

(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
2024年5月30日 (30.05.2024)



(10) 国际公布号
WO 2024/109056 A1

- (51) 国际专利分类号:
H04L 47/10 (2022.01)
- (21) 国际申请号: PCT/CN2023/104095
- (22) 国际申请日: 2023年6月29日 (29.06.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211473367.6 2022年11月21日 (21.11.2022) CN
202310065248.5 2023年1月17日 (17.01.2023) CN
- (71) 申请人: 华为云计算技术有限公司 (HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.) [CN/CN]; 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心, Guizhou 550025 (CN)。
- (72) 发明人: 任海 (REN, Hai); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心, Guizhou 550025 (CN)。张杰 (ZHANG, Jie); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心, Guizhou 550025 (CN)。郑奇峰 (ZHENG, Qifeng); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心, Guizhou 550025 (CN)。
- (74) 代理人: 北京三高永信知识产权代理有限责任公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区上地信息产业基地三街1号楼四层C段457, Beijing 100085 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,

(54) Title: TRAFFIC SCHEDULING METHOD AND APPARATUS, AND DEVICE CLUSTER AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 流量调度方法、装置、设备集群及计算机可读存储介质

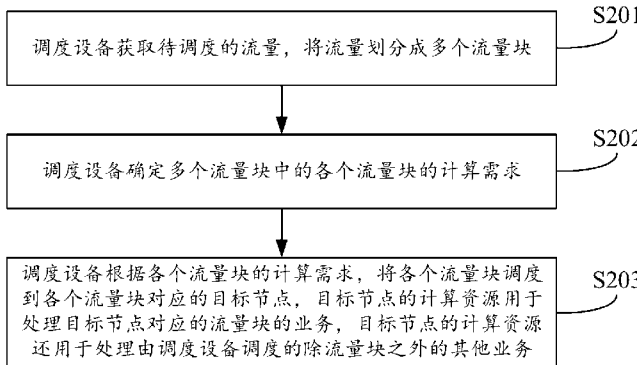


图 2

- S201 A scheduling device acquires traffic to be scheduled, and divides the traffic into a plurality of traffic blocks
- S202 The scheduling device determines a computation requirement of each traffic block among the plurality of traffic blocks
- S203 According to the computation requirement of each traffic block, the scheduling device schedules each traffic block to a target node corresponding to each traffic block, wherein computing resources of the target node are used for processing a service of the traffic block corresponding to the target node, and the computing resources of the target node are further used for processing other services than the traffic block that are scheduled by the scheduling device

(57) Abstract: The present application belongs to the technical field of communications. Disclosed are a traffic scheduling method and apparatus, and a device cluster and a computer-readable storage medium. The method comprises: a scheduling device acquiring traffic to be scheduled, and dividing the traffic into a plurality of traffic blocks; then, determining a computation requirement of each traffic block; and according to the computation requirement of each traffic block, scheduling each traffic block to a target node corresponding to each traffic block, wherein computing resources of the target node are used for processing a service of the traffic block corresponding

GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, 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, ME, 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))。

to the target node, the computing resources of the target node are further used for processing other services than the traffic block, and the other services are also scheduled onto the target node by the scheduling device. One scheduling device is used to schedule traffic or other services to a target node, and one scheduling device allocates computing resources of the target node to a traffic service and other services, thereby improving the allocation rationality of the computing resources, avoiding the waste and overload use of the computing resources, and improving the resource utilization rate.

(57) 摘要: 本申请公开了流量调度方法、装置、设备集群及计算机可读存储介质, 属于通信技术领域。方法包括: 调度设备获取待调度的流量, 并将流量划分成多个流量块; 之后, 确定各个流量块的计算需求; 再根据各个流量块的计算需求, 将各个流量块调度到各个流量块对应的目标节点, 其中, 目标节点的计算资源用于处理目标节点对应的流量块的业务, 目标节点的计算资源还用于处理除流量块之外的其他业务, 其他业务也是由调度设备调度到目标节点上。用一个调度设备调度流量或其他业务到目标节点, 由一个调度设备分配目标节点的计算资源给流量业务和其他业务, 提高了计算资源的分配合理性, 避免了计算资源的浪费和超负荷使用, 提高资源利用率。

流量调度方法、装置、设备集群及计算机可读存储介质

本申请要求于 2022 年 11 月 21 日提交的申请号为 202211473367.6、发明名称为“边缘云资源融合调度方法、装置、设备及可读存储介质”的中国专利申请的优先权，本申请要求于 2023 年 01 月 17 日提交的申请号为 202310065248.5、发明名称为“流量调度方法、装置、设备集群及计算机可读存储介质”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

本申请涉及通信技术领域，特别涉及流量调度方法、装置、设备集群及计算机可读存储介质。

背景技术

计算节点是一种进行业务数据处理的平台。计算节点可以提供带宽、中央处理器（central processing unit, CPU）和内存等计算资源来支持多种类型的业务。其中，计算节点例如是靠近业务数据源的边缘节点，计算节点支持的业务例如是流量业务和计算业务等。其中，流量业务是利用计算节点的计算资源处理调度到该计算节点的流量的业务。计算业务是利用计算节点的计算资源运行调度的容器和虚拟机等计算单元的业务。而在计算节点处理流量业务和计算业务之前，还需要调度流量或计算业务到计算节点上。

发明内容

本申请提供了一种流量调度方法、装置、设备集群及计算机可读存储介质，以将流量调度到计算节点上，技术方案如下：

第一方面，提供了一种流量调度方法，该方法由提供云服务的调度设备执行，该方法包括：调度设备获取待调度的流量，并将流量划分成多个流量块；之后，调度设备确定各个流量块的计算需求；再根据各个流量块的计算需求，将各个流量块调度到各个流量块对应的目标节点，其中，目标节点的计算资源可以用于处理目标节点对应的流量块的业务，目标节点的计算资源还可以用于处理除流量块之外的其他业务，其他业务也是由调度设备调度到目标节点上。

将流量划分成多个流量块，再确定各个流量块的计算需求，从而可以利用一个调度设备完成流量调度和其他业务的调度，调度设备具有通用调度能力。由于调度设备既可以调度流量，又可以调度其他业务，由同一个调度设备分配目标节点的计算资源给流量业务和其他业务，提高了计算资源的分配合理性，避免了计算节点的计算资源的浪费和超负荷使用，提升计算资源的资源利用率。

在一种可能的实现方式中，获取待调度的流量，包括：调度设备获取多个计算节点中的各个计算节点的资源使用情况；当多个计算节点中存在任一计算节点的资源使用情况指示任一计算节点出现异常时，将任一计算节点承载的流量作为待调度的流量，任一计算节点与目标节点为不同的计算节点。将出现异常的任一计算节点上承载的流量调度到其他计算节点上，实现任一计算节点的流量分担，并释放了该任一计算节点上被流量占用的计算资源，避免因为计算资源超负荷导致任一计算节点承载的业务受损。

在一种可能的实现方式中，基于多个计算节点中存在任一计算节点的资源使用情况指示任一计算节点出现异常，将任一计算节点承载的流量作为待调度的流量之前，还包括：基于任一计算节点的资源使用情况指示流量大小超过任一计算节点的第一阈值，或任一计算节点的数据处理单元的使用率超过任一计算节点的第二阈值中的至少一种，确定任一计算节点出现异常。不限定确定任一计算节点出现异常的确定方式，灵活度高。

在一种可能的实现方式中，将流量划分成多个流量块，包括：根据流量对应的解析协议确定参考数量；将位于同一区域内的流量划分成参考数量个流量块。根据解析协议确定参考数量，使得流量的划分过程更精准。

在一种可能的实现方式中，确定多个流量块中的各个流量块的计算需求，包括：获取各个流量块的节点区域要求与计算资源需求，将节点区域要求与计算资源需求作为计算需求。确定的计算需求不光限定了流量块所需要的计算资源，还限定了流量块需要调度到的节点的区域，得到的计算需求精确度高。

在一种可能的实现方式中,获取各个流量块的节点区域要求,包括:构建各个流量块对应的区域标签,其中,流量块对应的区域标签用于指示流量块对应的计算节点位于的区域;将各个流量块与各个流量块对应的区域标签之间的亲和关系作为各个流量块的节点区域要求,亲和关系用于确定流量块需要调度到的区域。通过区域标签和亲和关系即可描述流量块的节点区域要求,节点区域要求简单,基于简单的节点区域要求选择目标节点时的操作简单,更便捷。

在一种可能的实现方式中,区域标签包括地区标签和运营商标签中的至少一种。本申请不限定区域标签的类型和数量,通用度高。

在一种可能的实现方式中,计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。其中,内存处理单元例如是CPU。计算资源需求中可以包括多种数据,灵活度高。

在一种可能的实现方式中,当各个流量块的计算需求包括节点区域要求和计算资源需求时,根据各个流量块的计算需求,将各个流量块调度到各个流量块对应的目标节点,包括:对于任一流量块,根据任一流量块的计算需求筛选可调度的计算节点,得到筛选后的计算节点,筛选后的计算节点的资源使用情况满足任一流量块的计算资源需求,且筛选后的计算节点的区域信息满足任一流量块的节点区域要求;从筛选后的计算节点中选择计算节点作为任一流量块对应的目标节点;将任一流量块调度到任一流量块对应的目标节点。从筛选后的计算节点中选择的目标节点可以满足任一流量块的计算资源需求和节点区域要求,保证了任一流量块调度到目标节点后,该任一流量块的业务可以在目标节点顺利处理。

在一种可能的实现方式中,其他业务包括计算业务和存储业务中的至少一种。不限定调度设备调度的其他业务,通用度高,适用范围广。

在一种可能的实现方式中,目标节点属于边缘节点。用于处理流量业务的目标节点为靠近业务数据源的边缘节点,拉近了目标节点和业务数据源的距离,缩短了业务数据源上的流量到目标节点之间的传输距离,降低了流量的传输时间,有效提高目标节点对流量的响应速度和连接速度。

第二方面,提供了一种流量调度装置,该装置应用于提供云服务的调度设备,装置包括:划分模块,用于获取待调度的流量,将流量划分成多个流量块;确定模块,用于确定多个流量块中的各个流量块的计算需求;调度模块,用于根据各个流量块的计算需求,将各个流量块调度到各个流量块对应的目标节点,目标节点的计算资源用于处理目标节点对应的流量块的业务,目标节点的计算资源还用于处理由调度设备调度的除流量块之外的其他业务。

