(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104301257 B
(45)授权公告日 2018.01.02

(21)申请号 201410476386.3

(22)申请日 2014.09.17

(65)同一申请的已公布的文献号
申请公布号 CN 104301257 A

(43)申请公布日 2015.01.21

(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 甘嘉栋

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291
代理人 冯艳莲

(51)Int.Cl.
H04L 12/917(2013.01)

(56)对比文件
CN 101035013A A,2007.09.12,
US 2004244001 A,2004.12.02,
CN 103713955 A,2014.04.09,

审查员 马旗超

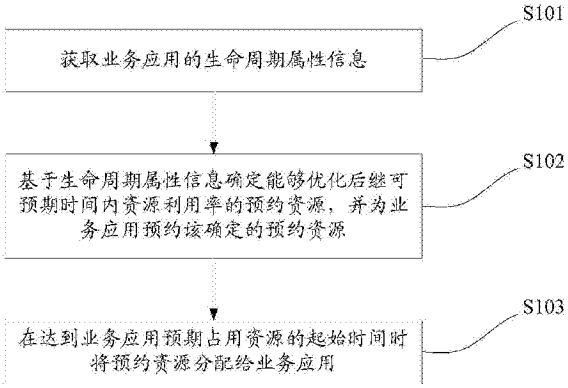
权利要求书2页 说明书9页 附图4页

(54)发明名称

一种资源分配方法、装置及设备

(57)摘要

本发明公开了一种资源分配方法、装置及设备，本发明中获取业务应用的生命周期属性信息，所述生命周期属性信息反映所述业务应用预期占用资源的时间信息；基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为所述业务应用预约所述预约资源；在达到所述业务应用预期占用资源的起始时间时，将所述预约资源分配给所述业务应用。通过本发明能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。



CN 104301257 B

1.一种资源分配方法,其特征在于,包括:

获取业务应用的生命周期属性信息,所述生命周期属性信息反映所述业务应用预期占用资源的时间信息;

确定预设资源集合中已启用基础设施资源实体,并确定所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间,选择在后继可预期时间内所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述生命周期属性信息中所述业务应用预期占用资源时间相近、且所述已启用基础设施资源实体上未分配资源满足所述业务应用所需资源条件的资源,作为预约资源,并为所述业务应用预约所述预约资源,其中,所述时间相近,是指所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述生命周期属性信息中所述业务应用预期占用资源时间的差值,小于等于约定值;

在达到所述业务应用预期占用资源的起始时间时,将所述预约资源分配给所述业务应用。

2.如权利要求1所述的方法,其特征在于,为所述业务应用预约所述预约资源之前,所述方法还包括:

在后继可预期时间内资源利用率满足进一步优化条件的前提下,对已预约资源进行重预约。

3.如权利要求1或2所述的方法,其特征在于,所述获取业务应用的生命周期属性信息,包括:

调用业务资源请求接口携带业务应用生命周期属性信息功能;

通过调用的业务资源请求接口携带业务应用生命周期属性信息功能,获取业务应用的生命周期属性信息。

4.如权利要求1所述的方法,其特征在于,基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源,包括:

在设定的部分资源集合中,基于所述生命周期属性信息,优先选择能够优化后继可预期时间内资源利用率的预约资源。

5.一种资源分配装置,其特征在于,包括:

获取单元,用于获取业务应用的生命周期属性信息,所述生命周期属性信息反映所述业务应用预期占用资源的时间信息;

预约单元,用于确定预设资源集合中已启用基础设施资源实体,并确定所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间;选择在后继可预期时间内所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述获取单元获取的所述生命周期属性信息中所述业务应用预期占用资源时间相近、且所述已启用基础设施资源实体上未分配资源满足所述业务应用所需资源条件的资源,作为预约资源,并为所述业务应用预约所述预约资源,其中,所述时间相近,是指所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述生命周期属性信息中所述业务应用预期占用资源时间的差值,小于等于约定值;

分配单元,用于在达到所述业务应用预期占用资源的起始时间时,将所述预约单元预约的预约资源分配给所述业务应用。

6.如权利要求5所述的装置，其特征在于，所述装置还包括重预约单元，用于在后继可预期时间内资源利用率满足进一步优化条件的前提下，在所述预约单元为所述业务应用预约所述预约资源之前，对已预约资源进行重预约。

7.如权利要求5或6所述的装置，其特征在于，所述获取单元，具体用于按如下方式获取业务应用的生命周期属性信息：

调用业务资源请求接口携带业务应用生命周期属性信息功能；

通过调用的业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

8.如权利要求5所述的装置，其特征在于，所述预约单元，具体用于按如下方式基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源：

在设定的部分资源集合中，基于所述生命周期属性信息，优先选择能够优化后继可预期时间内资源利用率的预约资源。

9.一种资源分配设备，其特征在于，包括通信接口、存储器和处理器，其中：

所述存储器，用于存储所述处理器执行的程序代码；

所述处理器，用于调用所述存储器存储的程序代码，实现如下功能：

通过通信接口获取业务应用的生命周期属性信息，所述生命周期属性信息反映业务应用预期占用资源的时间信息，确定预设资源集合中已启用基础设施资源实体，并确定所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间，选择在后继可预期时间内所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述生命周期属性信息中所述业务应用预期占用资源时间相近、且所述已启用基础设施资源实体上未分配资源满足所述业务应用所需资源条件的资源，作为预约资源，并为所述业务应用预约所述预约资源，其中，所述时间相近，是指所述已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间与所述生命周期属性信息中所述业务应用预期占用资源时间的差值，小于等于约定值，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给所述业务应用。

