

(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
2016 年 1 月 28 日 (28.01.2016)



(10) 国际公布号
WO 2016/011826 A1

(51) 国际专利分类号:
H04L 29/08 (2006.01)

(21) 国际申请号: PCT/CN2015/074872

(22) 国际申请日: 2015 年 3 月 23 日 (23.03.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201410357045.4 2014 年 7 月 24 日 (24.07.2014) CN

(71) 申请人: 深圳天珑无线科技有限公司 (SHENZHEN TINNO WIRELESS TECHNOLOGY CO., LTD) [CN/CN]; 中国广东省深圳市南山区华侨城东部工业区 H3 栋 2-4 楼, Guangdong 518053 (CN)。 江苏省天珑电子科技有限公司 (JIANGSU TINNO ELECTRONICS TECHNOLOGY CO., LTD) [CN/CN]; 中国江苏省南京市秦淮区应天大街 388 号 1865 创意园 E13 东座, Jiangsu 210000 (CN)。

(72) 发明人: 经华 (JING, Hua); 中国广东省深圳市南山区华侨城东部工业区 H3 栋 2-4 楼, Guangdong 518053 (CN)。 项浩 (XIANG, Hao); 中国广东省深

圳市南山区华侨城东部工业区 H3 栋 2-4 楼, Guangdong 518053 (CN)。

(74) 代理人: 深圳市威世博知识产权代理事务所 (普通合伙) (CHINA WISPRO INTELLECTUAL PROPERTY LLP.); 中国广东省深圳市南山区高新区粤兴三道 8 号中国地质大学产学研基地中地大楼 A806, Guangdong 518057 (CN)。

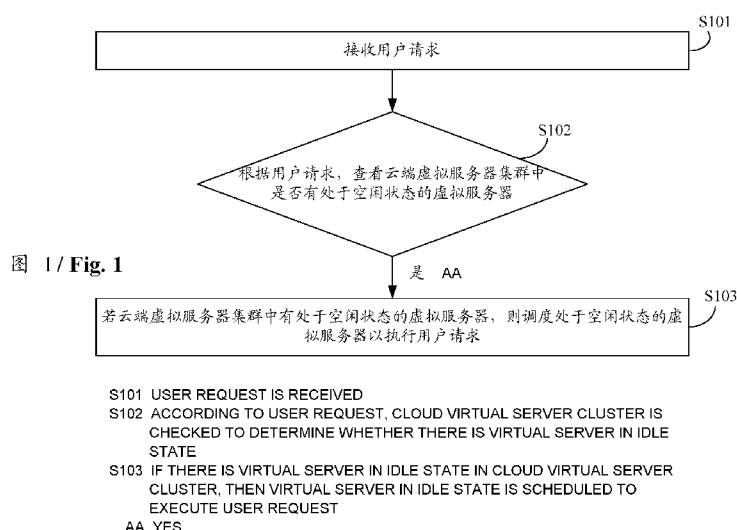
(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH,

[见续页]

(54) Title: CLOUD VIRTUAL SERVER SCHEDULING METHOD AND DEVICE

(54) 发明名称: 云端虚拟服务器的调度方法及装置



(57) Abstract: The present invention discloses a cloud virtual server scheduling method and device, the method comprising: according to a user request, a cloud virtual server cluster is checked to determine whether there is a virtual server in an idle state, a virtual server in an idle state being taken to mean a virtual server not currently being used; if there is a virtual server in an idle state in the cloud virtual server cluster, then the virtual server in an idle state is scheduled to execute the user request. In the described manner, the present invention improves the utilization rate of virtual servers, thereby conserving resources.

(57) 摘要: 本发明公开了一种云端虚拟服务器的调度方法及装置, 该方法包括: 根据用户请求, 查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器, 处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器; 若云端虚拟服务器集群中有处于空闲状态的虚拟服务器, 则调度处于空闲状态的虚拟服务器以执行用户请求。通过上述方式, 本发明能够更好地提高虚拟服务器的使用率, 从而节约资源。



WO 2016/011826 A1



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))。

云端虚拟服务器的调度方法及装置

[1] **【技术领域】**

[2] 本发明涉及云计算技术领域，特别是涉及一种云端虚拟服务器的调度方法及装置。

[3] **【背景技术】**

[4] 云计算可以让资源共享，因此可以提高硬件资源的使用率。在云端虚拟机的调度上，由于用户的使用情况不同，需求配置不同，使得资源的利用率并不高。一个合理的调度能够极大提升用户体验和执行效率。

[5] 目前，大多数云服务供应商，比如亚马逊和阿里云，在云虚拟服务器的管理上，采用的是通过用户自己来控制的方法来实现管理的，即用户需要时启动虚拟服务器，并开始计费，不需要时关闭虚拟服务器，并停止计费。

[6] 但是，本申请的发明人在长期的研发中发现：采用上述传统的方式主要是因为便于计费，缺点也是显而易见的，即启动虚拟服务器后在很多情况下用户并没有连续使用虚拟服务器，在用户没有使用虚拟服务器的情况下，仍然会占用虚拟服务器资源。

[7] **【发明内容】**

[8] 本发明主要解决的技术问题是提供一种云端虚拟服务器的调度方法及装置，能够提高虚拟服务器的使用率。

[9] 为解决上述技术问题，本发明采用的一个技术方案是：提供一种云端虚拟服务器的调度方法，包括：创建所述云端虚拟服务器集群；初始化所述云端虚拟服务器集群中的虚拟服务器；接收用户请求；根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定数量的虚拟服务器；若所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度所述处于空闲状态的虚拟服务器以执行所述用户请求；若所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器，则创建并初始化n个虚拟

服务器，所述 n 为所述用户请求的个数；调度所述 n 个虚拟服务器以执行所述用户请求。

[10] 其中，该方法还包括若所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量，则逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。

