



(12) 发明专利申请

(10) 申请公布号 CN 102197385 A

(43) 申请公布日 2011. 09. 21

(21) 申请号 200880131724. 2

(22) 申请日 2008. 08. 27

(85) PCT申请进入国家阶段日
2011. 04. 27

(86) PCT申请的申请数据
PCT/US2008/074393 2008. 08. 27

(87) PCT申请的公布数据
W02010/024805 EN 2010. 03. 04

(71) 申请人 惠普开发有限公司
地址 美国德克萨斯州

(72) 发明人 C. E. 巴什 G. H. 富尔马

(74) 专利代理机构 中国专利代理(香港)有限公司 72001
代理人 胡莉莉 王洪斌

(51) Int. Cl.
G06F 15/16(2006. 01)
G06F 9/46(2006. 01)

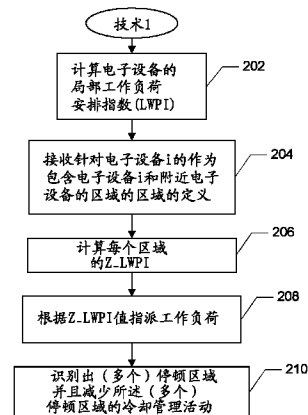
权利要求书 2 页 说明书 10 页 附图 6 页

(54) 发明名称

根据环境条件执行基于区域的工作负荷调度

(57) 摘要

为了根据具有电子设备的系统中的环境条件执行基于区域的工作负荷调度, 累计相对应的区域内的电子设备的冷却效率指标, 以形成相应区域的累计指标, 其中所述区域包括相应的电子设备子集。根据所述累计指标给电子设备指派工作负荷。



1. 一种根据具有电子设备的系统中的环境条件执行基于区域的工作负荷调度的方法,其包括:

累计相对应的区域内的电子设备的冷却效率指标,以形成相应区域的累计指标,其中所述区域包括相应的电子设备子集;以及

根据所述累计指标给电子设备指派工作负荷。

2. 权利要求 1 的方法,其还包括接收区域的定义,其中所述接收包括以下各项之一:

(1) 接收其中每个区域都包括相对应的特定电子设备和附近电子设备的定义;以及

(2) 接收规定预先划分的区域的定义。

3. 权利要求 1 的方法,其还包括基于所述累计指标给电子设备指派基于区域的冷却效率,

其中,给电子设备指派工作负荷包括根据基于区域的冷却效率对电子设备进行排序。

4. 权利要求 3 的方法,其中,指派基于区域的冷却效率包括使用所述累计指标作为所述基于区域的冷却效率。

5. 权利要求 3 的方法,其中,指派基于区域的冷却效率包括基于把累计指标与电子设备的冷却效率的相对应的单独指标相组合来计算所述基于区域的冷却效率。

6. 权利要求 1 的方法,其中,累计区域中的特定区域的指标包括以下各项之一:计算指标的平均值、确定指标的中值、确定指标的最大值以及确定指标的最小值。

7. 权利要求 1 的方法,其还包括:

识别出作为工作负荷指派的结果而变为停顿的至少一个区域;以及

减少所述至少一个区域的冷却活动。

8. 权利要求 7 的方法,其中,所述系统在区域中的每个区域内包括至少一个传感器,所述方法还包括基于所述至少一个传感器的温度测量来计算至少一个区域内的电子设备的冷却效率指标,

其中,减少至少一个区域的冷却活动包括解除分配至少一个区域内的至少一个传感器。

9. 权利要求 7 的方法,其中,所述系统在区域中的每个区域内包括至少一个传感器,所述方法还包括基于所述至少一个传感器的温度测量来计算至少一个区域内的电子设备的冷却效率指标,