10.如权利要求9所述的设备，其特征在于，所述处理器，还用于：

在后继可预期时间内资源利用率满足进一步优化条件的前提下，在为所述业务应用预约所述预约资源之前，对已预约资源进行重预约。

一种资源分配方法、装置及设备

技术领域

[0001] 本发明涉及资源分配技术领域，尤其涉及一种资源分配方法、装置及设备。

背景技术

[0002] 随着云概念的推广，业务应用层与管理层的分层架构广泛应用于要求资源动态分配的解决方案中。

[0003] 业务应用向基础设施管理系统发送业务资源请求，以申请处理特定业务所需的基础设施资源以运行其实例。基础设施管理系统针对业务应用发送的业务资源请求中请求的资源规格(资源规格例如可以是CPU、内存和网络连接等物理规格，也可以是其它一些质量属性要求)，以及系统内可用的资源分布情况，进行综合分析并进行资源的分配。

[0004] 目前，基础设施管理系统进行资源分配时，在满足资源规格的前提下，优先选取已启用的基础设施资源实体进行资源分配，尽量不使用未启用的基础设施资源实体，并在选取的基础设施资源实体中优先选择剩余资源最少的基础设施资源实体进行资源分配，在业务应用占用资源时间达到后，资源被释放。

[0005] 上述进行资源分配的过程在一定程度上能够提高资源利用率，然而，不同业务应用占用资源的时间不尽相同，例如图1中基础设施资源实体1上的资源(1/2)在短时间后被释放，然而基础设施资源实体1上的资源(1/4)和基础设施资源实体2上的资源(1/2)需要长期被占用，虽然资源(1/4)占用基础设施资源实体的资源较少，但是还是需要运行基础设施资源实体1，故需要同时运行基础设施资源实体1和基础设施资源实体2，基础设施管理系统中的资源利用率很难在长时间内保持优化的状态。

发明内容

[0006] 本发明实施例提供一种资源分配方法、装置及设备，以优化基础设施管理系统中的资源利用率。

[0007] 第一方面，提供一种资源分配方法，包括：

[0008] 获取业务应用的生命周期属性信息，所述生命周期属性信息反映所述业务应用预期占用资源的时间信息；

[0009] 基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为所述业务应用预约所述预约资源；

[0010] 在达到所述业务应用预期占用资源的起始时间时，将所述预约资源分配给所述业务应用。

[0011] 结合第一方面，在第一种实现方式中，为所述业务应用预约所述预约资源之前，所述方法还包括：

[0012] 在后继可预期时间内资源利用率满足进一步优化条件的前提下，对已预约资源进行重预约。

[0013] 结合第一方面或者第一方面的第一种实现方式，在第二种实现方式中，所述获取

业务应用的生命周期属性信息，包括：

[0014] 调用业务资源请求接口携带业务应用生命周期属性信息功能；

[0015] 通过调用的业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

[0016] 结合第一方面的上述任一种实现方式，在第三种实现方式中，基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，包括：

[0017] 在设定的部分资源集合中，基于所述生命周期属性信息，优先选择能够优化后继可预期时间内资源利用率的预约资源。

[0018] 第二方面，提供一种资源分配装置，包括获取单元、预约单元和分配单元，其中：

[0019] 所述获取单元，用于获取业务应用的生命周期属性信息，所述生命周期属性信息反映所述业务应用预期占用资源的时间信息；

[0020] 所述预约单元，用于基于所述获取单元获取的所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为所述业务应用预约所述预约资源；

[0021] 所述分配单元，用于在达到所述业务应用预期占用资源的起始时间时，将所述预约单元预约的预约资源分配给所述业务应用。

[0022] 结合第二方面，在第一种实现方式中，所述装置还包括重预约单元，其中：

[0023] 所述重预约单元，用于在后继可预期时间内资源利用率满足进一步优化条件的前提下，在所述预约单元为所述业务应用预约所述预约资源之前，对已预约资源进行重预约。

[0024] 结合第二方面或者第二方面的第一种实现方式，在第二种实现方式中，所述获取单元，具体用于按如下方式获取业务应用的生命周期属性信息：

[0025] 调用业务资源请求接口携带业务应用生命周期属性信息功能；

[0026] 通过调用的业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

[0027] 结合第二方面的任一种实现方式，在第三种实现方式中，所述预约单元，具体用于按如下方式基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源：

[0028] 在设定的部分资源集合中，基于所述生命周期属性信息，优先选择能够优化后继可预期时间内资源利用率的预约资源。

[0029] 第三方面，提供一种资源分配设备，包括通信接口、存储器和处理器，其中：

[0030] 所述存储器，用于存储所述处理器执行的程序代码；

[0031] 所述处理器，用于调用所述存储器存储的程序代码，实现如下功能：

[0032] 通过通信接口获取业务应用的生命周期属性信息，所述生命周期属性信息反映业务应用预期占用资源的时间信息，并基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为所述业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给所述业务应用。

[0033] 结合第三方面，在第一种实现方式中，所述处理器，还用于：

[0034] 在后继可预期时间内资源利用率满足进一步优化条件的前提下，在为所述业务应用预约所述预约资源之前，对已预约资源进行重预约。

[0035] 本发明实施例提供的资源分配方法、装置及设备，获取业务应用的生命周期属性

信息，该生命周期属性信息中能够反映业务应用预期占用资源的时间信息，故可预先获知后继可预期时间内资源被占用的时间。然后基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给业务应用，能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。