在一种可能的实现方式中,划分模块,用于获取多个计算节点中的各个计算节点的资源使用情况;基于多个计算节点中存在任一计算节点的资源使用情况指示任一计算节点出现异常,将任一计算节点承载的流量作为待调度的流量,任一计算节点与目标节点为不同的计算节点。

在一种可能的实现方式中,确定模块,还用于基于任一计算节点的资源使用情况指示流量大小超过任一计算节点的第一阈值,或任一计算节点的数据处理单元的使用率超过任一计算节点的第二阈值中的至少一种,确定任一计算节点出现异常。

在一种可能的实现方式中,划分模块,用于根据流量对应的解析协议确定参考数量;将位于同一区域内的流量划分为参考数量个流量块。

在一种可能的实现方式中,确定模块,用于获取各个流量块的节点区域要求与计算资源需求,将节点区域要求与计算资源需求作为计算需求。

在一种可能的实现方式中,确定模块,用于构建各个流量块对应的区域标签,流量块对应的区域标签用于指示流量块对应的计算节点位于的区域;将各个流量块与各个流量块对应的区域标签之间的亲和关系作为各个流量块的节点区域要求,亲和关系用于确定流量块需要调度到的区域。

在一种可能的实现方式中,区域标签包括地区标签和运营商标签中的至少一种。

在一种可能的实现方式中,计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。

在一种可能的实现方式中,各个流量块的计算需求包括节点区域要求和计算资源需求,调度模块,用于对于任一流量块,根据任一流量块的计算需求筛选可调度的计算节点,得到筛选后的计算节点,筛选后的计算节点的资源使用情况满足任一流量块的计算资源需求,且筛选后的计算节点的区域信息满足任一流量块的节点区域要求;从筛选后的计算节点中选择计算节点作为任一流量块对应的目标节点;将任一流量块调度到任一流量块对应的目标节点。

在一种可能的实现方式中，其他业务包括计算业务和存储业务中的至少一种。

在一种可能的实现方式中，目标节点属于边缘节点。

第三方面，提供了一种计算设备集群，该计算设备集群包括至少一个计算设备，每个计算设备包括处理器和存储器；至少一个计算设备的处理器用于执行至少一个计算设备的存储器中存储的指令，以使得计算设备集群执行上述第一方面中任一种的流量调度方法。

第四方面，提供了一种计算机可读存储介质，该计算机可读存储介质包括计算机程序指令，当计算机程序指令由计算设备集群执行时，计算设备集群执行上述第一方面中任一种的流量调度方法。

第五方面，提供了一种包含指令的计算机程序（产品），当该指令被计算设备集群运行时，使得计算设备集群执行上述第一方面任一种的流量调度方法。

第六方面，提供了一种通信装置，该装置包括：收发器、存储器和处理器。其中，该收发器、该存储器和该处理器通过内部连接通路互相通信，该存储器用于存储指令，该处理器用于执行该存储器存储的指令，以控制收发器接收信号，并控制收发器发送信号，并且当该处理器执行该存储器存储的指令时，使得该处理器执行第一方面或第一方面的任一种可能的实施方式中的方法。

可选地，处理器为一个或多个，存储器为一个或多个。

可选地，存储器可以与处理器集成在一起，或者存储器与处理器分离设置。

在具体实现过程中，存储器可以为非瞬态性（non-transitory）存储器，例如只读存储器（read only memory, ROM），其可以与处理器集成在同一块芯片上，也可以分别设置在不同的芯片上，本申请对存储器的类型以及存储器与处理器的设置方式不做限定。

第七方面，提供了一种芯片，包括处理器，用于从存储器中调用并运行存储器中存储的指令，使得安装有该芯片的通信设备执行上述各方面中的方法。

第八方面，提供另一种芯片，包括：输入接口、输出接口、处理器和存储器，输入接口、输出接口、处理器以及存储器之间通过内部连接通路相连，处理器用于执行存储器中的代码，当代码被执行时，处理器用于执行上述各方面中的方法。

应当理解的是，本申请的第二方面至第八方面的技术方案及对应的可能的实现方式所取得的有益效果可以参见上述对第一方面及其对应的可能的实现方式的技术效果，此处不再赘述。

附图说明

图1为本申请实施例提供了一种实施环境的示意图；

图2为本申请实施例提供了一种流量调度方法的流程图；

图3为本申请实施例提供的另一种流量调度方法的流程图；

图4为本申请实施例提供了一种流量调度装置的结构示意图；

图5为本申请实施例提供了一种计算设备的结构示意图；

图6为本申请实施例提供了一种计算设备的连接示意图。

具体实施方式

本申请的实施方式部分使用的术语仅用于对本申请的具体实施例进行解释，而非旨在限定本申请。为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

计算节点是一种利用带宽、CPU和内存等计算资源来进行业务数据处理的平台。随着通信技术的发展，提供云服务的计算节点的构建也日趋成熟，越来越多的业务基于计算节点实现服务，计算节点上承载的业务也开始向多元化演进。其中，计算节点例如是内容分发网络（content delivery network, CDN）中存在的互联网数据中心（internet data center, IDC）节点。当计算节点配置在靠近业务数据源时，该计算节点也可以被称为边缘节点。

在一种可能的情况下，一个计算节点不光可以提供处理流量的业务所需要的计算资源，以实现流量业务的承载，还可以提供处理计算业务所需要的计算资源。由此可知，同一个计算节点的计算资源需要被分配给不同业务，在多元业务的场景下，可能存在计算资源冲突和争抢，进而导致该计算节点的线上业务质量受损。本申请实施例提供了一种流量调度方法，用于将流量合理调度到计算节点上，以实现计算节点的计算资源的合理分配。

请参考图 1，其示出了本申请实施例提供的流量调度方法的实施环境示意图。该实施环境为计算设备集群 10，计算设备集群 10 包括至少一台用于提供云服务的调度设备，调度设备之间可通过有线或无线网络的方式进行通信连接。可选地，流量调度方法可以由一台调度设备独立执行，也可由计算设备集群 10 中包括的多台调度设备交互执行，本申请实施例不限定计算设备集群 10 包括的调度设备的数量，图 1 中仅以两台调度设备为例进行举例说明。此外，图 1 中的计算设备集群 10 除了可以用于执行本申请实施例提供的流量调度方法，将流量块调度到流量块对应的目标节点上，还可以将其他业务调度到其他业务对应的目标节点上。也即计算设备集群 10 执行的是一种调度流量业务和其他业务的融合调度。

示例性地，计算设备集群 10 包括的调度设备可以是服务器，例如是中心服务器、边缘服务器，或者是本地数据中心中的本地服务器。服务器可以是物理服务器，还可以是提供云计算服务的云服务器。在一些实施例中，计算设备集群 10 包括的调度设备也可以是云服务中涉及到的台式机、笔记本电脑或者智能手机等终端设备。

本申请实施例提供一种流量调度方法，该流量调度方法可应用于上述图 1 所示的实施环境，该方法可由计算设备集群 10 中包括的调度设备执行，该方法的流程图如图 2 所示，包括 S201-S203。

S201，调度设备获取待调度的流量，将流量划分成多个流量块。

示例性地，待调度的流量根据访问请求产生，且待调度的流量需要借助计算节点进行处理。例如，来自某个区域的用户请求访问网站 A，调度设备根据用户的访问请求接收到流量。

上述为对待调度的流量的一个举例说明，调度设备可以接收根据用户访问请求产生的流量，将根据用户访问请求产生的流量作为待调度的流量，也可以将被分配到其他计算节点的流量作为待调度的流量。可选地，调度设备获取多个计算节点中的各个计算节点的资源使用情况；基于多个计算节点中存在任一计算节点的资源使用情况指示任一计算节点出现异常，将任一计算节点承载的流量作为待调度的流量，任一计算节点与目标节点为不同的计算节点。

其中，资源使用情况用于说明该计算节点上的计算资源的使用量，通过计算资源的使用量，调度设备可以确定该计算节点是否出现异常。基于一个计算节点提供的计算资源包括带宽、内存和数据处理单元等多种，调度设备也可以通过包括但不限于如下两种方式根据资源使用情况确定计算节点是否异常。

确定方式一、基于任一计算节点的资源使用情况指示流量大小超过任一计算节点的第一阈值，确定任一计算节点出现异常。

由于单个计算节点的带宽有限，而带宽又决定了该计算节点可以流通的流量最大值。因此，一个计算节点能承载的流量也是有限的，需要对计算节点承载的流量进行控制。可选地，任一计算节点与调度设备之间建立通信连接，调度设备可基于通信连接接收由任一计算节点发送的任一计算节点上承载的流量大小。调度设备统计任一计算节点承载的流量大小，将流量大小与任一计算节点的第一阈值进行比较，并根据比较结果确定该任一计算节点上承载的流量是否超过负荷能力。在流量大小大于第一阈值，也即任一计算节点承载的流量超过负荷能力时，确定该任一计算节点出现异常。

其中，第一阈值为基于经验设置的任意大小的数值，第一阈值的单位与流量的单位保持一致。在一种可能的情况下，调度设备在设置第一阈值时还会参考该计算节点可以流通的流量最大值，例如将流量最大值作为第一阈值，以保证计算节点的计算资源的充分利用。或者，将流量最大值乘以 90% 的计算结果作为第一阈值，在流量未达到计算节点的承载上限时就将流量调度到其他计算节点上。以保证调度设备在调度流量期间，该计算节点可以正常工作，从而避免该计算节点的流量业务受损。此外，不同的计算节点对应的第一阈值可以相同，也可以不同。不同可以是部分不同，或者是完全不同。

确定方式二、基于任一计算节点的资源使用情况指示数据处理单元的使用率超过任一计算节点的第二阈值，确定任一计算节点出现异常。

其中，数据处理单元可以是 CPU，还可以是其他支持数据处理和程序运行的单元。与第一阈值类似，第二阈值也为基于经验设置的任意大小的数值。例如，设置第二阈值为 80% 或者 90%。同样地，不同的计算节点对应的第二阈值可以相同，也可以不同，本申请实施例对此不进行限定。调度设备接收任一计算节点上传的数据处理单元的使用率，由此获得任一计算节点的资源使用情况，并将数据处理单元的使用率与第二阈值进行大小比较，进而确定任一计算节点是否出现异常。