[11] 其中，监控所述云端虚拟服务器集群中虚拟服务器的状态；若监控到所述虚拟服务器当前没有被使用，则将所述虚拟服务器的状态标记为空闲状态。

[12] 为解决上述技术问题，本发明采用的另一个技术方案是：提供一种云端虚拟服务器的调度方法，包括：接收用户请求；根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定数量的虚拟服务器；若所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度所述处于空闲状态的虚拟服务器以执行所述用户请求。

[13] 其中，所述接收用户请求的步骤之前，包括：创建所述云端虚拟服务器集群；初始化所述云端虚拟服务器集群中的虚拟服务器。

[14] 其中，根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器的步骤之后，包括：若所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器，则创建并初始化 n 个虚拟服务器，所述 n 为所述用户请求的个数；调度所述 n 个虚拟服务器以执行所述用户请求。

[15] 其中，所述方法还包括：若所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量，则逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。

[16] 其中，所述方法还包括：监控所述云端虚拟服务器集群中虚拟服务器的状态；若监控到所述虚拟服务器当前没有被使用，则将所述虚拟服务器的状态标记为空闲状态。

[17] 为解决上述技术问题，本发明采用的又一个技术方案是：提供一种云端虚拟服

务器的调度装置，所述装置包括：接收模块，用于接收用户请求；查看模块，用于根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定数量的虚拟服务器；第一调度模块，用于在所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器时，调度所述处于空闲状态的虚拟服务器以执行所述用户请求。

[18] 其中，所述装置还包括：第一创建模块，用于创建所述云端虚拟服务器集群；第一初始化模块，用于初始化所述云端虚拟服务器集群中的虚拟服务器。

[19] 其中，所述装置还包括：第二创建模块，用于在所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器时，创建n个虚拟服务器，所述n为所述用户请求的个数；第二初始化模块，用于初始化所述n个虚拟服务器；第二调度模块，用于调度所述n个虚拟服务器以执行所述用户请求。

[20] 其中，所述装置还包括回收模块，所述回收模块用于在所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量时，逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。

[21] 其中，所述装置还包括：监控模块，用于监控所述云端虚拟服务器集群中虚拟服务器的状态；标记模块，用于在监控到所述虚拟服务器当前没有被使用时，将所述虚拟服务器的状态标记为空闲状态。

[22] 本发明的有益效果是：区别于现有技术的情况，本发明根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器；若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。由于不考虑虚拟服务器是否被其它用户释放，只要有处于空闲状态的虚拟服务器时，直接调用处于空闲状态的虚拟服务器以执行用户请求，通过这种方式，可以更好地提高虚拟服务器的使用率，从而节约资源。

[23] 【附图说明】

[24] 图1是本发明云端虚拟服务器的调度方法一实施方式的流程图；

- [25] 图2是本发明云端虚拟服务器的调度方法另一实施方式的流程图；
- [26] 图3是本发明云端虚拟服务器的调度方法又一实施方式的流程图；
- [27] 图4是本发明云端虚拟服务器的调度方法又一实施方式的流程图；
- [28] 图5是本发明云端虚拟服务器的调度装置一实施方式的结构示意图；
- [29] 图6是本发明云端虚拟服务器的调度装置另一实施方式的结构示意图；
- [30] 图7是本发明云端虚拟服务器的调度装置又一实施方式的结构示意图；
- [31] 图8是本发明云端虚拟服务器的调度装置又一实施方式的结构示意图。

[32] **【具体实施方式】**

[33] 下面结合附图和实施方式对本发明进行详细说明。

[34] 参阅图1，图1是本发明云端虚拟服务器的调度方法一实施方式的流程图，包括：

[35] 步骤S101：接收用户请求。

[36] 用户请求是指用户请求使用虚拟服务器，以使得虚拟服务器向用户提供服务。

[37] 步骤S102：根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，云端虚拟服务器集群包括第一预定数量的虚拟服务器。

[38] 云端虚拟服务器集群是包括第一预定数量的虚拟服务器的集群，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，例如：被其它用户释放当前处于空闲状态的虚拟服务器；初始化后一直处于空闲状态的虚拟服务器；其它用户没有释放、但当前没有被使用的虚拟服务器等等。

[39] 在接收到用户请求后，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，以便于调度虚拟服务器。

[40] 步骤S103：若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。

[41] 如果云端虚拟服务器集群中有处于空闲状态的虚拟服务器，那么这些虚拟服务器可以调度来执行用户请求，从而有效地提高虚拟服务器的使用率。

[42] 本发明实施方式根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务

器；若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。由于不考虑虚拟服务器是否被其它用户释放，只要有处于空闲状态的虚拟服务器时，直接调用处于空闲状态的虚拟服务器以执行用户请求，通过这种方式，可以更好地提高虚拟服务器的使用率，从而节约资源。

[43] 参阅图2至图4，图2至图4是本发明云端虚拟服务器的调度方法三个实施方式的流程图，本三个实施方式和图1的方法基本相同，相同之处请参见图1和相关的文字说明，不同之处具体请参见如下内容：

[44] 步骤S201：创建云端虚拟服务器集群。

[45] 步骤S202：初始化云端虚拟服务器集群中的虚拟服务器。

[46] 预先创建云端虚拟服务器集群，并初始化虚拟服务器，构成一批处于空闲状态的虚拟服务器，等待请求调用。由于虚拟服务器提前已经创建并初始化了，需要调用的时候，直接可以调用，从而可以避免在调用时再创建并初始化虚拟服务器所需要等待的时间，通过这种方式，可以提高用户体验。

[47] 步骤S203：接收用户请求。

[48] 步骤S204：根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，云端虚拟服务器集群包括第一预定数量的虚拟服务器。

[49] 步骤S205：若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。

[50] 步骤S206：若云端虚拟服务器集群中没有处于空闲状态的虚拟服务器，则创建并初始化n个虚拟服务器，n为用户请求的个数。

[51] 步骤S207：调度n个虚拟服务器以执行用户请求。

[52] 如果云端虚拟服务器集群中所有的虚拟服务器都已经被用户使用，没有处于空闲状态的虚拟机，那么根据用户请求的个数另外创建n个虚拟服务器并初始化，然后调度这n个虚拟服务器以执行用户请求，为用户提供服务。