附图说明

[0036] 图1为现有技术中资源分配示意图；

[0037] 图2为本发明实施例可应用的系统架构图；

[0038] 图3为为本发明实施例提供的资源分配设备构成示意图；

[0039] 图4为本发明实施例提供的资源分配方法流程图；

[0040] 图5A为现有资源分配过程示意图；

[0041] 图5B为本发明实施例提供的资源分配过程示意图；

[0042] 图6A为本发明实施例提供的资源分配装置构成示意图；

[0043] 图6B为本发明实施例提供的另一资源分配装置构成示意图。

具体实施方式

[0044] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，并不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0045] 图2所示为本发明实施例提供的资源分配方法可应用的系统架构示意图，如图2所示，基础设施管理系统的业务资源请求接口具有获取业务应用的业务资源请求功能，基础设施管理系统通过资源请求接口获取业务应用的业务资源请求，资源分配与管理模块基于获取的业务资源请求，在资源池中选择资源为业务应用分配。本发明实施例中可扩展业务资源请求接口的功能，使其具有获取业务应用的生命周期属性信息的功能，获取业务应用的生命周期属性信息，生命周期属性信息反映业务应用预期占用资源的时间信息。资源分配与管理模块基于获取的生命周期属性信息进行资源分配，以达到在预期时间内对基础设施管理系统中的资源利用率进行优化的目的。

[0046] 需要说明的是，本发明实施例中提供的资源分配方法并不局限于图2所示的系统架构图，例如本发明实施例中还可应用于资源分配功能与资源管理功能通过两个单独的模块实现。

[0047] 本发明实施例提供一种资源分配设备300，如图3所示该资源分配设备300包括通信接口301、存储器302和处理器303。当然根据实际情况，本发明实施例中提供的资源分配设备300可能还包括通信总线，本发明实施例不做限定。

[0048] 本发明实施例中通信接口301，使用诸如收发器一类的装置，与其他设备或通信网络通信，如以太网，无线接入网(RAN)，无线局域网(Wireless Local Area Networks，WLAN)等。

[0049] 本发明实施例中存储器302，可以是只读存储器(read-only memory，ROM)或可存

储静态信息和指令的其他类型的静态存储设备，随机存取存储器(random access memory，RAM)或者可存储信息和指令的其他类型的动态存储设备，也可以是电可擦可编程只读存储器(Electrically Erasable Programmable Read-Only Memory，EEPROM)、只读光盘(Compact Disc Read-Only Memory，CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。

[0050] 本发明实施例中处理器303，可以是一个通用中央处理器(CPU)，微处理器，特定应用集成电路(application-specific integrated circuit，ASIC)，或一个或多个用于控制本发明方案程序执行的集成电路。

[0051] 本发明实施例中存储器302，用于存储处理器303执行的程序代码。

[0052] 处理器303，用于调用存储器302存储的程序代码，实现如下功能：

[0053] 通过通信接口301获取业务应用的生命周期属性信息，该生命周期属性信息反映业务应用预期占用资源的时间信息。

[0054] 基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为业务应用预约该确定的预约资源；在达到业务应用预期占用资源的起始时间时，将确定的预约资源分配给业务应用。

[0055] 在一种实现方式中，本发明实施例中处理器303还用于：

[0056] 在后继可预期时间内资源利用率满足进一步优化条件的前提下，在为所述业务应用预约所述预约资源之前，对已预约资源进行重预约；其中，进一步优化条件是指可以对后继可预期时间内的资源分配进行调优，例如可以使整体资源利用率最高。

[0057] 具体的，本发明实施例中通信接口301可以是业务资源请求接口，该业务资源请求接口具有携带业务应用生命周期属性信息的扩展功能，处理器303可调用业务资源请求接口携带业务应用生命周期属性信息功能；通过调用的业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

[0058] 具体的，本发明实施例中处理器303可采用如下方式基于所述生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源：

[0059] A：确定设定资源集合中已启用基础设施资源实体。

[0060] 本发明实施例中设定资源集合可以是基础设施管理系统中的全部资源，也可以是基础设施管理系统中的部分资源。

[0061] 本发明实施例中优选在基础设施管理系统中设定部分资源，进行基于生命周期属性信息进行资源分配，以使基础设施管理系统对外可同时支持原有分配机制(即不考虑资源占用的生命周期)，以及基于生命周期信息进行资源分配的机制。并且本发明实施例中设定部分资源进行基于生命周期信息进行资源分配，可避免原有资源分配机制的不确定性对基于生命周期属性信息进行资源分配的影响。

[0062] B：在设定的部分资源集合中，基于生命周期属性信息，优先选择能够优化后继可预期时间内资源利用率的预约资源。

[0063] 本发明实施例中，例如可以选择在后继可预期时间内资源被占用时间与生命周期属性信息中预期占用资源时间相近、且未分配资源满足业务应用所需资源条件的已启用基

础设施资源实体上的资源，作为预约资源。

[0064] 本发明实施例上述提供的资源分配设备300，可以是基础设施管理系统，也可以是基础设施管理系统中的部件，本发明实施例不做限定。

[0065] 本发明实施例提供的资源分配设备300，获取业务应用的生命周期属性信息，该生命周期属性信息中能够反映业务应用预期占用资源的时间信息，故可预先获知后继可预期时间内资源被占用的时间。然后基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给业务应用，能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。

[0066] 本发明实施例以下将对资源分配方法的实现方法进行详细说明。

[0067] 图4所示为本发明实施例提供的资源分配方法实现流程图，图4所示方法的执行主体例如可以是基础设施管理系统，也可以是基础设施管理系统的部件，本发明实施例不做限定。如图4所示，该方法包括：

[0068] S101：获取业务应用的生命周期属性信息。

[0069] 本发明实施例中生命周期属性信息反映业务应用预期占用资源的时间信息。

[0070] 本发明实施例中可扩展资源请求接口具有携带业务应用生命周期属性信息的功能，通过调用业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

[0071] 本发明实施例中生命周期属性信息例如可采用如下表1所示的结构进行表示，

[0072]

信息项	可选	说明
开始时间	是	资源预期被占用的开始时间。缺省表示将马上使用该资源
结束时间	是	计划结束占用该资源的时间。缺省表示不确定结束时间
重复性周期	是	预期该资源会被周期性占用的时间，例如每天的 4:00~22:00

