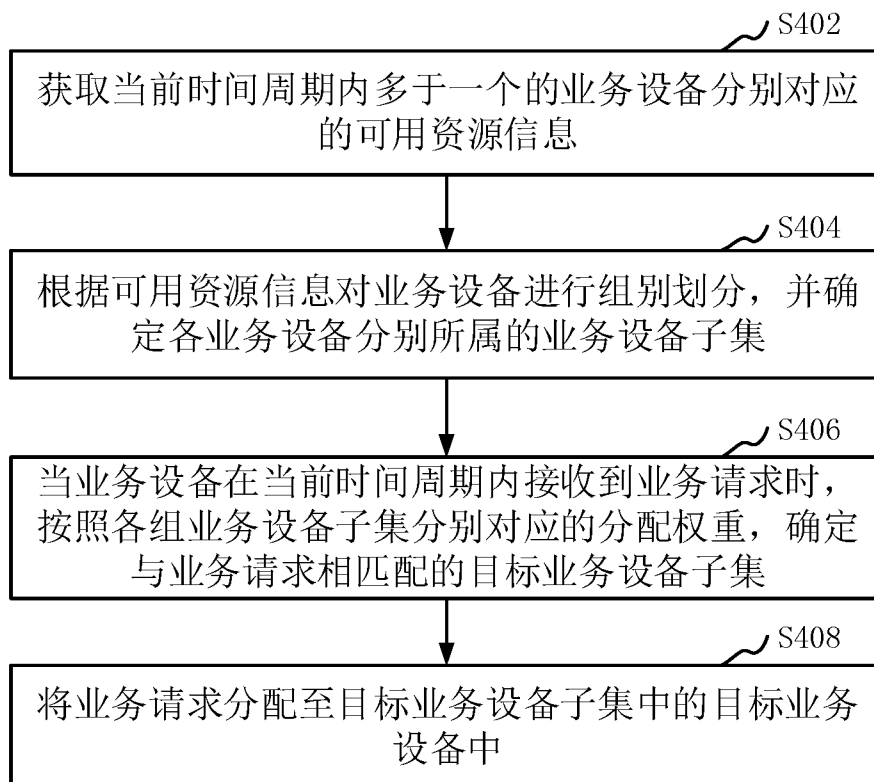


图 4

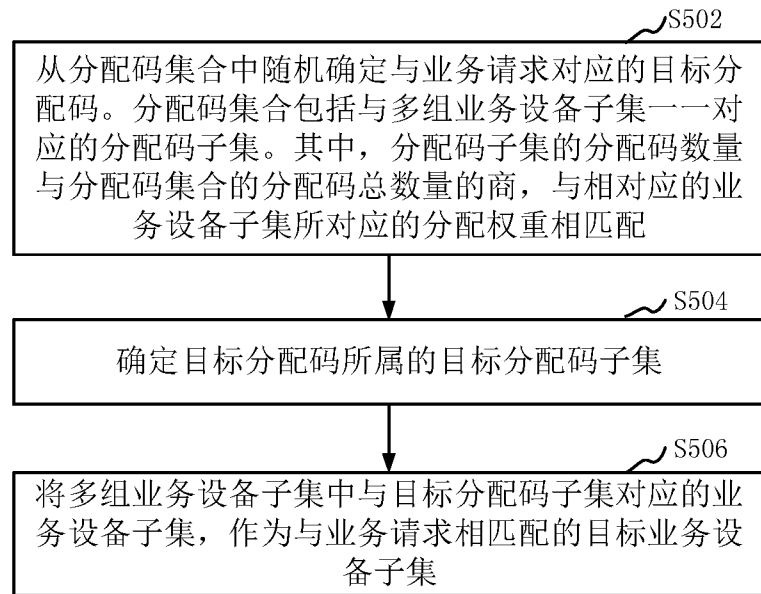


图 5

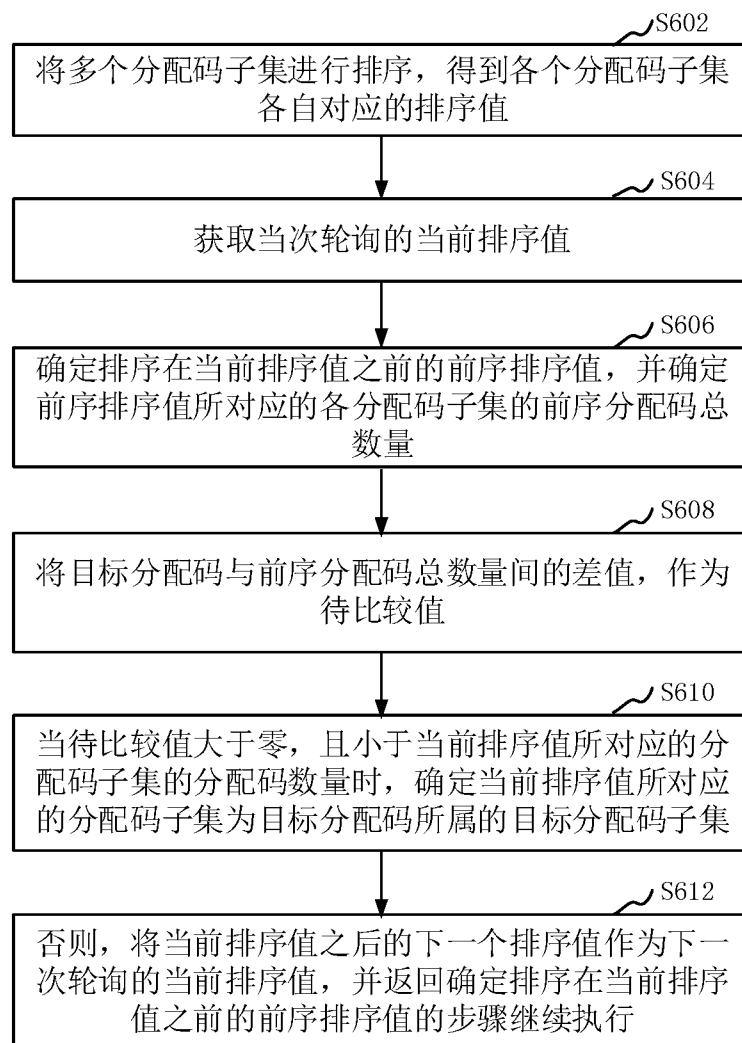


图 6

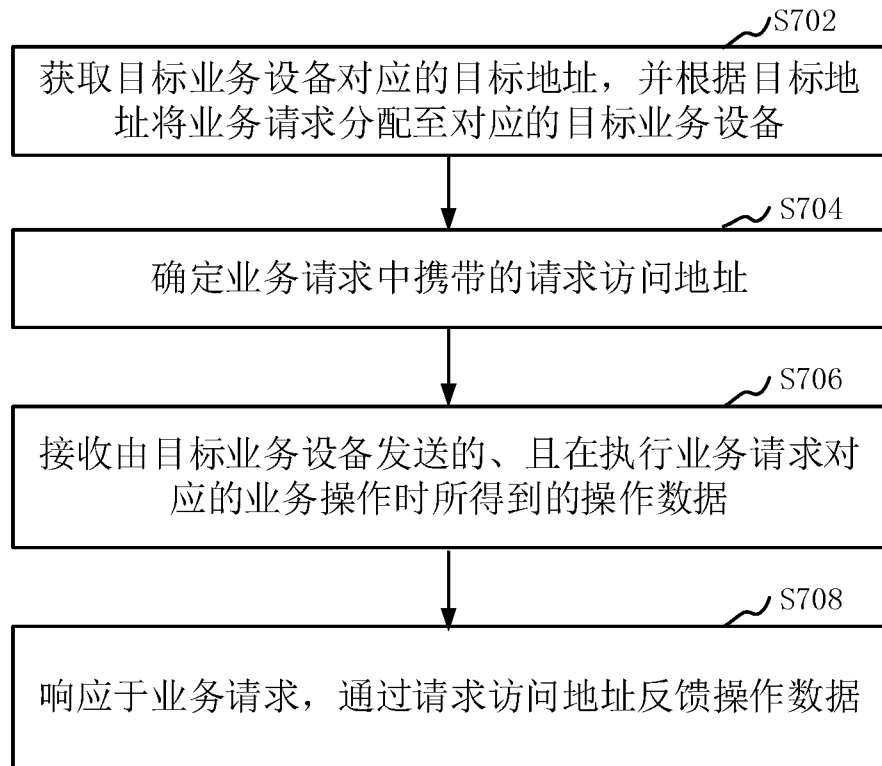


图 7

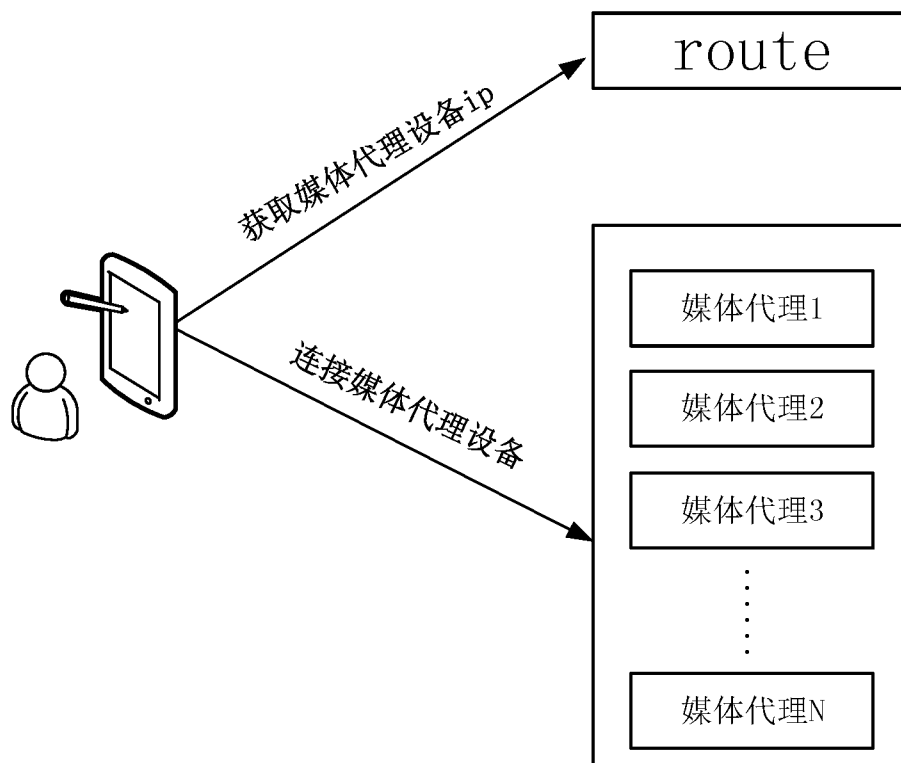


图 8

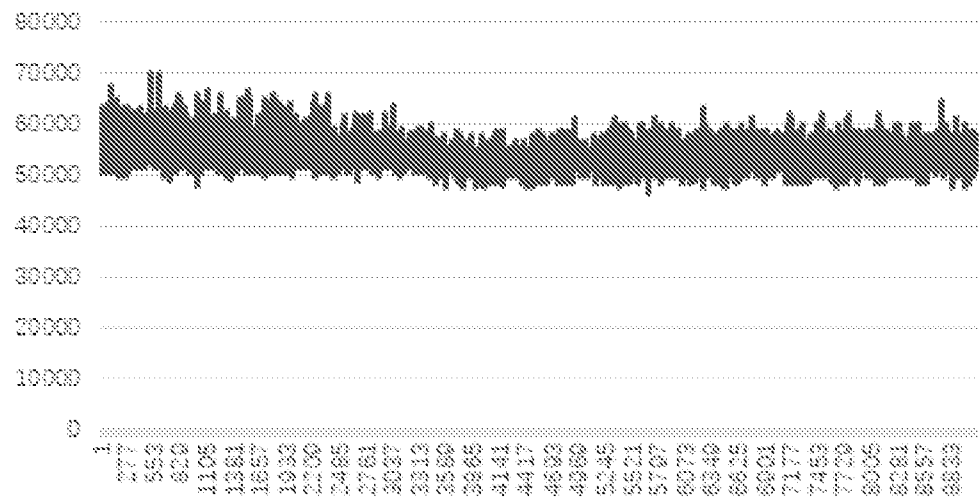


图 9

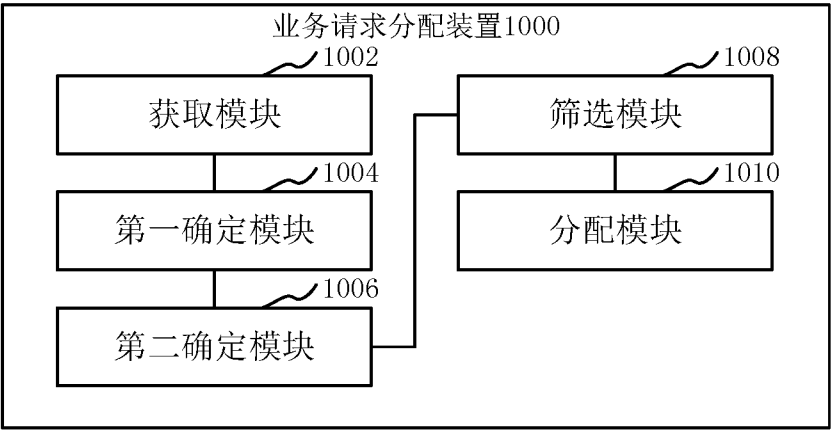


图 10

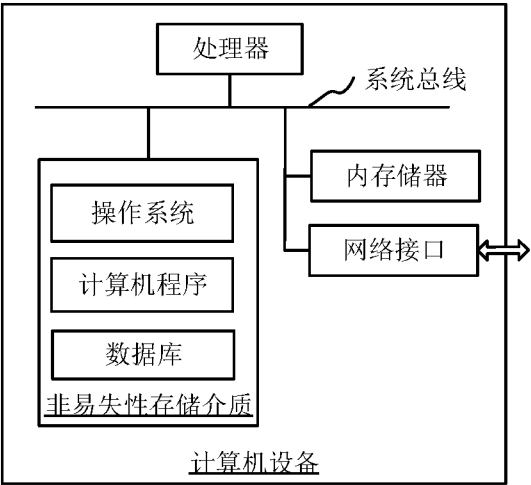


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/114001

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/50(2006.01)i; H04L 29/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, IEEE: 业务, 资源, 分配, 空闲, 划分, 权重, 排序, 最大, 可用, 分配码, 随机, 负载, 均衡, 平衡, business, resource, allocation, idle, partition, weight, rank, max, available, allocation w code, random, load, balance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112131006 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 25 December 2020 (2020-12-25) claims 1-15, description paragraphs [0114]-[0121]	1-20