[53] 步骤S208：监控云端虚拟服务器集群中虚拟服务器的状态。

[54] 步骤S209：若监控到虚拟服务器当前没有被使用，则将虚拟服务器的状态标记

为空闲状态。

[55] 在整个方法流程过程中，随时监控云端虚拟服务器集群中虚拟服务器的状态，如果监控到存在当前没有被使用的虚拟服务器，则将虚拟服务器的状态标记为空闲状态，其中，当前没有被使用的虚拟服务器，例如可以是虚拟服务器被释放、初始化后没有被使用、或当前暂时没有被使用等等。

[56] 步骤S210：若处于空闲状态的虚拟服务器的数量达到第二预定数量，且云端虚拟服务器集群中虚拟服务器的数量大于第一预定数量，则逐渐回收虚拟服务器，以使得云端虚拟服务器集群中虚拟服务器的数量大于或等于第一预定数量。

[57] 在整个方法流程过程中，如果执行步骤S206和步骤S207，那么还可以执行步骤S210，即随时监控处于空闲状态的虚拟服务器的数量是否达到第二预定数量，当处于空闲状态的虚拟服务器的数量达到第二预定数量时，说明很多虚拟服务器闲置着，没有被调度使用，此时如果云端虚拟服务器集群中虚拟服务器的数量大于第一预定数量，那么可以回收部分虚拟服务器，使得云端虚拟服务器集群中虚拟服务器的数量在第一预定数量内，或者稍稍大于第一预定数量。回收虚拟服务器时需要逐步回收，不能突然回收大量处于空闲状态的虚拟机，因为根据统计学上的规律，使用高峰的过去不会戛然而止，是个逐渐的过程，所以突然回收大量处于空闲状态的虚拟机，必然造成下一个时间点，需要重新创建大量的虚拟机，短时间内重复的创建和回收必然造成资源的浪费。所以合理的做法应该是逐渐的回收处于空闲状态的虚拟机，而不是一次回收大量处于空闲状态的虚拟机。

[58] 需要说明的是，步骤S208、步骤S209以及步骤S210在整个方法流程中随时可以根据需要或预先的设置进行执行，因此，步骤S208、步骤S209以及步骤S210在整个方法流程中没有固定的先后顺序，其中，步骤S210执行的条件是步骤S206和步骤S207预先执行。

[59] 另外，在实际应用中，可以根据具体应用环境和条件，对上述步骤进行相应的增减，在此不再进行赘叙。

[60] 本发明实施方式根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务

器；若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。由于不考虑虚拟服务器是否被其它用户释放，只要有处于空闲状态的虚拟服务器时，直接调用处于空闲状态的虚拟服务器以执行用户请求，通过这种方式，可以更好地提高虚拟服务器的使用率，从而节约资源。另外，提前创建并初始化虚拟服务器，需要调用的时候，直接可以调用，从而缩短等待时间，提高用户体验；监控虚拟服务器的状态，可以随时获得能够被调度的虚拟服务器，并且在处于空闲状态的虚拟服务器的数量达到第二预定数量且虚拟服务器的总数量大于第一预定数量时，逐渐回收虚拟服务器，可以避免短时间内重复的创建和回收造成的资源浪费。

[61] 参阅图5，图5是本发明云端虚拟服务器的调度装置一实施方式的结构示意图，该装置包括：接收模块101、查看模块102以及第一调度模块103。

[62] 需要说明的是，本实施方式的装置可以执行图1至图4中的步骤。

[63] 接收模块101用于接收用户请求。

[64] 用户请求是指用户请求使用虚拟服务器，以使得虚拟服务器向用户提供服务。

[65] 查看模块102用于根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，云端虚拟服务器集群包括第一预定数量的虚拟服务器。

[66] 云端虚拟服务器集群是包括第一预定数量的虚拟服务器的集群，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，例如：被其它用户释放当前处于空闲状态的虚拟服务器；初始化后一直处于空闲状态的虚拟服务器；其它用户没有释放、但当前没有被使用的虚拟服务器等等。

[67] 在接收到用户请求后，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，以便于调度虚拟服务器。

[68] 第一调度模块103用于在云端虚拟服务器集群中有处于空闲状态的虚拟服务器时，调度处于空闲状态的虚拟服务器以执行用户请求。

[69] 如果云端虚拟服务器集群中有处于空闲状态的虚拟服务器，那么这些虚拟服务器可以调度来执行用户请求，从而有效地提高虚拟服务器的使用率。

[70] 本发明实施方式根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状

态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器；若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。由于不考虑虚拟服务器是否被其它用户释放，只要在有处于空闲状态的虚拟服务器时，直接调用处于空闲状态的虚拟服务器以执行用户请求，通过这种方式，可以更好地提高虚拟服务器的使用率，从而节约资源。

[71] 参阅图6至图8，图6至图8是本发明云端虚拟服务器的调度装置三个实施方式的结构示意图，本三个实施方式的装置和图5的装置基本相同，相同之处请参见图5和相关的文字说明，不同之处具体请参见如下内容：

[72] 参阅图6，该装置包括：接收模块201、查看模块202、第一调度模块203、第一创建模块204、第一初始化模块205。

[73] 需要说明的是，图6实施方式的装置可以执行图2中的步骤。

[74] 第一创建模块204用于创建所述云端虚拟服务器集群。

[75] 第一初始化模块205用于初始化所述云端虚拟服务器集群中的虚拟服务器。

[76] 预先创建云端虚拟服务器集群，并初始化虚拟服务器，构成一批处于空闲状态的虚拟服务器，等待请求调用。由于虚拟服务器提前已经创建并初始化了，需要调用的时候，直接可以调用，从而可以避免在调用时再创建并初始化虚拟服务器所需要等待的时间，通过这种方式，可以提高用户体验。

[77] 接收模块201用于接收用户请求。

[78] 查看模块202用于根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，云端虚拟服务器集群包括第一预定数量的虚拟服务器。