[0073] 表1

[0074] 需要说明的是，本发明实施例中生命周期属性信息并不限于表1所列举的若干例子，还应包括任何可用于描述资源预期被占用的时间的方式。例如还可以是一个时间列表用于列出预期的各个不规则(非重复性周期)的使用时间。

[0075] S102：基于生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为业务应用预约该确定的预约资源。

[0076] 本发明实施例中为业务应用预约确定的预约资源，该被预约的预约资源没有真正被分配并占用，但是该预约资源在被预约的时间内，不能再被预约或分配给除当前被预约的业务应用以外的其它业务应用。

[0077] 本发明实施例中基于生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源过程中，可采用以减少资源碎片、启用的基础设施资源实体数目最少、长时间处于优化状态中的至少一个为优化目标，进行预约资源的确定。

[0078] 例如，本发明实施例中基于生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源的过程，可采用如下方式：

[0079] 确定设定资源集合中已启用基础设施资源实体，并确定已启用基础设施资源实体上已分配资源和已预约资源预期占用资源的时间。

[0080] 优先选择在后继可预期时间内资源被占用时间与所述生命周期属性信息中预期占用资源时间相近、且未分配资源满足业务应用所需资源条件的已启用基础设施资源实体上的资源，作为预约资源。

[0081] S103：在达到业务应用预期占用资源的起始时间时，将预约资源分配给业务应用。

[0082] 本发明实施例提供的资源分配方法，获取业务应用的生命周期属性信息，该生命周期属性信息中能够反映业务应用预期占用资源的时间信息，故可预先获知后继可预期时间内资源被占用的时间。然后基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给业务应用，能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。

[0083] 本发明实施例中基于业务应用的生命周期属性信息为业务应用进行资源预约的预约资源，未被真正分配并占用，故本发明实施例中通过对预约信息的更改，即可实现将另一基础设施资源实体的资源预约给业务应用，而不涉及任何资源迁移技术，技术实现简单易行。

[0084] 可选的，本发明实施例中为了进一步达到基础设施管理系统中资源长时间内处于优化状态，即资源利用率满足进一步优化条件的前提下，可对已预约资源进行重预约，其中，进一步优化条件，例如可以是能够减少资源碎片、能够使启用的基础设施资源实体数目最少或者能够使基础设施管理系统中的基础设施资源实体能够在长时间内处于优化状态。

[0085] 例如本发明实施例中可在当前基础设施资源实体上已预约资源在后继可预期时间内资源占用时间与当前业务应用预期占用资源的时间相近、且当前基础设施资源实体的未分配资源不满足当前业务应用所需资源的情况下，对当前基础设施资源实体上已预约资源进行重预约，将已预约的资源和当前应用所需的资源分配到同一基础设施资源实体上，优化资源利用率。

[0086] 本发明实施例以下将结合实际应用对上述进行资源分配的过程进行详细说明。

[0087] 本发明实施例中假设基础设施管理系统中的基础设施资源实体的资源规格相同，资源总量记为1S，3个基础设施资源实体对应的资源分别为Sn，n＝1、2和3。业务应用申请的资源规格为0.25S、0.5S或0.75S三种(现实中资源规格是多维度的)。假设基础设施管理系统需要依次处理5次资源请求rX(X＝A、B、C、D、E)，其请求的资源规格和时序关系如图5A和图5B所示，其中相邻两个时刻间的间隔均为时间T，且rA和rD对应的资源在使用一段时间后

被释放，rB、rC和rE则继续长期占用。

[0088] 图5A为按照不考虑生命周期属性信息的原有分配机制进行资源分配的示意图，图5B为按照本发明实施例提供的基于生命周期属性信息进行资源分配的示意图。

[0089] 图5A中，在无法预计后继可预期时间内的资源被占用时间的情况下，按照优先从已有部分资源被分配的基础设施资源实体上分配资源，也即尽量启用新的基础设施资源实体，以及在满足上述条件下，优先选择剩余资源最少、已被分配业务资源最少的基础设施资源实体进行资源分配。这点的作用是避免资源“碎片化”，提高资源利用率。按照资源利用率＝实际分配的资源总量/已启用基础设施资源实体的资源总量的确定方式，简单计算可得到t1～t6期间总体资源利用率只有约56％。在t6之后如果分配情况没有改变，则利用率将长期维持在约33％的低水平。

[0090] 图5B中，按照本发明实施例提供的基于生命周期属性信息进行资源分配的过程如下：

[0091] A：在t1之前，rA和rC提前进行了资源预约，虽然两者的预期占用资源的时间差异很大，但当时并没有其它资源分布可进行优化调配，所以它们都被预约到s1(资源未真正分配占用，仍处于下电状态)。

[0092] B：t1时，rA按照原来的预约进行了资源的分配，s1上电。而在处理rB时，系统分析到rB与rC在后继长时间的预期占用资源的时间基本相同，分配到一起能带来长期的资源优化效果。因此，启动s2分配rB资源，并将rC重预约到s2。

[0093] C：t4时处理rE，在不考虑生命周期属性信息的情况下，按照通用策略应该用s3分配。但由于rE和rB即rC在后继可预期时间内预期占用资源的时间基本相同(均长时间运行)，所以对分配策略进行适当调整，指定由s2分配。

[0094] 通过上述资源分配方式，按照前述资源利用率的计算方式，采用本发明方法的在该示例中的结果是，t1～t6时间内，资源利用率为67.5％(原来为56％)；t6后的长时间运行的资源利用率为100％。

[0095] 本发明实施例提供的资源分配方法，获取业务应用的生命周期属性信息，该生命周期属性信息中能够反映业务应用预期占用资源的时间信息，故可预先获知后继可预期时间内资源被占用的时间。然后基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给业务应用，能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。进一步的，本发明实施例中还可在必要时执行简单的无风险的“重预约”调整，最终使总体资源的使用能在可预期的一段时间内保持优化状态，技术实现简单。