调度设备可以选择上述任一种方法确定任一计算节点是否出现异常，调度设备还可以选择结合上述两

种方法进行异常的判断。也就是说，对于一个计算节点提供多种计算资源的情况，每一种计算资源都会存在对应的阈值。调度设备比较资源使用情况包括的各个计算资源的使用量与各个计算资源对应阈值的大小，确定任一计算节点是否出现异常。可选地，调度设备包括业务感知模块，用于执行上述获取资源使用情况，根据资源使用情况确定任一计算节点是否出现异常的操作。

除了上述举例的根据计算资源的使用量进行异常情况的判断，调度设备还可以统计计算资源的剩余量进行异常情况的判断。例如，统计数据处理单元的剩余率，基于剩余率小于或等于根据经验设置的参考阈值，确定任一计算节点出现异常。并且，用于确定是否出现异常的计算资源除了上述实施例示出的数据处理单元和流量，还可以是包括带宽在内的其他计算资源。则确定任一计算节点出现异常也可以是任一计算节点的带宽使用率超过基于经验设置的第三阈值。

可以理解，根据资源使用情况确定任一计算节点出现异常，是指该任一计算节点承载的业务超过负荷能力。因此，调度设备需要将出现异常的任一计算节点承载的流量及时调度到其他节点上，实现流量业务的分摊，计算资源的释放，进而避免由于计算节点的工作负荷导致该任一计算节点的承载的业务受损，保证任一计算节点提供高质量业务。

此外，虽然任一计算节点上会承载多种业务，例如任一计算节点的计算资源除了用于支持流量业务中涉及的流量解析，还会用于支持计算业务涉及到的计算单元的运行。但调度设备在释放该任一计算节点的计算资源时，还是会选择调度任一计算节点承载的流量到其他计算节点上。其原因是由于计算业务对计算资源的需求相对流量业务对计算业务的需求是固定的。以计算业务的计算单元为容器为例，在给每个计算节点分配容器时，容器对计算资源的占用已经确定，属于有限可控的，比如明确知道一个容器需要占用 10 个 CPU。而流量业务对计算资源的需求会随着时间存在变化。例如，在 20:00-22:00 晚高峰，观看视频和浏览网站的人数会增多，由此产生的流量也会增多，单个计算节点需要承载的流量增多，从而造成计算资源的负荷。基于流量为导致计算节点出现异常的主要原因，调整计算节点上承载的流量可以有效解决计算资源的负荷。

基于上述原理，调度设备会在检测出任一计算节点出现异常后，将任一计算节点承载的流量提取出来，作为待调度的流量。之后开始对待调度的流量进行标准化处理，使得处理后的流量属于和计算单元一样的标准化业务对象，进而可以利用同一个调度设备完成流量调度和包括计算业务在内的其他业务的调度。其中，标准化业务对象是指可以采用统一的描述单位对该业务对象进行描述。例如标准化后的流量可以采用统一描述单位描述计算需求，标准化后的计算单元可以采用统一描述单位描述该计算单元的计算能力。

以上为调度设备需要对流量进行标准化处理的原因，接下来开始介绍调度设备对流量进行标准化处理的过程。示例性地，调度设备根据流量对应的解析协议确定参考数量；将位于同一区域内的流量划分为参考数量个流量块。在一种可能的实现方式中，调度设备调度流量到计算节点是基于解析协议进行。调度设备根据该流量中携带的访问请求，利用解析协议确定访问请求对应的网际互联网协议（internet protocol, IP）地址，联合各个计算节点的资源使用情况等多种因素确定流量需要调度到的计算节点。而上述用于解析流量的解析协议也就是该待调度的流量对应的解析协议，解析协议例如是域名系统（domain name system, DNS）协议，也可以是其他协议。

在利用解析协议确定访问情况对应的 IP 地址时，由于运营商本地（Local）DNS 的限制，每个 DNS 包内最多可以包括 N 个 IP 地址，N 为正整数。此种情况下，根据解析协议确定参考数量，也就是将解析协议对应的 N 确定为参考数量。由于 N 反应了最多可以将流量划分成多少个流量块，将 N 作为参考数量，调度设备再将待调度的流量划分为参考数量个流量块，实现待调度的流量的最细粒度的划分，有效控制了流量块的大小，以提高后续调度流量块的调度效率。示例性地，划分过程可以是均分，例如，同一区域内的流量大小为 200 吉字节（gigabyte, G），N 为 10，那么得到的 10 个流量块的大小均为 20G。

在一种可能的情况下，被划分的流量属于同一区域，且该流量请求访问的域名也相同。流量属于的区域根据用户在请求访问时的区域决定。区域可以是根据省份和运营商中的至少一种确定，例如来自某地区的用户采用某运营商进行请求访问，则请求访问产生的流量位于的区域为某地区+某运营商。

地区可以根据省份划分，例如河北属于一个地区，安徽属于另一个地区，则河北+运营商 A 属于一个区域，安徽+运营商 A 属于另一个区域。地区也可以是根据城市、地区等地域划分单位和确定。例如，朝阳区属于一个地区，海淀区属于另一个地区，则朝阳区+运营商 A 属于一个区域，海淀区+运营商 A 属于另一个区域。当然，区域可以指国内的区域，也可以指国外的区域，国外区域的划分与国内区域的划分类

似，可参考相关描述，暂不再赘述。还需要注意的是，上述通过地区和运营商确定区域旨在举例说明区域的概念，而非用于限定区域的划分方式，流量位于的区域还可以有其他更多的划分方式。

由于流量本身具备粘滞性，也即某个区域中请求访问某个域名的流量不会在短时间内大量突增。而切分流量时的参考数量也是恒定的，所以可以认为在参考时间内划分某个区域中请求某个域名的流量得到的流量块的大小是不变的，或者说流量块的大小变化幅度不会超过基于经验设置的误差阈值。其中，参考时间为根据经验设置的时间，如参考时间为5分钟，应理解，本申请采用参考时间进行描述，旨在说明单个流量块在短时间内大小是稳定的，而非用于限定单个流量块的稳定时间，在一些情况下，单个流量块的稳定时间可以为更长或更短的时间。

由于流量块的大小在参考时间内稳定，计算节点处理该流量块所需要的计算资源也是稳定的。也就是说，将流量切分为块后，得到的流量块可以类比为网荚（pod），虚拟机（virtual machine, VM）等计算单元。对流量块的调度，在场景上就可以和计算业务调度等归为一类，本申请可以采用同一个调度设备执行流量块的调度和计算业务等其他业务的调度。

S202，调度设备确定多个流量块中的各个流量块的计算需求。

调度设备确定流量块的计算需求的过程包括但不限于：获取各个流量块的节点区域要求与计算资源需求，将节点区域要求与计算资源需求作为计算需求。在一种可能的实现方式中，基于网络链路限制或者流量传输限制，各个流量块会对计算节点存在一个节点区域要求。而调度设备确定节点区域要求的过程包括但不限于：构建各个流量块对应的区域标签，流量块对应的区域标签用于指示流量块对应的计算节点位于的区域；将各个流量块与各个流量块对应的区域标签之间的亲和关系作为各个流量块的节点区域要求，亲和关系用于确定流量块需要调度到的区域。

在一种可能的实现方式中，调度设备根据网络链路限制或流量传输限制中的至少一种，构建流量块对应的区域标签。针对区域是根据地区和运营商中的至少一种确定的情况，区域标签也包括地区标签和运营商标签中的至少一种。以流量块A为例，构建的区域标签为：地区标签：北京；运营商标签：运营商A。当然，区域标签还可以包括节点名，例如，除了上述两个举例，流量块A构建的区域标签还包括：节点名：计算节点A。

调度设备为流量块构建的区域标签，可能是用于指示流量块需要调度到的区域，也可以是用于指示流量块不能调度到的区域。也就是说，流量块和区域标签之间是具有一种亲和关系。当流量块与该区域标签之间的亲和关系为亲和性时，则指示该流量块期望调度到此种区域标签指示的区域。继续以上述流量块A对应的区域标签为例，在流量调度过程中，业务都会有本省覆盖的要求，即本省的流量要调度到本省的节点上去，且不能跨运营商覆盖。这种本省覆盖的调度要求可以通过亲和性表述，得到的表达结果为，亲和性：地区标签==北京；亲和性：运营商标签==运营商A。

而当流量块与该区域标签之间的亲和关系为反亲和性时，则指示该流量块不期望调度到该区域标签指示的区域。仍以上述流量块A对应的区域标签为例，计算节点A为原本承载流量块A的计算节点，由于调度设备是需要将流量块A调度到除计算节点A之外的目标节点，因此，流量块A不期望调度到计算节点A。此种调度需求通过亲和性表述为：反亲和性：节点名!=计算节点A。当然，节点区域要求中包括的节点名，还可以是指流量块期望调度到的计算节点，也即节点名和流量块之间的亲和关系除了上述举例的反亲和性，也可以为亲和性。

之后，调度设备可以将上述通过亲和关系表述的结果作为流量块的节点区域要求。也即，调度设备获取的流量块A的节点区域要求为：亲和性：地区标签==北京；亲和性：运营商标签==运营商A；反亲和性：节点名!=计算节点A。

由于调度流量块到计算节点除了需要考虑该流量块对区域的限制，还需要考虑处理该流量块所需要的计算资源。因此，调度设备还需要统计该流量块的计算资源需求。与计算资源包括带宽、数据处理单元和内存中的至少一种相对应，计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。仍以流量块为流量块A，数据处理单元为CPU为例，流量块A的计算资源需求为：带宽：10G；CPU：1核。调度设备由此得到包括计算资源需求和节点区域要求的计算需求。

通过上述操作，调度设备可以采用统一的描述单位去描述不同的流量块的计算需求，实现了流量块的标准化处理。流量块的描述单位例如是上述举例的地区标签、运营商标签、带宽、CPU和流量块大小等。

S203，调度设备根据各个流量块的计算需求，将各个流量块调度到各个流量块对应的目标节点，目标

节点的计算资源用于处理目标节点对应的流量块的业务，目标节点的计算资源还用于处理由调度设备调度的除流量块之外的其他业务。

在一种可能的情况下，调度设备统计的各个计算节点的资源使用情况，还会用于判断该计算节点是否可以处理流量块的业务。并且调度设备在确定流量块对应的目标节点时，除了根据各个计算节点的资源使用情况，还会获取各个计算节点的区域信息，也即该计算节点位于哪个地区，属于哪个运营商，再根据各个计算节点的区域信息进行目标节点的确定。