[79] 第一调度模块203用于在云端虚拟服务器集群中有处于空闲状态的虚拟服务器时，调度处于空闲状态的虚拟服务器以执行用户请求。

[80] 监控模块209用于监控云端虚拟服务器集群中虚拟服务器的状态。

[81] 标记模块210用于在监控到虚拟服务器当前没有被使用时，将虚拟服务器的状态标记为空闲状态。

[82] 在整个装置中，随时监控云端虚拟服务器集群中虚拟服务器的状态，如果监控

到存在当前没有被使用的虚拟服务器，则将虚拟服务器的状态标记为空闲状态，其中，当前没有被使用的虚拟服务器，例如可以是虚拟服务器被释放、初始化后没有被使用、或当前暂时没有被使用等等。

[83] 参阅图7，该装置还包括：第二创建模块206、第二初始化模块207、第二调度模块208。

[84] 需要说明的是，图7实施方式的装置可以执行图3中的步骤。

[85] 第二创建模块206用于在云端虚拟服务器集群中没有处于空闲状态的虚拟服务器时，创建 n 个虚拟服务器， n 为用户请求的个数。

[86] 第二初始化模块207用于初始化 n 个虚拟服务器。

[87] 第二调度模块208用于调度 n 个虚拟服务器以执行用户请求。

[88] 如果云端虚拟服务器集群中所有的虚拟服务器都已经被用户使用，没有处于空闲状态的虚拟机，那么根据用户请求的个数另外创建 n 个虚拟服务器并初始化，然后调度这 n 个虚拟服务器以执行用户请求，为用户提供服务。

[89] 参阅图8，装置还包括：回收模块211。

[90] 需要说明的是，图8实施方式的装置可以执行图4中的步骤。

[91] 回收模块211用于在处于空闲状态的虚拟服务器的数量达到第二预定数量，且云端虚拟服务器集群中虚拟服务器的数量大于第一预定数量时，逐渐回收虚拟服务器，以使得云端虚拟服务器集群中虚拟服务器的数量大于或等于第一预定数量。

[92] 在整个装置中，如果装置包括：第二创建模块206、第二初始化模块207、第二调度模块208，那么随时监控处于空闲状态的虚拟服务器的数量是否达到第二预定数量，当处于空闲状态的虚拟服务器的数量达到第二预定数量时，说明很多虚拟服务器闲置着，没有被调度使用，此时如果云端虚拟服务器集群中虚拟服务器的数量大于第一预定数量，那么可以回收部分虚拟服务器，使得云端虚拟服务器集群中虚拟服务器的数量在第一预定数量内，或者稍稍大于第一预定数量。回收虚拟服务器时需要逐步回收，不能突然回收大量处于空闲状态的虚拟机。

[93] 另外，在实际应用中，可以根据具体应用环境和条件，对上述模块或单元进行

相应的增减，在此不再进行赘叙。

- [94] 本发明实施方式根据用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器；若云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度处于空闲状态的虚拟服务器以执行用户请求。由于不考虑虚拟服务器是否被其它用户释放，只要在有处于空闲状态的虚拟服务器时，直接调用处于空闲状态的虚拟服务器以执行用户请求，通过这种方式，可以更好地提高虚拟服务器的使用率，从而节约资源。另外，提前创建并初始化虚拟服务器，需要调用的时候，直接可以调用，从而缩短等待时间，提高用户体验；监控虚拟服务器的状态，可以随时获得能够被调度的虚拟服务器，并且在处于空闲状态的虚拟服务器的数量达到第二预定数量且虚拟服务器的总数量大于第一预定数量时，逐渐回收虚拟服务器，可以避免短时间内重复的创建和回收造成的资源浪费。

- [95] 以上所述仅为本发明的实施方式，并非因此限制本发明的专利范围，凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本发明的专利保护范围内。

权利要求书

- [权利要求 1] 一种云端虚拟服务器的调度方法，其特征在于，包括：
- 创建所述云端虚拟服务器集群；
- 初始化所述云端虚拟服务器集群中的虚拟服务器；
- 接收用户请求；
- 根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定数量的虚拟服务器；
- 若所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度所述处于空闲状态的虚拟服务器以执行所述用户请求；
- 若所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器，则创建并初始化 n 个虚拟服务器，所述 n 为所述用户请求的个数；
- 调度所述 n 个虚拟服务器以执行所述用户请求。
- [权利要求 2] 根据权利要求1所述的方法，其特征在于，所述方法还包括：
- 若所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量，则逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。
- [权利要求 3] 根据权利要求1所述的方法，其特征在于，所述方法还包括：
- 监控所述云端虚拟服务器集群中虚拟服务器的状态；
- 若监控到所述虚拟服务器当前没有被使用，则将所述虚拟服务器的状态标记为空闲状态。
- [权利要求 4] 一种云端虚拟服务器的调度方法，其特征在于，包括：
- 接收用户请求；
- 根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定

数量的虚拟服务器；

若所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器，则调度所述处于空闲状态的虚拟服务器以执行所述用户请求。

[权利要求 5] 根据权利要求4所述的方法，其特征在于，所述接收用户请求的步骤之前，包括：

创建所述云端虚拟服务器集群；

初始化所述云端虚拟服务器集群中的虚拟服务器。

[权利要求 6] 根据权利要求4所述的方法，其特征在于，根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器的步骤之后，包括：

若所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器，则创建并初始化n个虚拟服务器，所述n为所述用户请求的个数；调度所述n个虚拟服务器以执行所述用户请求。

[权利要求 7] 根据权利要求5所述的方法，其特征在于，所述方法还包括：
若所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量，则逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。

[权利要求 8] 根据权利要求4所述的方法，其特征在于，所述方法还包括：
监控所述云端虚拟服务器集群中虚拟服务器的状态；
若监控到所述虚拟服务器当前没有被使用，则将所述虚拟服务器的状态标记为空闲状态。

[权利要求 9] 一种云端虚拟服务器的调度装置，其特征在于，所述装置包括：
接收模块，用于接收用户请求；
查看模块，用于根据所述用户请求，查看云端虚拟服务器集群中是否有处于空闲状态的虚拟服务器，所述处于空闲状态的虚拟服务器是指当前没有被使用的虚拟服务器，所述云端虚拟服务器集群包括第一预定数量的虚拟服务器；