[0096] 需要说明的是，本发明实施例中基础设施管理系统中的基础设施资源实体上的资源可以是基础设施管理系统所管理上层业务应用使用的资源或其组合，例如可以是网络资源(网络连接和通信带宽等)、存储资源、某类特殊处理能力(例如特殊硬件提供的数据加解密处理能力)等等。并且，资源分配的形式也是多样的，例如以计算资源为例(其它类型资源类似)，所分配的资源形式也不只限于使用虚拟化技术的虚拟机，还可包括任何可将物理资源的部分或整体提供给申请者使用的形式，例如Linux内核提供的CGroups技术即能将一个OS管理下的资源(一般为一个物理服务器)划分为多个含有一定资源的实体。进一步的，基

础设施资源实体也是多样的，以计算资源为例(其它类型资源类似)，其物理资源实体所指的范围也不一定为一台物理服务器，而可以是任何能提供一定量计算资源的物理或逻辑集合。例如可以是一个机框内若干刀片服务器、一个机房内的服务器，或者被逻辑划分到一个集群内的服务器等等。

[0097] 进一步的，基于本发明实施例上述提供的资源分配方法，可根据实际情况将基于生命周期属性信息进行资源分配的方式灵活运用于目前的基础设施

[0098] 管理系统中，例如：

[0099] A：生命周期属性信息作为必选属性

[0100] 例如在一些企业或部门内部的基础设施管理系统，为了最大程度地获得本发明方法的效果，可能会要求其用户(如部门内人员)必须在提出资源申请时提供其对应的预计的生命周期信息。

[0101] 还例如在一些企业或部门内部的基础设施管理系统，为了最大程度地获得本发明方法的效果，会严格按照其用户(如部门内人员)在提出资源申请时指定的命周期来分配资源，即在所预先指定的生命周期范围外的时间，用户不能访问使用该资源。当然采用此种方式需要将该约束条件通过使用例如规范等方式提前知会用户。

[0102] B：生命周期属性信息作为可选属性

[0103] 本发明实施例提供的基于生命周期属性信息进行资源分配的方式可以只应用于基础设施管理系统中的设定部分基础设施资源管理实体，而在其余部分应用原有的资源分配方法。一个典型的例子可以是，基础设施管理系统对外可同时支持资源请求携带或不携带生命周期信息的情况，但在内部将两类请求区分处理，对携带了生命周期信息的请求使用本发明的方法进行资源分配和优化，并为其指定一个专用的资源集合；其余请求继续使用原有分配机制(即不考虑资源占用的生命周期)，并指定另外一个资源集合承担资源分配。将两者区分处理的好处是，前者的资源分配均基于资源生命周期进行分析和优化，避免了后者(不考虑资源占用的生命周期)的资源分配和使用的不确定性对前者效果的影响。

[0104] 基于上述实施例提供的资源分配方法，本发明实施例还提供一种资源分配装置600，如图6A所示，该资源分配装置600包括获取单元601、预约单元602和分配单元603，其中：

[0105] 获取单元601，用于获取业务应用的生命周期属性信息，其中，生命周期属性信息反映业务应用预期占用资源的时间信息。

[0106] 预约单元602，用于基于获取单元601获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为业务应用预约预约资源。

[0107] 分配单元603，用于在达到业务应用预期占用资源的起始时间时，将预约单元602预约的预约资源分配给业务应用。

[0108] 在第一种实现方式中，资源分配装置600还包括重预约单元604，如图6B所示，其中：

[0109] 重预约单元604，用于在后继可预期时间内资源利用率满足进一步优化条件的前提下，在预约单元602为业务应用预约预约资源之前，对已预约资源进行重预约。

[0110] 在第二种实现方式中，获取单元601，具体用于按如下方式获取业务应用的生命周期属性信息：

[0111] 调用业务资源请求接口携带业务应用生命周期属性信息功能。

[0112] 通过调用的业务资源请求接口携带业务应用生命周期属性信息功能，获取业务应用的生命周期属性信息。

[0113] 在第三种实现方式中，预约单元602，具体用于按如下方式基于生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源：

[0114] 在设定的部分资源集合中，基于生命周期属性信息，优先选择能够优化后继可预期时间内资源利用率的预约资源。

[0115] 本发明实施例上述提供的资源分配装置600，可以是基础设施管理系统，也可以是基础设施管理系统中的部件，本发明实施例不做限定。

[0116] 本发明实施例提供的资源分配装置600，获取业务应用的生命周期属性信息，该生命周期属性信息中能够反映业务应用预期占用资源的时间信息，故可预先获知后继可预期时间内资源被占用的时间。然后基于获取的生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，为业务应用预约所述预约资源，在达到业务应用预期占用资源的起始时间时，将所述预约资源分配给业务应用，能够使得基础设施管理系统中的资源利用率在可预期时间内得到优化，并保持优化状态。进一步的，本发明实施例中还可在必要时执行简单的无风险的“重预约”调整，最终使总体资源的使用能在可预期的一段时间内保持优化状态，技术实现简单。

[0117] 需要说明的是，本发明实施例提供的资源分配装置600，可用于实现图4、图5B的资源分配方法，故本发明实施例中对资源分配装置600描述不够详尽的地方，可参考相关方式实施例的描述，在此不再赘述。