由于根据计算节点的区域信息和资源使用情况，调度不同的流量块到对应的目标节点的过程类似，接下来暂以调度任一流量块到任一流量块对应的目标节点为例进行举例说明，关于其他流量块的调度过程，可参考相关描述，此处不再赘述。

示例性地，调度设备根据任一流量块的计算需求筛选可调度的计算节点，得到筛选后的计算节点，筛选后的计算节点的资源使用情况满足任一流量块的计算资源需求，且筛选后的计算节点的区域信息满足任一流量块的节点区域要求；从筛选后的计算节点中选择计算节点作为任一流量块对应的目标节点；将任一流量块调度到任一流量块对应的目标节点。

其中，可调度的计算节点是指可以正常进行业务处理的计算节点，可调度的计算节点上的计算资源运行正常。例如，计算节点 B 上的 CPU 运行异常，则计算节点 B 就不属于可调度的计算节点。

调度设备比较任一流量块的节点区域要求和任一计算节点的区域信息，当该任一计算节点位于的地区属于节点区域要求中期望调度到的地区，且任一计算节点属于的运营商，属于节点区域要求中期望调度到的运营商，确定任一计算节点的区域信息满足任一流量块的节点区域要求。当然，如果该流量块的节点区域要求中还明确该流量块不期望调度到的计算节点，调度设备还会在比较任一计算节点和流量块的地区和运营商的基础上，进一步比较任一计算节点是否是不期望调度到的计算节点，并在该任一计算节点不是不期望调度到的计算节点时，确定任一计算节点的区域信息满足任一流量块的节点区域要求。

在一种可能的实现方式中，调度设备也会对计算节点进行标准化处理。通过构建计算节点对应的区域标签，明确该计算节点位于的地区或属于的运营商中的至少一种。将计算节点对应的区域标签作为区域信息，实现采用统一描述单位对计算节点所属区域进行描述。此种情况下，上述节点区域要求和区域信息的比较可以通过匹配流量块对应的区域标签和计算节点对应的区域标签实现。

此外，调度设备还会根据资源使用情况确定计算节点的计算资源剩余量，也即该计算节点能提供的计算资源的数量。由于任一流量块的计算资源需求指示了任一流量块需要的计算资源的数量。通过比较任一流量块的计算资源需求和任一计算节点的资源使用情况，当资源使用情况中指示的该计算节点能提供的计算资源的数量，不小于任一流量块的计算资源需求中指示的所需要的计算资源的数量，则确定任一计算节点的资源使用情况满足任一流量块的计算资源需求。

通过计算资源和区域信息这两个维度进行计算节点的筛选，保证筛选后的计算节点符合任一流量块的计算需求。之后，调度设备会从筛选后的计算节点中选择计算节点作为任一流量块对应的目标节点。针对筛选后的计算节点的数量为一个的情况，调度设备直接将筛选后的计算节点作为该任一流量块对应的目标节点。针对筛选后的计算节点的数量为多个的情况，调度设备可以随机选择一个计算节点作为任一流量块对应的目标节点，也可以根据各个计算节点的资源使用情况，将计算资源使用最少的计算节点确定为任一流量块对应的目标节点。当然，调度设备还可以基于其他原则进行目标节点的选择，本申请实施例不进行限定。

无论调度设备基于何种方式确定任一流量块对应的目标节点，均可以将任一流量块调度到任一流量块对应的目标节点。可选地，调度设备将任一流量块调度到对应的目标节点的过程包括：根据流量块对应的目标节点确定该流量块的调度策略；调度设备向下游设备发送流量块的调度策略，下游设备用于根据调度策略实现流量块的调度。其中，流量块的调度策略用于描述该流量块会被调度到的目标节点是什么。例如流量块 A 的调度策略为流量块 A 调度到计算节点 C。下游设备例如是 DNS。

此外，将流量调度到目标节点是指用户后续请求访问时，可以直接向目标节点请求访问，根据访问请求产生的流量直接发往目标节点，而不用再发往调度设备。并且，在将用户的访问请求作为流量的情况下，调度设备下发到 DNS 的调度策略会被转义为：域名+区域+运营商的形式，且该调度策略还会请求解析到目标节点的 IP 地址。此种情况下，流量调度过程为 DNS 负责解析调度策略，将来自不同区域不同运营商的用户的访问请求解析到调度策略指定的目标节点上。

基于此，用户后续可以通过目标节点进行访问请求的处理。目标节点接收该用户的访问请求，再利用CPU进行访问请求的处理，根据用户的访问请求返回相关数据。例如，目标节点解析访问请求，基于解析确定用户请求的网站为网站A。由于目标节点中缓存了网站A的网站界面，目标节点可以直接将网站A的网站界面返回给用户，用户由此实现对网站A的访问。其中，返回给用户的相关数据除了上述举例的网站界面，还可以是用户请求的图片数据、音乐数据或者视频数据等。上述为目标节点调用计算资源处理流量的业务的举例，基于一个计算节点可以支持多种业务，目标节点的计算资源除了用于处理上述流量块的业务，还可以处理除流量块之外的其他业务。可选地，其他业务包括计算业务和存储业务中的至少一种。

示例性地，目标节点处理的其他业务也是由调度设备调度到目标节点上。也就是说，调度设备还会接收其他业务，根据其他业务的调度需求确定其他业务对应的目标节点，将其他业务调度到其他业务对应的目标节点上。

在其他业务存在调度需求的情况下，会通过其他业务的请求设备向调度设备发送待调度的其他业务。以待调度的其他业务为计算业务为例，则请求设备向调度设备提交的其他业务就是待调度的计算单元的信息。在一种可能的实现方式中，其他业务的调度请求是基于业务触发。例如，需要启用新的计算业务，该计算业务需要10个容器，每个容器需要2核CPU，500兆的内存。请求设备将上述待调度的计算单元的信息提交给调度设备，调度设备由此接收到待调度的计算业务。或者，基于计算节点上的计算单元的数量不满足计算需求，还需要在计算节点上配置新的计算单元，每个计算单元需要占用200兆的内存，请求设备也会向调度设备发送需要配置的计算单元的信息，调度设备由此接收到待调度的计算业务。

在一种可能的实现方式中，请求设备发送给调度设备其他业务也是经过标准化处理的，也即用统一的描述单位描述了计算业务的计算能力或存储业务的存储能力。以计算业务的计算单元为例，对标准化处理计算业务进行解释说明。由于不同型号的计算单元的计算能力存在区别，可能有的计算单元的计算能力强，有的计算单元的计算能力弱，所以需要采用统一的描述单位去描述计算单元。此外，对于同样型号的计算单元的计算能力也可能存在区别，这些区别可能是使用时长不同导致。例如，计算单元的CPU在使用过程中，会因为年限产生损耗，所以同一型号的新旧计算单元，能承载的计算能力不同。基于此种情况，也需要利用描述单位，描述计算单元的不同计算能力，实现对计算单元进行标准化处理。

无论调度设备接收到的其他业务为何种情况，调度设备均可在接收到待调度的其他业务后，根据其他业务对计算资源的需求，确定其他业务对应的目标节点，进而生成用于指示其他业务对应的目标节点的调度策略。其中，确定其他业务对应的目标节点的过程与上述实施例示出的确定流量块对应的目标节点的过程类似，可参考相关描述，此处不再赘述。之后，调度设备将其他业务的调度策略下发至边缘云平台，启动其他业务的部署流程，以完成将其他业务调度至其他业务对应的目标节点。本申请实施例不限定其他业务的部署方式，例如，计算业务中的计算单元可以采用一个开源的容器编排引擎(kubernetes, K8s)实现自动化部署。

综上所述，本申请实施例提供的流量调度方法，将流量划分成在参考时间内大小稳定的流量块，实现通过一个调度设备完成流量调度和其他业务的调度，调度设备具有通用调度能力。由于调度设备既可以调度流量，又可以调度其他业务，由同一个调度设备分配目标节点的计算资源给流量业务和其他业务，调度设备可全面掌握各类业务使用计算资源的情况，提高了计算资源的分配合理性，避免了计算节点的计算资源的浪费和超负荷使用，提升计算资源的资源利用率，降低了资源成本。

并且，由一个调度系统进行多业务调度，可极大的提升不同业务之间的计算资源复用率，也能及时的处理单个业务的突发，避免多业务之间的计算资源的冲突，提升了计算节点承载的业务质量。本申请的流量调度方法可以让计算节点在同样的计算资源下承载更多的业务，有效降低计算节点承载业务的成本。

图3为本申请实施例提供的一种流量调度方法的流程图，以调度设备包括业务感知模块、标准化模块和核心调度器为例，示出了计算节点、调度设备和请求设备之间的交互过程。其中，请求设备用于处理其他业务的调度请求，其他业务例如是计算业务。

S301，计算节点向业务感知模块上报资源使用情况。

可选地，计算节点和调度设备之间可通过有线或无线网络的方式建立通信连接，各个计算节点基于通信连接向调度设备包括的业务感知模块发送资源使用情况。

S302，业务感知模块根据资源使用情况触发调度事件。

在资源使用情况指示不存在计算节点出现异常的情况下，调度设备重复 S301 的操作，通过接收计算节点上报的资源使用情况，进行各个计算节点的异常检测。示例性地，在资源使用情况指示任一计算节点出现异常的情况下，业务感知模块由此触发调度事件。触发调度事件是指开始调度任一计算节点上承载的流量。在一种可能的实现方式中，根据资源使用情况触发调度事件的过程与上述图 2 所示的实施例 S201 中示出的根据资源使用情况获取待调度的流量的过程类似，可参考相关描述，此处不再赘述。

S303，在调度事件触发后，业务感知模块向标准化模块发送待调度的流量。

示例性地，业务感知模块将在 S302 中检测出的出现异常的任一计算节点上承载的流量确定为待调度的流量，并将该任一计算节点上承载的流量提取出来，向标准化模块发送。

S304，标准化模块标准化待调度的流量。

对待调度的流量的标准化处理包括但不限于将流量切分成参考时间内大小稳定的流量块，以及确定流量块的计算需求。在一种可能的情况下，标准化模块标准化待调度的流量的过程与上述图 2 所示的实施例 S201-S202 中示出的切分待调度的流量和确定流量块的计算需求的过程类似，可参考相关描述，此处不再赘述。通过对流量进行标准化处理，使得流量业务转换为云原生平台的 K8s 可识别的通用业务模型，进而可以通过同一个调度设备调度流量和其他业务。