第一调度模块，用于在所述云端虚拟服务器集群中有处于空闲状态的虚拟服务器时，调度所述处于空闲状态的虚拟服务器以执行所述用户请求。

[权利要求 10] 根据权利要求9所述的装置，其特征在于，所述装置还包括：
第一创建模块，用于创建所述云端虚拟服务器集群；
第一初始化模块，用于初始化所述云端虚拟服务器集群中的虚拟服务器。

[权利要求 11] 根据权利要求9所述的装置，其特征在于，所述装置还包括：
第二创建模块，用于在所述云端虚拟服务器集群中没有处于空闲状态的虚拟服务器时，创建n个虚拟服务器，所述n为所述用户请求的个数；
第二初始化模块，用于初始化所述n个虚拟服务器；
第二调度模块，用于调度所述n个虚拟服务器以执行所述用户请求。

[权利要求 12] 根据权利要求11所述的装置，其特征在于，所述装置还包括回收模块，所述回收模块用于在所述处于空闲状态的虚拟服务器的数量达到第二预定数量，且所述云端虚拟服务器集群中虚拟服务器的数量大于所述第一预定数量时，逐渐回收所述虚拟服务器，以使得所述云端虚拟服务器集群中虚拟服务器的数量大于或等于所述第一预定数量。

[权利要求 13] 根据权利要求9所述的装置，其特征在于，所述装置还包括：
监控模块，用于监控所述云端虚拟服务器集群中虚拟服务器的状态；
标记模块，用于在监控到所述虚拟服务器当前没有被使用时，将所述虚拟服务器的状态标记为空闲状态。