Y	CN 110858161 A (ALIBABA GROUP HOLDING LIMITED) 03 March 2020 (2020-03-03) description, paragraphs [0043]-[0082]	1-6, 9-20
Y	CN 102611735 A (QIZHI SOFTWARE (BEIJING) CO., LTD.) 25 July 2012 (2012-07-25) description paragraphs [0043]-[0044], [0055]-[0085]	1-6, 9-20
A	US 2017251064 A1 (ALIBABA GROUP HOLDING LIMITED) 31 August 2017 (2017-08-31) entire document	1-20
A	CN 110430278 A (PING AN PUHUI ENTERPRISE MANAGEMENT CO., LTD.) 08 November 2019 (2019-11-08) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

15 October 2021

Date of mailing of the international search report

23 November 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/114001

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112131006	A	25 December 2020	None			
CN	110858161	A	03 March 2020	None			
CN	102611735	A	25 July 2012	None			
US	2017251064	A1	31 August 2017	WO	2017151510	A1	08 September 2017
				CN	107133087	A	05 September 2017
				EP	3423940	A1	09 January 2019
CN	110430278	A	08 November 2019	None			

国际检索报告

国际申请号

PCT/CN2021/114001

A. 主题的分类

G06F 9/50(2006.01)i; H04L 29/08(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, WPI, EPDOC, CNKI, IEEE: 业务, 资源, 分配, 空闲, 划分, 权重, 排序, 最大, 可用, 分配码, 随机, 负载, 均衡, 平衡, business, resource, allocation, idle, partition, weight, rank, max, available, allocation w code, random, load, balance

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 112131006 A (腾讯科技深圳有限公司) 2020年 12月 25日 (2020 - 12 - 25) 权利要求1-15、说明书第[0114]-[0121]段	1-20
Y	CN 110858161 A (阿里巴巴集团控股有限公司) 2020年 3月 3日 (2020 - 03 - 03) 说明书第[0043]-[0082]段	1-6、9-20
Y	CN 102611735 A (奇智软件北京有限公司) 2012年 7月 25日 (2012 - 07 - 25) 说明书第 [0043]-[0044]、[0055]-[0085]段	1-6、9-20
A	US 2017251064 A1 (ALIBABA GROUP HOLDING LIMITED) 2017年 8月 31日 (2017 - 08 - 31) 全文	1-20
A	CN 110430278 A (平安普惠企业管理有限公司) 2019年 11月 8日 (2019 - 11 - 08) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2021年 10月 15日

国际检索报告邮寄日期

2021年 11月 23日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国 北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

李玉坤

电话号码 86-(10)-53961358

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/114001

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	112131006	A	2020年 12月 25日	无			
CN	110858161	A	2020年 3月 3日	无			
CN	102611735	A	2012年 7月 25日	无			
US	2017251064	A1	2017年 8月 31日	W0	2017151510	A1	2017年 9月 8日
				CN	107133087	A	2017年 9月 5日
				EP	3423940	A1	2019年 1月 9日
CN	110430278	A	2019年 11月 8日	无			