S305，标准化模块向核心调度器提交标准化流量块，请求设备向核心调度器提交其他业务。

示例性地，标准化流量块是指 S304 中进行标准化处理后的待调度的流量，也就是确定了计算需求的流量块。由于调度设备不光可以进行流量调度，还可以实现计算单元等其他业务的调度。因此，核心调度器还会接收请求设备发送的待调度的其他业务，以将其他业务调度到计算节点上。关于请求设备向核心调度器提交其他业务的过程与上述图 2 所示的实施例 S203 中请求设备向调度设备提交其他业务的过程类似，可参考相关描述，此处不再赘述。

S306，核心调度器调度逻辑执行。

示例性地，核心调度器调度逻辑执行包括核心调度器调度标准化流量块和调度其他业务。可选地，核心调度器调度标准化流量块的过程与上述图 2 所示的 S203 中调度设备调度流量块的过程类似，包括确定流量块对应的目标节点，和调度流量块到对应的目标节点。核心调度器调度其他业务的过程与上述图 2 所示的 S203 中调度设备调度其他业务的过程类似，包括确定其他业务对应的目标节点和调度其他业务到对应的目标节点的过程，均可参考相关描述，此处不再赘述。

此外，针对核心调度器需要确定流量块对应的目标节点和其他业务对应的目标节点的情况，核心调度器可以选择同步执行目标节点的确定，也可以选择异步执行。异步执行的过程可以先完成流量块对应的目标节点的确定，再进行其他业务对应的目标节点的确定。异步执行的过程还可以是先完成其他业务对应的目标节点的确定，再进行流量块对应的目标节点的确定，或者是穿插执行流量块对应的目标节点的确定和其他业务对应的目标节点的确定。

S307，核心调度器分别向计算节点和请求设备反馈调度策略。

示例性地，调度策略是指核心调度器确定的流量块对应的目标节点或者确定的其他业务对应的目标节点。核心调度器向计算节点反馈调度策略的过程与上述图 2 所示的实施例 S203 中将流量调度到目标节点的过程类似，可参考相关描述，此处不再赘述。

以上介绍了本申请实施例的流量调度方法，与上述方法对应，本申请实施例还提供了流量调度装置。图 4 是本申请实施例提供的一种流量调度装置的结构示意图。基于图 4 所示的如下多个模块，该图 4 所示的流量调度装置能够执行上述图 2 所示的全部或部分操作。应理解到，该装置可以包括比所示模块更多的附加模块或者省略其中所示的一部分模块，本申请实施例对此并不进行限制。如图 4 所示，该装置应用于提供云服务的调度设备，装置包括：

划分模块 401，用于获取待调度的流量，将流量划分成多个流量块；

确定模块 402，用于确定多个流量块中的各个流量块的计算需求；

调度模块 403，用于根据各个流量块的计算需求，将各个流量块调度到各个流量块对应的目标节点，目标节点的计算资源用于处理目标节点对应的流量块的业务，目标节点的计算资源还用于处理由调度设备调度的除流量块之外的其他业务。

在一种可能的实现方式中，划分模块 401，用于获取多个计算节点中的各个计算节点的资源使用情况；

基于多个计算节点中存在任一计算节点的资源使用情况指示任一计算节点出现异常，将任一计算节点承载的流量作为待调度的流量，任一计算节点与目标节点为不同的计算节点。

在一种可能的实现方式中，确定模块 402，还用于基于任一计算节点的资源使用情况指示流量大小超过任一计算节点的第一阈值，或任一计算节点的数据处理单元的使用率超过任一计算节点的第二阈值中的至少一种，确定任一计算节点出现异常。

在一种可能的实现方式中，划分模块 401，用于根据流量对应的解析协议确定参考数量；将位于同一区域内的流量划分为参考数量个流量块。

在一种可能的实现方式中，确定模块 402，用于获取各个流量块的节点区域要求与计算资源需求，将节点区域要求与计算资源需求作为计算需求。

在一种可能的实现方式中，确定模块 402，用于构建各个流量块对应的区域标签，流量块对应的区域标签用于指示流量块对应的计算节点位于的区域；将各个流量块与各个流量块对应的区域标签之间的亲和关系作为各个流量块的节点区域要求，亲和关系用于确定流量块需要调度到的区域。

在一种可能的实现方式中，区域标签包括地区标签和运营商标签中的至少一种。

在一种可能的实现方式中，计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。

在一种可能的实现方式中，各个流量块的计算需求包括节点区域要求和计算资源需求，调度模块 403，用于对于任一流量的流量块，根据任一流量的流量块的计算需求筛选可调度的计算节点，得到筛选后的计算节点，筛选后的计算节点的资源使用情况满足任一流量的流量块的计算资源需求，且筛选后的计算节点的区域信息满足任一流量的流量块的节点区域要求；从筛选后的计算节点中选择计算节点作为任一流量的流量块对应的目标节点；将任一流量的流量块调度到任一流量的流量块对应的目标节点。

在一种可能的实现方式中，其他业务包括计算业务和存储业务中的至少一种。

在一种可能的实现方式中，目标节点属于边缘节点。

上述装置将流量划分成多个流量块，再确定各个流量块的计算需求，从而可以利用一个调度设备完成流量调度和其他业务的调度，调度设备具有通用调度能力。由于调度设备既可以调度流量，又可以调度其他业务，由同一个调度设备分配目标节点的计算资源给流量业务和其他业务，提高了计算资源的分配合理性，避免了计算节点的计算资源的浪费和超负荷使用，提升计算资源的资源利用率。

其中，划分模块 401、确定模块 402 和调度模块 403 均可以通过软件实现，或者可以通过硬件实现。示例性地，接下来以划分模块 401 为例，介绍划分模块 401 的实现方式。类似的，确定模块 402 和调度模块 403 的实现方式可以参考划分模块 401 的实现方式。

模块作为软件功能单元的一种举例，划分模块 401 可以包括运行在计算实例上的代码。其中，计算实例可以包括物理主机（计算设备）、虚拟机、容器中的至少一种。进一步地，上述计算实例可以是一台或者多台。例如，划分模块 401 可以包括运行在多个主机/虚拟机/容器上的代码。需要说明的是，用于运行该代码的多个主机/虚拟机/容器可以分布在相同的区域（region）中，也可以分布在不同的 region 中。进一步地，用于运行该代码的多个主机/虚拟机/容器可以分布在相同的可用区（availability zone, AZ）中，也可以分布在不同的 AZ 中，每个 AZ 包括一个数据中心或多个地理位置相近的数据中心。其中，通常一个 region 可以包括多个 AZ。

同样，用于运行该代码的多个主机/虚拟机/容器可以分布在同一个虚拟私有云（virtual private cloud, VPC）中，也可以分布在多个 VPC 中。其中，通常一个 VPC 设置在一个 region 内，同一 region 内两个 VPC 之间，以及不同 region 的 VPC 之间跨区通信需在每个 VPC 内设置通信网关，经通信网关实现 VPC 之间的互连。

模块作为硬件功能单元的一种举例，划分模块 401 可以包括至少一个计算设备，如服务器等。或者，划分模块 401 也可以是利用专用集成电路（application-specific integrated circuit, ASIC）实现、或可编程逻辑器件（programmable logic device, PLD）实现的设备等。其中，上述 PLD 可以是复杂程序逻辑器件（complex programmable logical device, CPLD）、现场可编程门阵列（field-programmable gate array, FPGA）、通用阵列逻辑（generic array logic, GAL）或其任意组合实现。

划分模块 401 包括的多个计算设备可以分布在相同的 region 中，也可以分布在不同的 region 中。划分模块 401 包括的多个计算设备可以分布在相同的 AZ 中，也可以分布在不同的 AZ 中。同样，划分模块 401 包括的多个计算设备可以分布在同一个 VPC 中，也可以分布在多个 VPC 中。其中，多个计算设备可以是

服务器、ASIC、PLD、CPLD、FPGA 和 GAL 等计算设备的任意组合。

需要说明的是，在其他实施例中，划分模块 401 可以用于执行流量调度方法中的任意步骤，确定模块 402 可以用于执行流量调度方法中的任意步骤，调度模块 403 可以用于执行流量调度方法中的任意步骤，划分模块 401、确定模块 402、以及调度模块 403 负责实现的步骤可根据需要指定，通过划分模块 401、确定模块 402、以及调度模块 403 分别实现流量调度方法中不同的步骤来实现流量调度装置的全部功能。

本申请还提供一种计算设备 500，可选地，该计算设备 500 为本申请实施例提供的调度设备。如图 5 所示，计算设备 500 包括：总线 502、处理器 504、存储器 506 和通信接口 508。处理器 504、存储器 506 和通信接口 508 之间通过总线 502 通信。计算设备 500 可以是服务器或终端设备。应理解，本申请不限定计算设备 500 中的处理器、存储器的个数。

总线 502 可以是外设部件互连标准（peripheral component interconnect, PCI）总线或扩展工业标准结构（extended industry standard architecture, EISA）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图 5 中仅用一条线表示，但并不表示仅有一根总线或一种类型的总线。总线 502 可包括在计算设备 500 各个部件（例如，存储器 506、处理器 504、通信接口 508）之间传送信息的通路。

处理器 504 可以包括中央处理器（central processing unit, CPU）、图形处理器（graphics processing unit, GPU）、微处理器（micro processor, MP）或者数字信号处理器（digital signal processor, DSP）等处理器中的任意一种或多种。

存储器 506 可以包括易失性存储器（volatile memory），例如随机存取存储器（random access memory, RAM）。处理器 504 还可以包括非易失性存储器（non-volatile memory），例如只读存储器（read-only memory, ROM），快闪存储器，机械硬盘（hard disk drive, HDD）或固态硬盘（solid state drive, SSD）。

存储器 506 中存储有可执行的程序代码，处理器 504 执行该可执行的程序代码以分别实现前述获取模块、确定模块和分组模块的功能，从而实现流量调度方法。也即，存储器 506 上存有用于执行流量调度方法的指令。