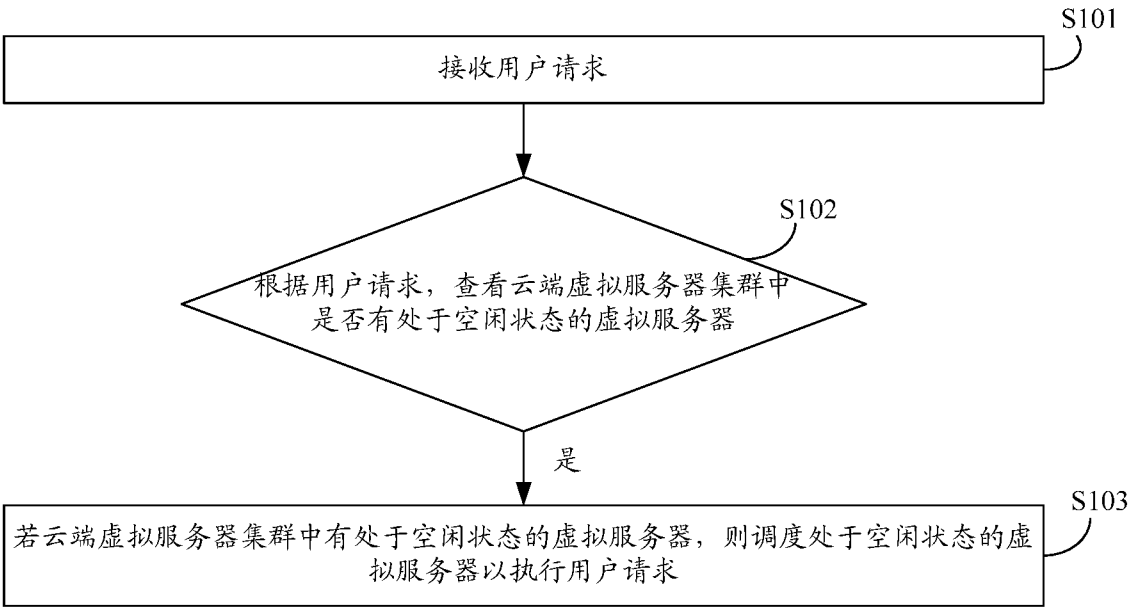


图 1

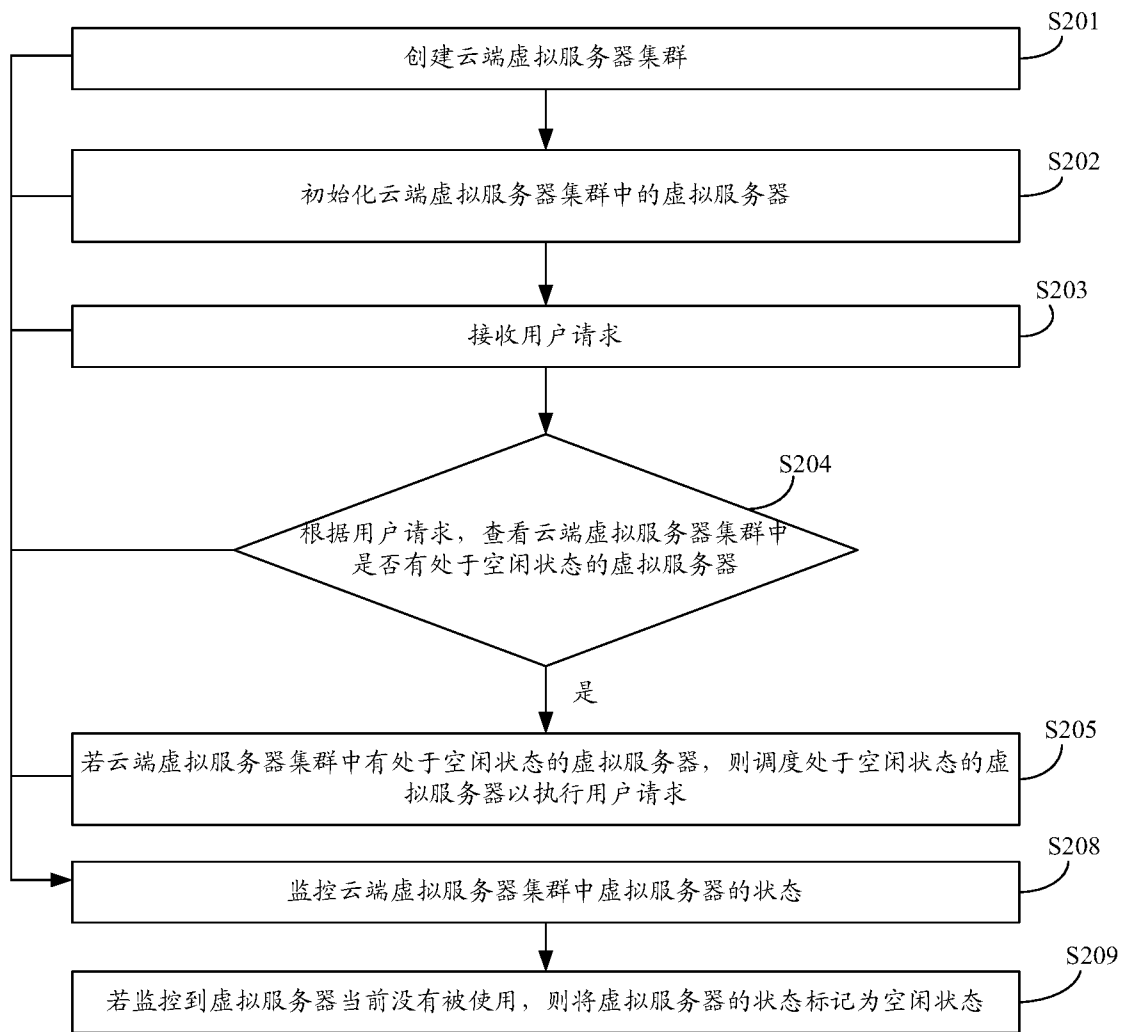


图 2

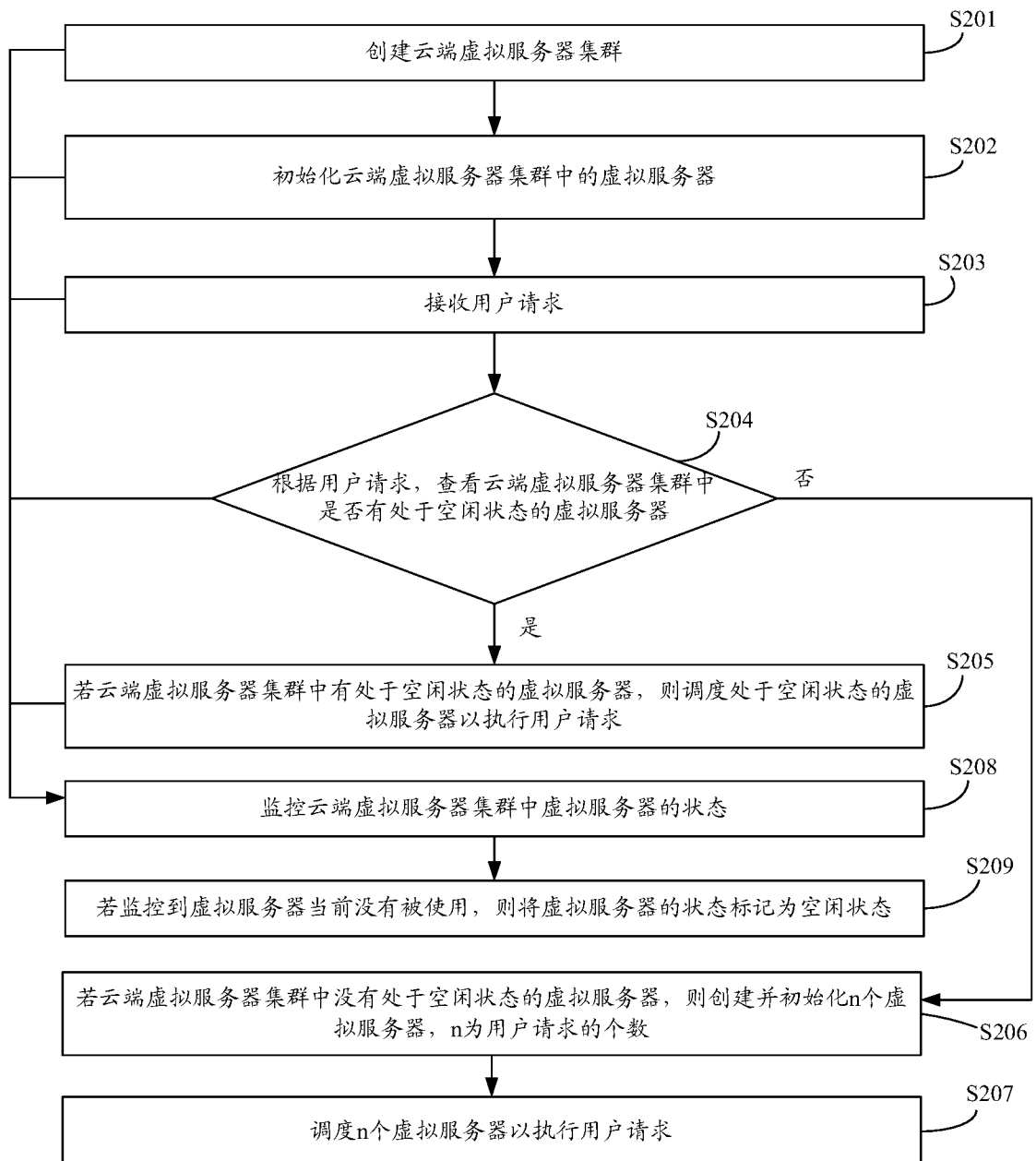


图 3

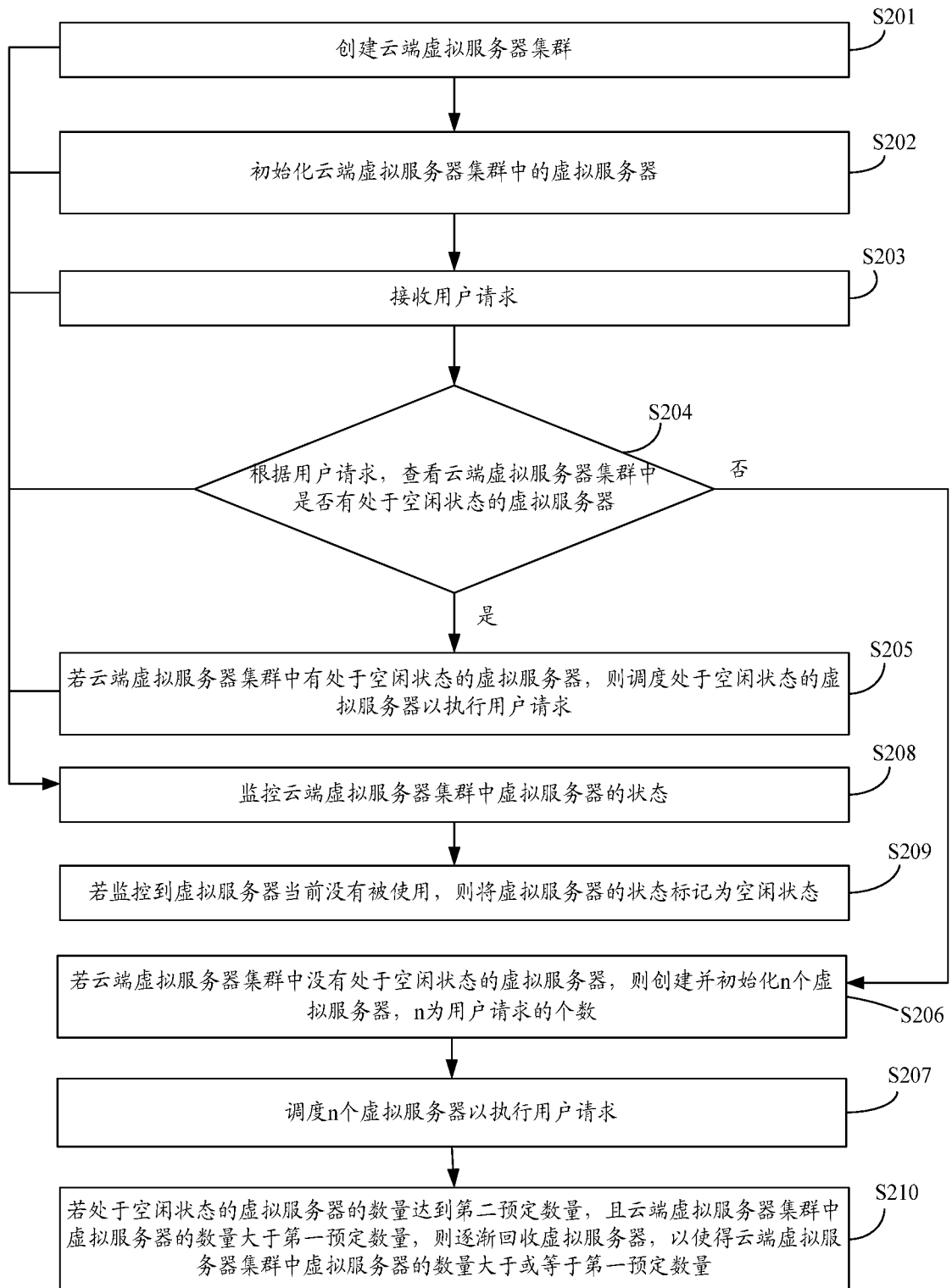


图 4

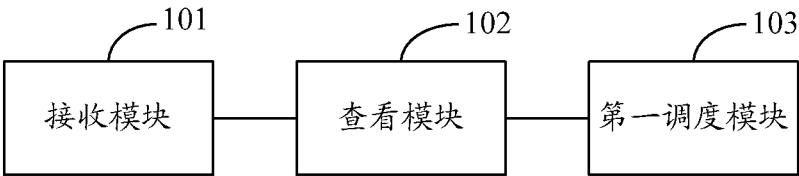


图 5

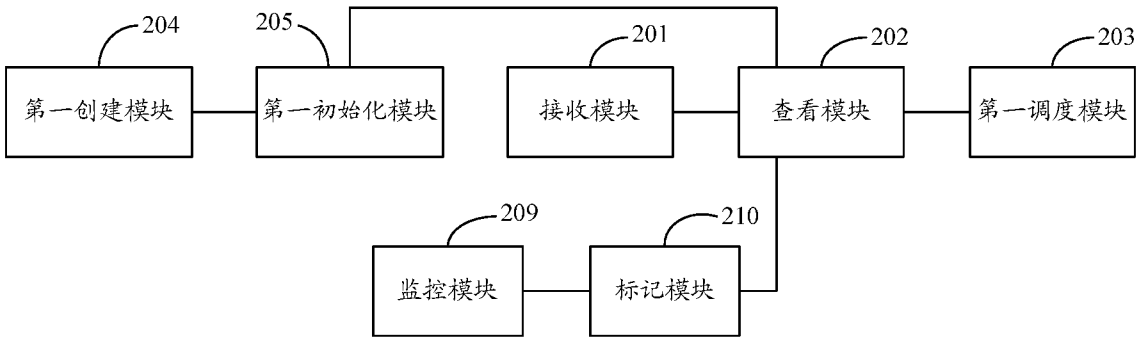


图 6

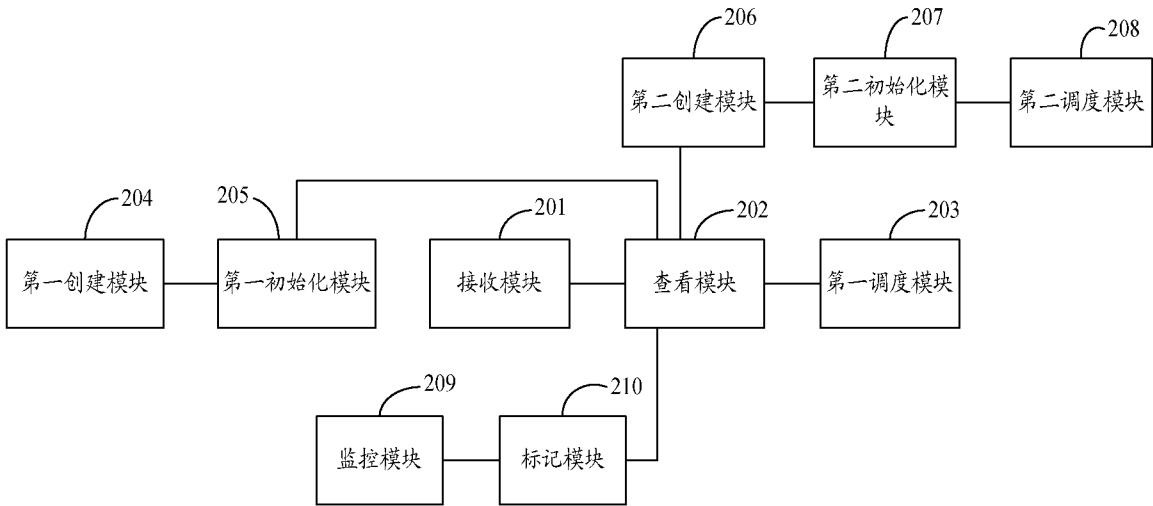


图 7

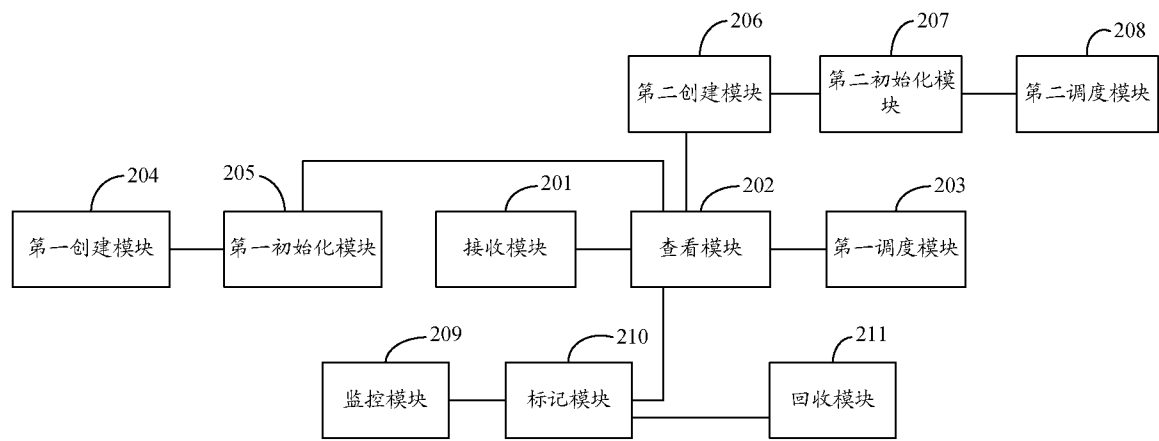


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/074872

A. CLASSIFICATION OF SUBJECT MATTER

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)

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)

CPRSABS, CNKI, CNTXT, DWPI, VEN: server, group, cluster, virtual, cloud, idle, new, creat+, request

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104168310 A (SHENZHEN TINNO WIRELESS TECHNOLOGY CO., LTD.), 26 November 2014 (26.11.2014), claims 1-10	1-13
Y	CN 102594912 A (BEIJING KALENDS TECHNOLOGY CO., LTD.), 18 July 2012 (18.07.2012), see description, paragraphs [0007], [0042]-[0044], [0049], [0052] and [0058]	1-13