[0118] 显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

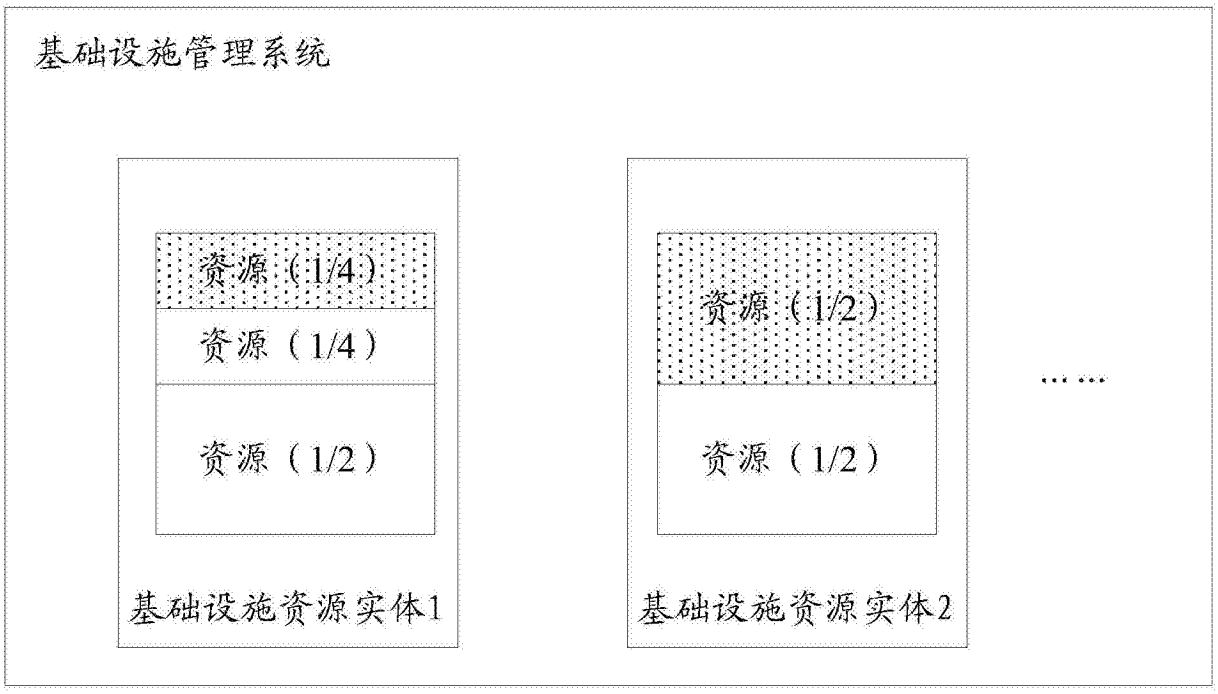

基础设施管理系统
资源（1/4）
资源（1/4）
资源（1/2）
基础设施资源实体1
资源（1/2）
资源（1/2）
基础设施资源实体2
……

图1

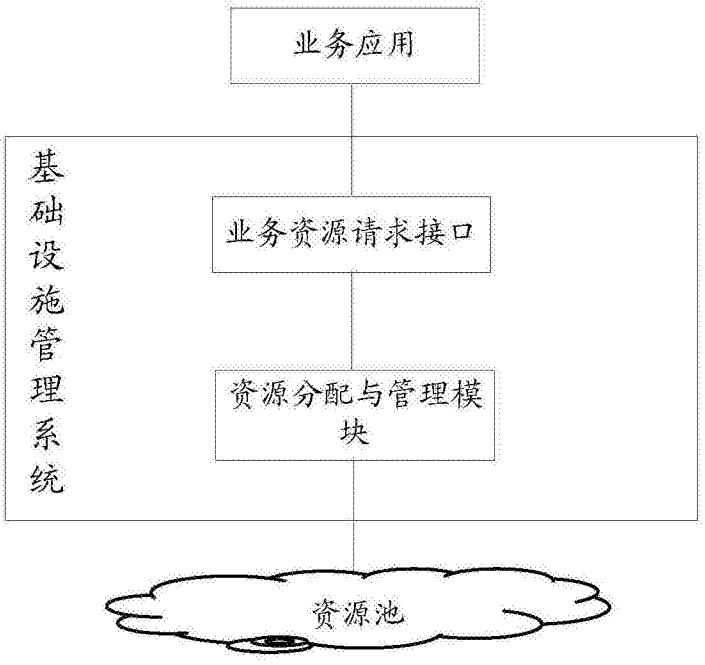

业务应用
基础设施管理系统
业务资源请求接口
资源分配与管理模块
资源池

图2

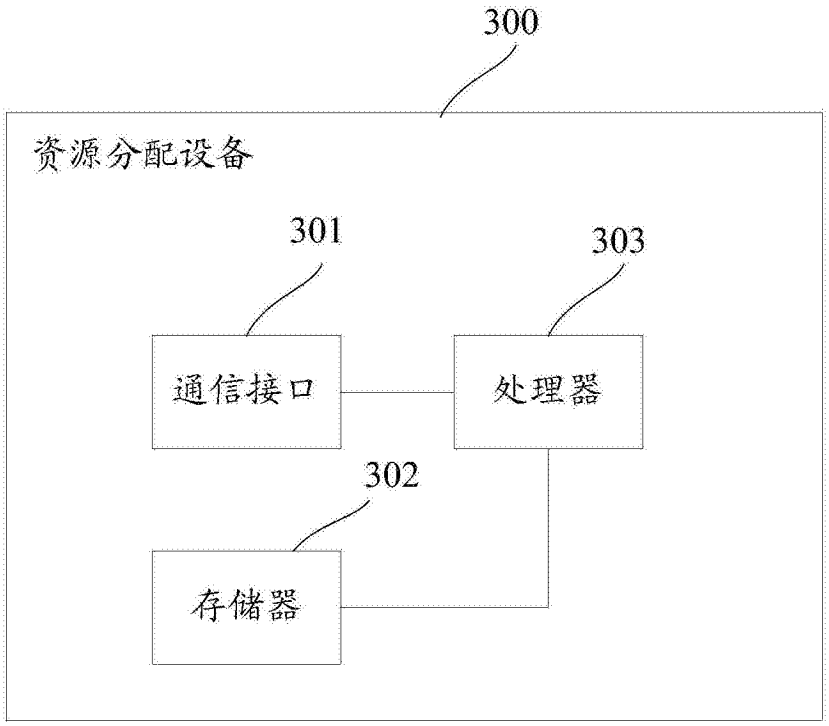

300
资源分配设备
301
303
通信接口
处理器
302
存储器