通信接口 508 使用例如但不限于网络接口卡、收发器一类的收发模块，来实现计算设备 500 与其他设备或通信网络之间的通信。

本申请实施例还提供了一种计算设备集群。该计算设备集群包括至少一台计算设备。该计算设备可以是服务器，例如是中心服务器、边缘服务器，或者是本地数据中心中的本地服务器。在一些实施例中，计算设备也可以是台式机、笔记本电脑或者智能手机等终端设备。

可选地，计算设备集群包括的至少一个计算设备的结构可参见图 5 示出的计算设备 500。计算设备集群中的一个或多个计算设备 500 中的存储器 506 中可以存有相同的用于执行流量调度方法的指令。

在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备 500 的存储器 506 中也可以分别存有用于执行流量调度方法的部分指令。换言之，一个或多个计算设备 500 的组合可以共同执行用于执行流量调度方法的指令。

需要说明的是，计算设备集群中的不同的计算设备 500 中的存储器 506 可以存储不同的指令，分别用于执行流量调度装置的部分功能。也即，不同的计算设备 500 中的存储器 506 存储的指令可以实现划分模块、确定模块和调度模块中的一个或多个模块的功能。

在一些可能的实现方式中，计算设备集群中的一个或多个计算设备可以通过网络连接。其中，网络可以是广域网或局域网等等。图 6 示出了一种可能的实现方式。如图 6 所示，两个计算设备 600A 和 600B 之间通过网络进行连接。具体地，通过各个计算设备中的通信接口与网络进行连接。在这一类可能的实现方式中，计算设备 600A 和 600B 包括总线 602、处理器 604、存储器 606 和通信接口 608。计算设备 600A 中的存储器 606 中存有执行划分模块的功能的指令。同时，计算设备 600B 中的存储器 606 中存有执行确定模块和调度模块的功能的指令。

图 6 所示的计算设备集群之间的连接方式可以是考虑到本申请提供的流量调度方法需要将划分流量得到的流量块调度到流量块对应的目标节点上，因此考虑将确定模块和调度模块实现的功能交由计算设备 600B 执行。

应理解，图 6 中示出的计算设备 600A 的功能也可以由多个计算设备 600 完成。同样，计算设备 600B

的功能也可以由多个计算设备 600 完成。

本申请实施例还提供了一种通信装置，该装置包括：收发器、存储器和处理器。其中，该收发器、该存储器和该处理器通过内部连接通路互相通信，该存储器用于存储指令，该处理器用于执行该存储器存储的指令，以控制收发器接收信号，并控制收发器发送信号，并且当该处理器执行该存储器存储的指令时，使得该处理器执行流量调度方法。

应理解的是，上述处理器可以是中央处理器（Central Processing Unit, CPU），还可以是其他通用处理器、数字信号处理器（digital signal processing, DSP）、专用集成电路（application specific integrated circuit, ASIC）、现场可编程门阵列（field-programmable gate array, FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者是任何常规的处理器等。值得说明的是，处理器可以是支持进阶精简指令集机器（advanced RISC machines, ARM）架构的处理器。

进一步地，在一种可选的实施例中，上述存储器可以包括只读存储器和随机存取存储器，并向处理器提供指令和数据。存储器还可以包括非易失性随机存取存储器。例如，存储器还可以存储设备类型的信息。

该存储器可以是易失性存储器或非易失性存储器，或可包括易失性和非易失性存储器两者。其中，非易失性存储器可以是只读存储器（read-only memory, ROM）、可编程只读存储器（programmable ROM, PROM）、可擦除可编程只读存储器（erasable PROM, EPROM）、电可擦除可编程只读存储器（electrically EPROM, EEPROM）或闪存。易失性存储器可以是随机存取存储器（random access memory, RAM），其用作外部高速缓存。通过示例性但不是限制性说明，许多形式的 RAM 可用。例如，静态随机存取存储器（static RAM, SRAM）、动态随机存取存储器（dynamic random access memory, DRAM）、同步动态随机存取存储器（synchronous DRAM, SDRAM）、双倍数据速率同步动态随机存取存储器（double data rate SDRAM, DDR SDRAM）、增强型同步动态随机存取存储器（enhanced SDRAM, ESDRAM）、同步连接动态随机存取存储器（synchlink DRAM, SDRAM）和直接内存总线随机存取存储器（direct rambus RAM, DR RAM）。

本申请实施例还提供了一种包含指令的计算机程序（产品）。计算机程序（产品）可以是包含指令的，能够运行在计算设备上或被储存在任何可用介质中的软件或程序产品。当计算机程序（产品）在至少一个计算设备上运行时，使得至少一个计算设备执行本申请实施例提供的流量调度方法。

本申请实施例还提供了一种计算机可读存储介质。计算机可读存储介质可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘）等。该计算机可读存储介质包括指令，指令指示计算设备执行流量调度方法。

本申请实施例还提供了一种芯片，包括处理器，用于从存储器中调用并运行存储器中存储的指令，使得安装有芯片的通信设备执行本申请实施例提供的流量调度方法。

本申请实施例还提供另一种芯片，包括：输入接口、输出接口、处理器和存储器，输入接口、输出接口、处理器以及存储器之间通过内部连接通路相连，处理器用于执行存储器中的代码，当代码被执行时，处理器用于执行本申请实施例提供的流量调度方法。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。计算机程序产品包括一个或多个计算机指令。在计算机上加載和执行计算机程序指令时，全部或部分地产生按照本申请所示的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线）或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。该可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘 Solid State Disk）等。

为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各实施例的步骤及组成。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。本领域普通技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

用于实现本申请实施例的方法的计算机程序代码可以用一种或多种编程语言编写。这些计算机程序代码可以提供给通用计算机、专用计算机或其他可编程的规则查找装置的处理器的处理器，使得程序代码在被计算机或其他可编程的规则查找装置执行的时候，引起在流程图和/或框图中规定的功能/操作被实施。程序代码可以完全在计算机上、部分在计算机上、作为独立的软件包、部分在计算机上且部分在远程计算机上或完全在远程计算机或服务器上执行。

在本申请实施例的上下文中，计算机程序代码或者相关数据可以由任意适当载体承载，以使得设备、装置或者处理器能够执行上文描述的各种处理和操作。载体的示例包括信号、计算机可读介质等等。信号的示例可以包括电、光、无线电、声音或其它形式的传播信号，诸如载波、红外信号等。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的系统、设备和模块的具体工作过程，可以参见前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、设备和方法，可以通过其它的方式实现。例如，以上所描述的设备实施例仅仅是示意性的，例如，该模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、设备或模块的间接耦合或通信连接，也可以是电的，机械的或其它的形式连接。

该作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。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申请实施例方案的目的。

另外，在本申请各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以是两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。

本申请中术语“第一”“第二”等字样用于对作用和功能基本相同的相同项或相似项进行区分，应理解，“第一”、“第二”、“第n”之间不具有逻辑或时序上的依赖关系，也不对数量和执行顺序进行限定。还应理解，尽管以下描述使用术语第一、第二等来描述各种元素，但这些元素不应受术语的限制。这些术语只是用于将一元素与另一元素区别分开。例如，在不脱离各种示例的范围的情况下，第一链路可以被称为第二链路，并且类似地，第二链路可以被称为第一链路。

还应理解，在本申请的各个实施例中，各个过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

本申请中术语“至少一个”的含义是指一个或多个，本申请中术语“多个”的含义是指两个或两个以上，例如，多个第二报文是指两个或两个以上的第二报文。本文中术语“系统”和“网络”经常可互换使用。

应理解，在本文中对各种示例的描述中所使用的术语只是为了描述特定示例，而并非旨在进行限制。如在对各种示例的描述和所附权利要求书中所使用的那样，单数形式“一个（“a”，“an”）”和“该”旨在也包括复数形式，除非上下文另外明确地指示。

还应理解，术语“包括”（也称“includes”、“including”、“comprises”和/或“comprising”）当在本说明书中使用指定存在所陈述的特征、整数、步骤、操作、元素、和/或部件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元素、部件、和/或其分组。

还应理解，根据上下文，短语“若确定...”或“若检测到[所陈述的条件或事件]”可被解释为意指“在确定...时”或“响应于确定...”或“在检测到[所陈述的条件或事件]时”或“响应于检测到[所陈述的条件或事件]”。

应理解，根据A确定B并不意味着仅仅根据A确定B，还可以根据A和/或其它信息确定B。

还应理解，说明书通篇中提到的“一个实施例”、“一实施例”、“一种可能的实现方式”意味着与实施例或实现方式有关的特定特征、结构或特性包括在本申请的至少一个实施例中。因此，在整个说明书各处

出现的“在一个实施例中”或“在一实施例中”、“一种可能的实现方式”未必一定指相同的实施例。此外，这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。

需要说明的是，本申请所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。例如，本申请中涉及到的待调度的流量等都是在充分授权的情况下获取的。