Y	CN 103428008 A (ZHEJIANG UNIVERSITY), 04 December 2013 (04.12.2013), see description, paragraphs [0006], [0010]-[0011], [0016], [0035]-[0037], [0045] and [0050]-[0053], and figure 1	1-13
A	CN 102868763 A (SHANDONG COMPUTER SCIENCE CENTER), 09 January 2013 (09.01.2013), see the whole document	1-13

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 08 June 2015 (08.06.2015)	Date of mailing of the international search report 26 June 2015 (26.06.2015)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer LI, Xiaoqian Telephone No.: (86-10) 62089456

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/074872

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104168310 A	26 November 2014	None	
CN 102594912 A	18 July 2012	None	
CN 103428008 A	04 December 2013	None	
CN 102868763 A	09 January 2013	None	

A. 主题的分类		
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)		
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))		
CPRSABS, CNKI, CNTXT, DWPI, VEN: 服务器 集群 虚拟 云 空闲 新 创建 请求; server, group, cluster, virtual, cloud, idle, new, creat+, request		
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 104168310 A (深圳天珑无线科技有限公司) 2014年 11月 26日 (2014 - 11 - 26) 权利要求1-10	1-13
Y	CN 102594912 A (北京昆仑万维科技股份有限公司) 2012年 7月 18日 (2012 - 07 - 18) 参见说明书第[0007]段, [0042]-[0044]段, [0049]段, 第[0052]段, [0058]段	1-13
Y	CN 103428008 A (浙江大学) 2013年 12月 4日 (2013 - 12 - 04) 参见说明书第[0006]段, [0010]-[0011]段, [0016]段, 第[0035]-[0037]段, [0045]段, 第[0050]-[0053]段, 图1	1-13
A	CN 102868763 A (山东省计算中心) 2013年 1月 9日 (2013 - 01 - 09) 参见全文	1-13
☐ 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		国际检索报告邮寄日期
2015年 6月 8日		2015年 6月 26日
ISA/CN的名称和邮寄地址		授权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国		李晓茜
传真号 (86-10)62019451		电话号码 (86-10)62089456

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/074872

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	104168310	A	2014年 11月 26日	无	
CN	102594912	A	2012年 7月 18日	无	
CN	103428008	A	2013年 12月 4日	无	
CN	102868763	A	2013年 1月 9日	无	