图3

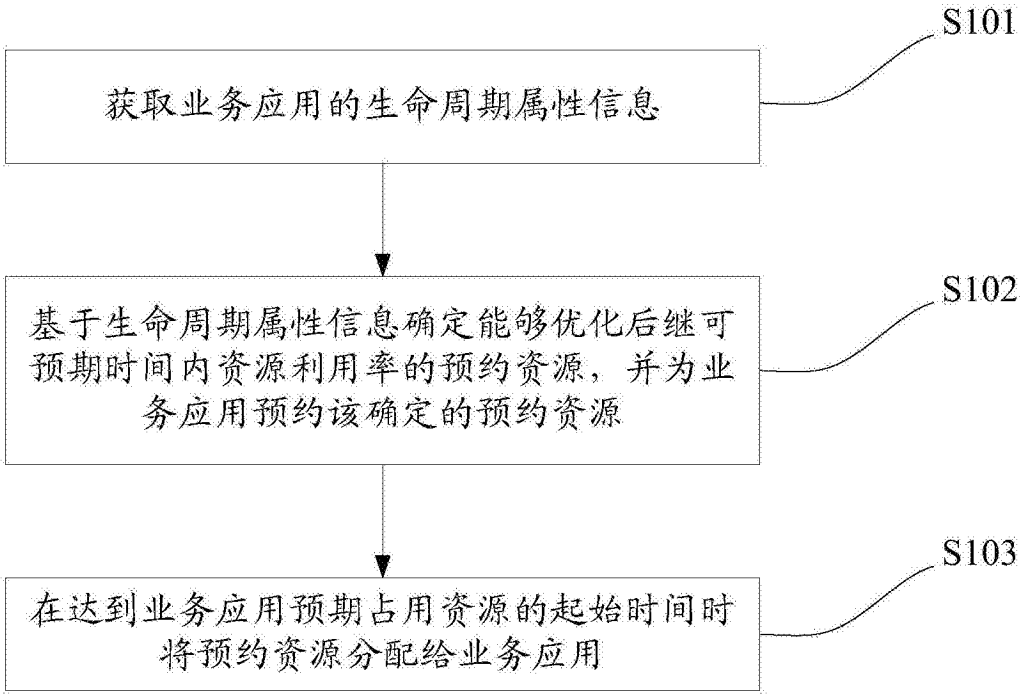

S101
获取业务应用的生命周期属性信息
S102
基于生命周期属性信息确定能够优化后继可预期时间内资源利用率的预约资源，并为业务应用预约该确定的预约资源
S103
在达到业务应用预期占用资源的起始时间时将预约资源分配给业务应用

图4

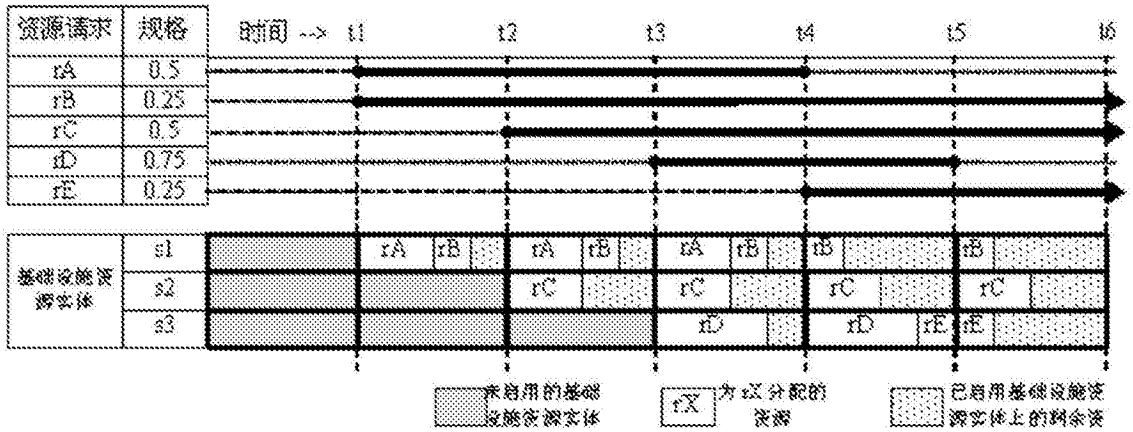

资源请求
规格
时间 -->
t1
t2
t3
t4
t5
t6
rA
0.5
rB
0.25
rC
0.5
rD
0.75
rE
0.25
基础设施资源实体
s1
s2
s3
未启用的基础设施资源实体
rX
为 rX 分配的资源
已启用基础设施资源实体上的剩余资源