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保护范围。

权利要求书

- 1.一种流量调度方法，其特征在于，所述方法由提供云服务的调度设备执行，所述方法包括：
所述调度设备获取待调度的流量，将所述流量划分成多个流量块；
所述调度设备确定所述多个流量块中的各个流量块的计算需求；
所述调度设备根据所述各个流量块的计算需求，将所述各个流量块调度到所述各个流量块对应的目标节点，所述目标节点的计算资源用于处理所述目标节点对应的流量块的业务，所述目标节点的计算资源还用于处理由所述调度设备调度的除所述流量块之外的其他业务。
- 2.根据权利要求1所述的方法，其特征在于，所述获取待调度的流量，包括：
获取多个计算节点中的各个计算节点的资源使用情况；
基于所述多个计算节点中存在任一计算节点的资源使用情况指示所述任一计算节点出现异常，将所述任一计算节点承载的流量作为所述待调度的流量，所述任一计算节点与所述目标节点为不同的计算节点。
- 3.根据权利要求2所述的方法，其特征在于，所述基于所述多个计算节点中存在任一计算节点的资源使用情况指示所述任一计算节点出现异常，将所述任一计算节点承载的流量作为所述待调度的流量之前，还包括：
基于所述任一计算节点的资源使用情况指示流量大小超过所述任一计算节点的第一阈值，或所述任一计算节点的数据处理单元的使用率超过所述任一计算节点的第二阈值中的至少一种，确定所述任一计算节点出现异常。
- 4.根据权利要求1-3任一所述的方法，其特征在于，所述将所述流量划分成多个流量块，包括：
根据所述流量对应的解析协议确定参考数量；
将位于同一区域内的流量划分为参考数量个流量块。
- 5.根据权利要求1-4任一所述的方法，其特征在于，所述确定所述多个流量块中的各个流量块的计算需求，包括：
获取所述各个流量块的节点区域要求与计算资源需求，将所述节点区域要求与所述计算资源需求作为所述计算需求。
- 6.根据权利要求5所述的方法，其特征在于，所述获取所述各个流量块的节点区域要求，包括：
构建所述各个流量块对应的区域标签，所述流量块对应的区域标签用于指示所述流量块对应的计算节点位于的区域；
将所述各个流量块与所述各个流量块对应的区域标签之间的亲和关系作为所述各个流量块的节点区域要求，所述亲和关系用于确定所述流量块需要调度到的区域。
- 7.根据权利要求6所述的方法，其特征在于，所述区域标签包括地区标签和运营商标签中的至少一种。
- 8.根据权利要求5-7任一所述的方法，其特征在于，所述计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。
- 9.根据权利要求1-8任一所述的方法，其特征在于，所述各个流量块的计算需求包括节点区域要求和计算资源需求，所述根据所述各个流量块的计算需求，将所述各个流量块调度到所述各个流量块对应的目标节点，包括：
对于任一流量块，根据所述任一流量块的计算需求筛选可调度的计算节点，得到筛选后的计算节点，所述筛选后的计算节点的资源使用情况满足所述任一流量块的计算资源需求，且所述筛选后的计算节点的

区域信息满足所述任一流量块的节点区域要求；

从所述筛选后的计算节点中选择计算节点作为所述任一流量块对应的目标节点；

将所述任一流量块调度到所述任一流量块对应的目标节点。

10.根据权利要求1-9任一所述的方法，其特征在于，所述其他业务包括计算业务和存储业务中的至少一种。

11.根据权利要求1-10任一所述的方法，其特征在于，所述目标节点属于边缘节点。

12.一种流量调度装置，其特征在于，所述装置应用于提供云服务的调度设备，所述装置包括：

划分模块，用于获取待调度的流量，将所述流量划分成多个流量块；

确定模块，用于确定所述多个流量块中的各个流量块的计算需求；

调度模块，用于根据所述各个流量块的计算需求，将所述各个流量块调度到所述各个流量块对应的目标节点，所述目标节点的计算资源用于处理所述目标节点对应的流量块的业务，所述目标节点的计算资源还用于处理由所述调度设备调度的除所述流量块之外的其他业务。

13.根据权利要求12所述的装置，其特征在于，所述划分模块，用于获取多个计算节点中的各个计算节点的资源使用情况；基于所述多个计算节点中存在任一计算节点的资源使用情况指示所述任一计算节点出现异常，将所述任一计算节点承载的流量作为所述待调度的流量，所述任一计算节点与所述目标节点为不同的计算节点。

14.根据权利要求13所述的装置，其特征在于，所述确定模块，还用于基于所述任一计算节点的资源使用情况指示流量大小超过所述任一计算节点的第一阈值，或所述任一计算节点的数据处理单元的使用率超过所述任一计算节点的第二阈值中的至少一种，确定所述任一计算节点出现异常。

15.根据权利要求12-14任一所述的装置，其特征在于，所述划分模块，用于根据所述流量对应的解析协议确定参考数量；将位于同一区域内的流量划分为参考数量个流量块。

16.根据权利要求12-15任一所述的装置，其特征在于，所述确定模块，用于获取所述各个流量块的节点区域要求与计算资源需求，将所述节点区域要求与所述计算资源需求作为所述计算需求。

17.根据权利要求16所述的装置，其特征在于，所述确定模块，用于构建所述各个流量块对应的区域标签，所述流量块对应的区域标签用于指示所述流量块对应的计算节点位于的区域；将所述各个流量块与所述各个流量块对应的区域标签之间的亲和关系作为所述各个流量块的节点区域要求，所述亲和关系用于确定所述流量块需要调度到的区域。

18.根据权利要求17所述的装置，其特征在于，所述区域标签包括地区标签和运营商标签中的至少一种。

19.根据权利要求16-18任一所述的装置，其特征在于，所述计算资源需求包括带宽需求、内存需求和数据处理单元需求中的至少一种。

20.根据权利要求12-19任一所述的装置，其特征在于，所述各个流量块的计算需求包括节点区域要求和计算资源需求，所述调度模块，用于对于任一流量块，根据所述任一流量块的计算需求筛选可调度的计算节点，得到筛选后的计算节点，所述筛选后的计算节点的资源使用情况满足所述任一流量块的计算资源需求，且所述筛选后的计算节点的区域信息满足所述任一流量块的节点区域要求；从所述筛选后的计算节点中选择计算节点作为所述任一流量块对应的目标节点；将所述任一流量块调度到所述任一流量块对应的目标节点。

21.根据权利要求12-20任一所述的装置，其特征在于，所述其他业务包括计算业务和存储业务中的至少一种。

22.根据权利要求12-21任一所述的装置，其特征在于，所述目标节点属于边缘节点。

23.一种计算设备集群，其特征在于，所述计算设备集群包括至少一个计算设备，每个计算设备包括处理器和存储器；所述至少一个计算设备的处理器用于执行所述至少一个计算设备的存储器中存储的指令，以使得所述计算设备集群执行如权利要求1-11中任一所述的流量调度方法。

24.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质包括计算机程序指令，当所述计算机程序指令由计算设备集群执行时，所述计算设备集群执行权利要求1-11中任一所述的流量调度方法。