图5A

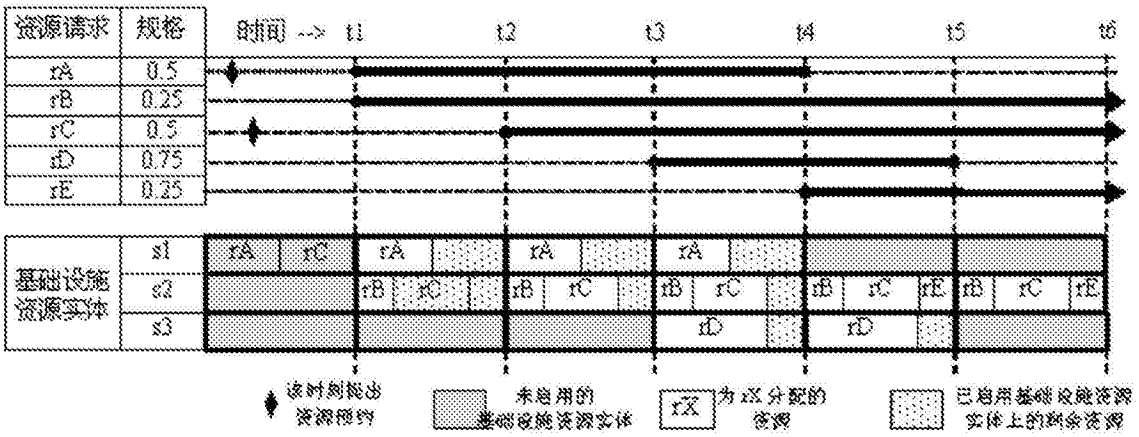

资源请求
规格
时间 -->
t1
t2
t3
t4
t5
t6
rA
0.5
rB
0.25
rC
0.5
rD
0.75
rE
0.25
基础设施资源实体
s1
s2
s3
该时刻预约资源预约
未启用的基础设施资源实体
rX
为 rX 分配的资源
已启用基础设施资源实体上的剩余资源

图5B

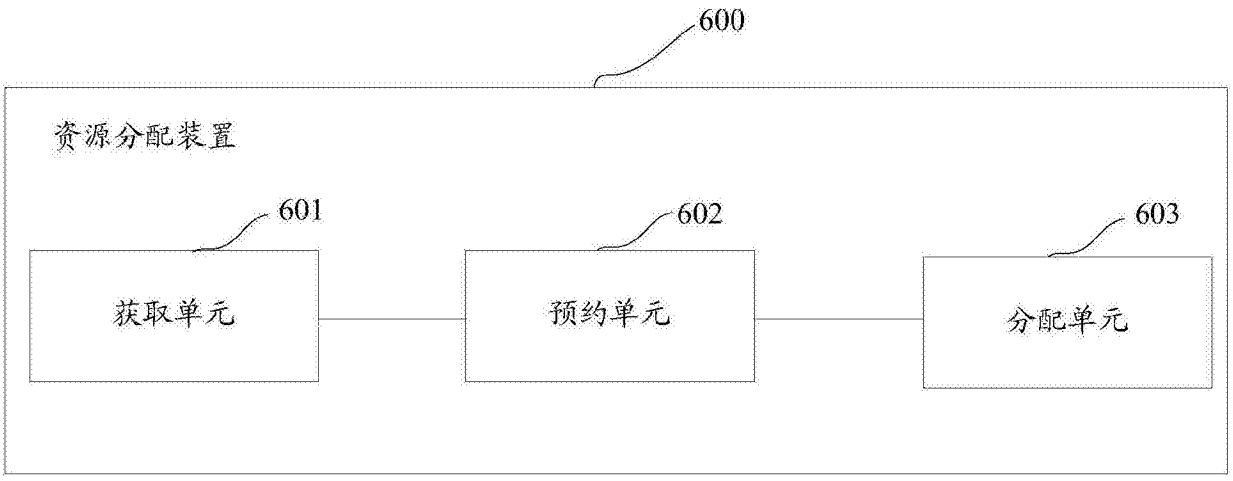

600
资源分配装置
601
获取单元
602
预约单元
603
分配单元

图6A

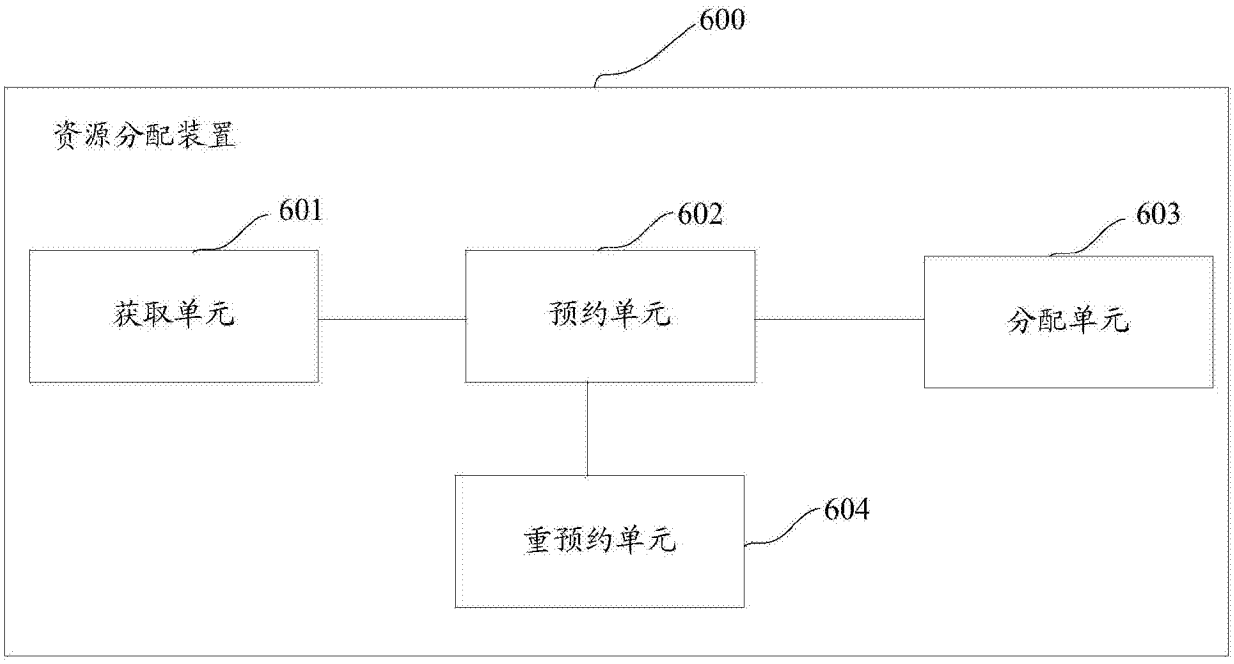


图6B