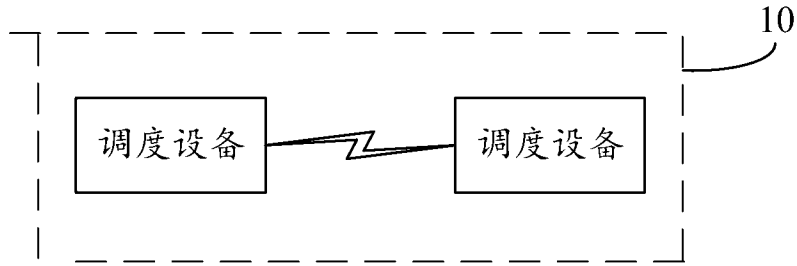


图 1

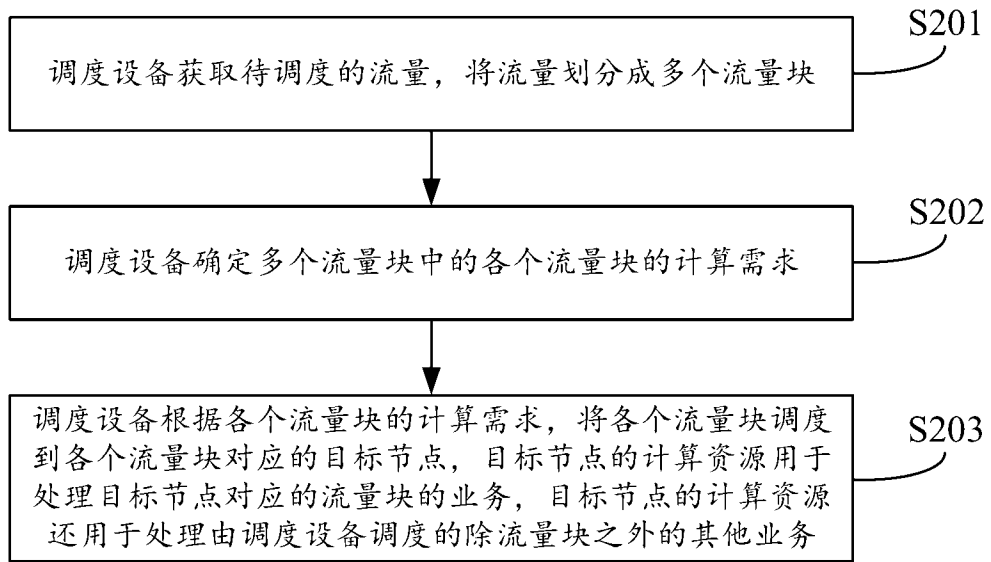


图 2

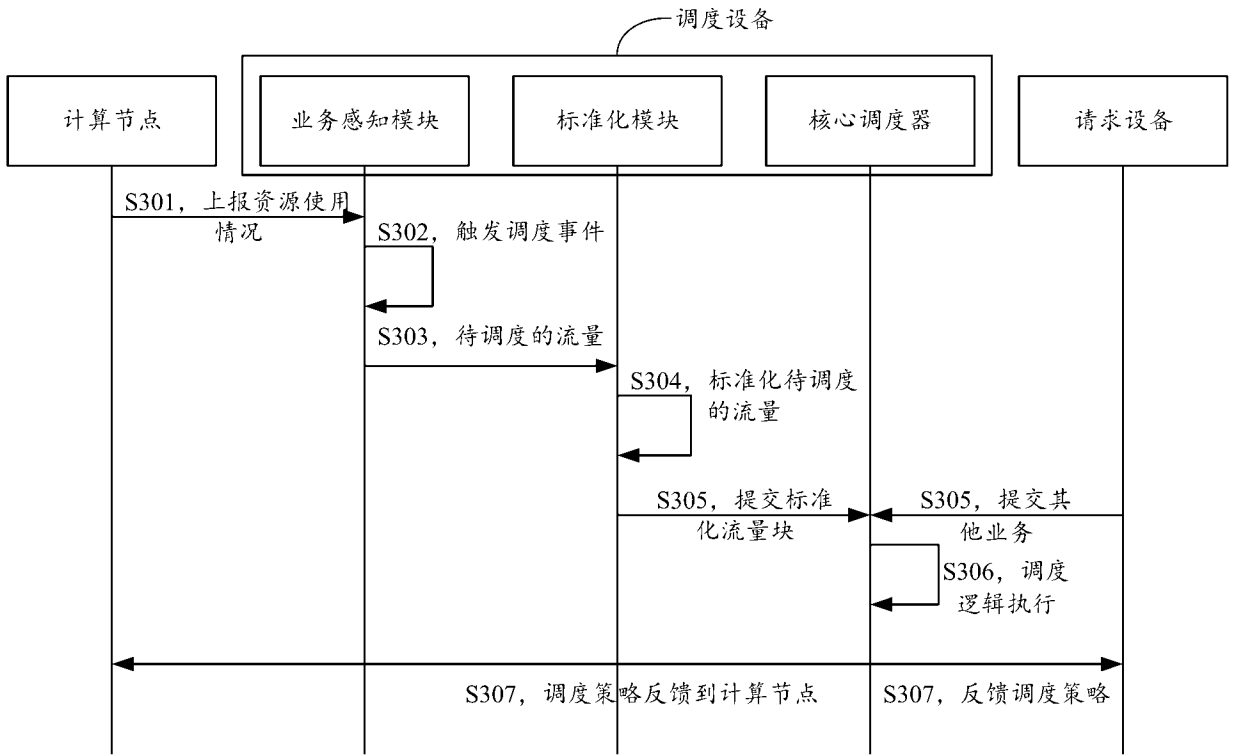


图 3

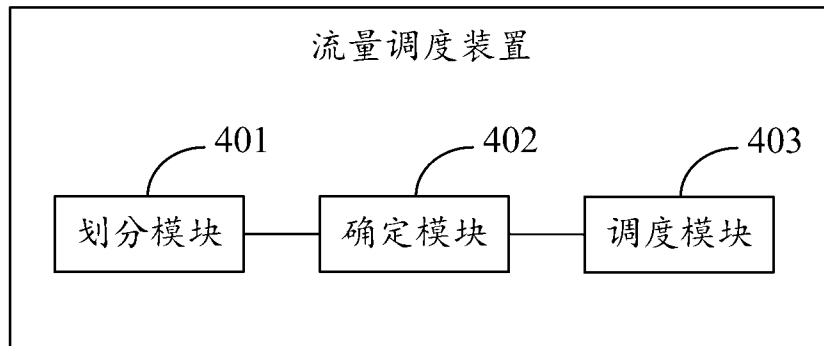


图 4

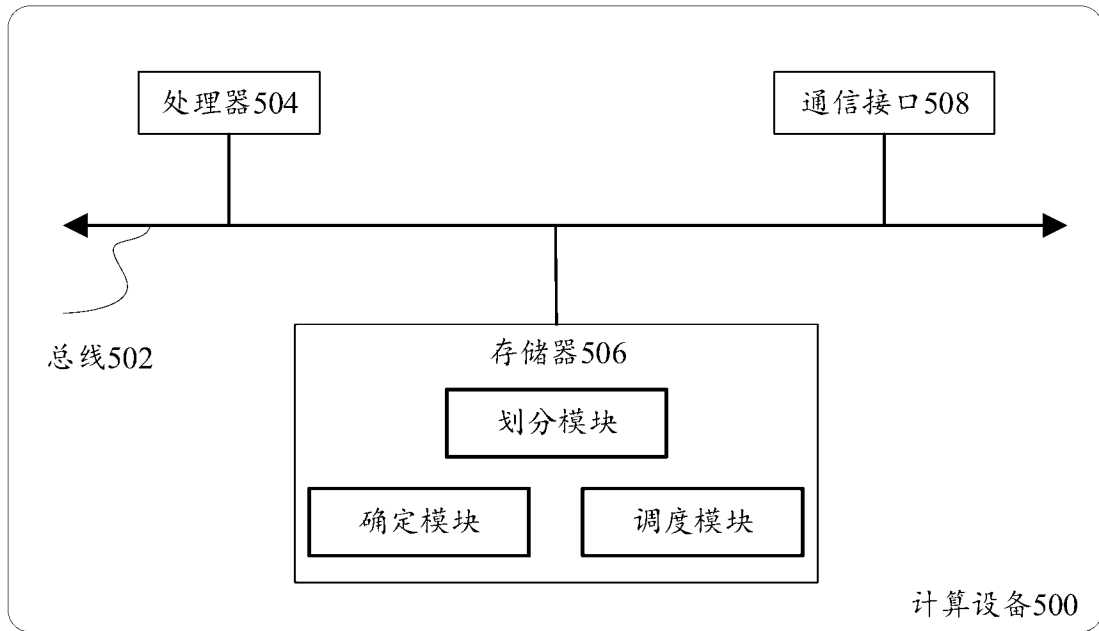


图 5

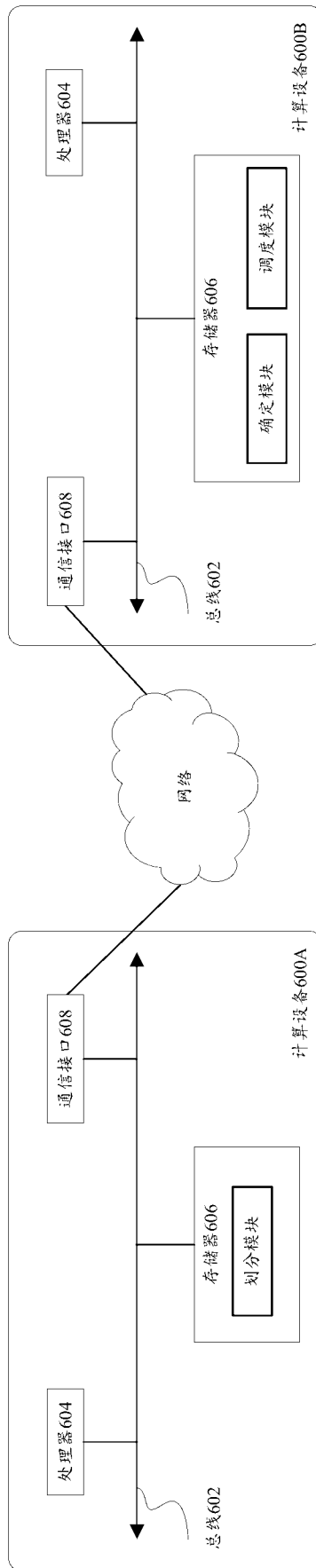


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/104095

A. CLASSIFICATION OF SUBJECT MATTER		
H04L 47/10(2022.01)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)		
IPC: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)		
CNABS; CNTXT; CNKI; VEN; USTXT; WOTXT; EPTXT; IEEE: 流量, 调度, 需求, 匹配, 资源, 节点, 控制, 负载均衡, 异常, flow, schedule, demand, match, control, node, load balance, fault		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 111432247 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 17 July 2020 (2020-07-17) description, paragraphs [0056]-[0104]	1-24
Y	CN 114710563 A (HUAWEI TECHNOLOGIES CO., LTD.) 05 July 2022 (2022-07-05) description, paragraphs [0051]-[0070]	1-24
Y	US 2011072437 A1 (IBM) 24 March 2011 (2011-03-24) claims 1-2, and description, paragraphs [0022]-[0037]	1-24
A	CN 107872402 A (BEIJING QIYI CENTURY SCIENCE & TECHNOLOGY CO., LTD.) 03 April 2018 (2018-04-03) entire document	1-24
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 "D" document cited by the applicant in the international application "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		Date of mailing of the international search report
18 September 2023		19 September 2023
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/CN2023/104095

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111432247	A	17 July 2020	WO	2021184589	A1	23 September 2021
CN	114710563	A	05 July 2022	None			
US	2011072437	A1	24 March 2011	US	9015724	B2	21 April 2015
CN	107872402	A	03 April 2018	CN	107872402	B	09 April 2021

国际检索报告

国际申请号

PCT/CN2023/104095

<p>A. 主题的分类</p> <p>H04L 47/10 (2022.01) i</p> <p>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</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>IPC:H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS;CNTXT;CNKI;VEN;USTXT;WOTXT;EPTXT;IEEE:流量、调度、需求、匹配、资源、节点、控制、负载均衡、异常、flow、schedule、demand、match、control、node、load balance、faul</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 111432247 A (平安科技(深圳)有限公司) 2020年7月17日 (2020 - 07 - 17) 说明书第[0056]-[0104]段</td> <td>1-24</td> </tr> <tr> <td>Y</td> <td>CN 114710563 A (华为技术有限公司) 2022年7月5日 (2022 - 07 - 05) 说明书第[0051]-[0070]段</td> <td>1-24</td> </tr> <tr> <td>Y</td> <td>US 2011072437 A1 (IBM) 2011年3月24日 (2011 - 03 - 24) 权利要求1-2, 说明书第[0022]-[0037]段</td> <td>1-24</td> </tr> <tr> <td>A</td> <td>CN 107872402 A (北京奇艺世纪科技有限公司) 2018年4月3日 (2018 - 04 - 03) 全文</td> <td>1-24</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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 “D” 申请人在国际申请中引证的文件 “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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 111432247 A (平安科技(深圳)有限公司) 2020年7月17日 (2020 - 07 - 17) 说明书第[0056]-[0104]段	1-24	Y	CN 114710563 A (华为技术有限公司) 2022年7月5日 (2022 - 07 - 05) 说明书第[0051]-[0070]段	1-24	Y	US 2011072437 A1 (IBM) 2011年3月24日 (2011 - 03 - 24) 权利要求1-2, 说明书第[0022]-[0037]段	1-24	A	CN 107872402 A (北京奇艺世纪科技有限公司) 2018年4月3日 (2018 - 04 - 03) 全文	1-24
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
Y	CN 111432247 A (平安科技(深圳)有限公司) 2020年7月17日 (2020 - 07 - 17) 说明书第[0056]-[0104]段	1-24															
Y	CN 114710563 A (华为技术有限公司) 2022年7月5日 (2022 - 07 - 05) 说明书第[0051]-[0070]段	1-24															
Y	US 2011072437 A1 (IBM) 2011年3月24日 (2011 - 03 - 24) 权利要求1-2, 说明书第[0022]-[0037]段	1-24															
A	CN 107872402 A (北京奇艺世纪科技有限公司) 2018年4月3日 (2018 - 04 - 03) 全文	1-24															
<p>国际检索实际完成的日期</p> <p>2023年9月18日</p>	<p>国际检索报告邮寄日期</p> <p>2023年9月19日</p>																
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088</p>	<p>授权官员</p> <p>陈相玫</p> <p>电话号码 (+86) 020-28957158</p>																

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/104095

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	111432247	A	2020年7月17日	WO	2021184589	A1	2021年9月23日
CN	114710563	A	2022年7月5日	无			
US	2011072437	A1	2011年3月24日	US	9015724	B2	2015年4月21日
CN	107872402	A	2018年4月3日	CN	107872402	B	2021年4月9日