其中,减少所述至少一个区域的冷却活动包括为所述至少一个区域内的设定点温度配置更大余量。

10. 权利要求 1 的方法,其中,所述指标包括用以把电子设备维持在预定温度范围内的能量使用的度量。

11. 一种根据系统中的环境条件执行基于区域的工作负荷调度的方法,其中所述系统具有电子设备和递送用于冷却所述电子设备的冷却流体的流体移动设备的,所述方法包括:

识别相应的流体移动设备的瓶颈传感器;

基于所识别出的瓶颈传感器来识别区域;

给所述区域内的电子设备指派基于区域的冷却效率;以及

根据所述基于区域的冷却效率给所述电子设备指派工作负荷。

12. 权利要求 11 的方法,其还包括:

选择瓶颈传感器之一;

识别为了使得所选择的瓶颈传感器能被忽略而不得不停顿的电子设备,其中所述不得不停顿的电子设备形成电子设备区域。

13. 权利要求 12 的方法,其还包括:重复所述选择以便选择瓶颈传感器中的另一瓶颈传感器,重复所述识别以便识别为了使得所述瓶颈传感器能被忽略而不得不停顿的另一电子设备区域。

14. 权利要求 12 的方法,其中,选择所述一个瓶颈传感器包括:

执行仿真,以便在假设相应的瓶颈传感器被忽略的情况下确定流体移动设备的新操作点,

其中,所选择的流体移动设备是在流体移动设备的新操作点下坚持具有最高能量节省的流体移动设备。

15. 一种包括至少一个计算机可读存储介质的产品,所述至少一个计算机可读存储介质包含指令,当运行所述指令时,所述指令使得计算机:

累计相对应的区域内的电子设备的冷却效率指标,以形成相应区域的累计指标,其中所述区域包括相应的电子设备子集;以及

根据所述累计指标给电子设备指派工作负荷。

根据环境条件执行基于区域的工作负荷调度

背景技术

[0001] 数据中心可以被定义为容纳被布置在多个机架中的计算机的场所、例如房间(或多个房间)。这些机架被配置成容纳相对大数目的电子设备(例如计算机、存储设备、电信设备等等),这些电子设备包含在操作期间散热的组件(例如印刷电路板、电源、处理器等等)。

[0002] 诸如空调单元之类的冷却单元通常被用来冷却经过加热的空气,并且被用来把经过冷却的空气供给所述电子设备。通常经由被安置在集气室(plenum)上方的地板中的一系列通风砖(vent tile)来供给经过冷却的空气,所述集气室把气流从空气冷却单元引导到所述通风砖。在传统上,通常针对最坏情况或峰值负荷情形来供应及操作冷却单元。由于通常的数据中心操作仅仅利用所述数据中心内的小部分电子设备,所以针对最坏情况或峰值负荷情形来供应冷却单元的做法常常不够高效。

[0003] 此外,在某些常规实施方式中,通常按照随机方式或者基于遵循电子设备可用性的方案来为所述机架中的电子设备安排(place)工作负荷。因此,常规的系统通常为电子设备安排工作负荷,并且随后根据被供给回到冷却单元的气流的温度变化来减少或增加冷却单元操作。按照这种方式操作冷却单元的效率不高,这是因为所述冷却单元通常消耗的能量量多于用于使电子设备足够冷却所需的能量量。

[0004] 在某些其它实施方式中,可以计算冷却效率指数,以使得能够根据所述指数来为计算机分配工作负荷。所述指数可以标识出局部热点和热空气再循环,或者可替换地,所述指数可以使得能够对数据中心内的最为高效的位置进行排序以便安排工作负荷。但是这样的方法通常集中于在每个电子设备的基础上安排新的工作负荷,而并没有考虑到以下事实:从中去除工作负荷的一个电子设备(“第一电子设备”)可被具有相对高的工作负荷(并且因此可能具有高冷却需求)的其它电子设备所包围。即使已从第一电子设备中去除了工作负荷,周围的电子设备仍然可继续有高冷却需求。

附图说明

[0005] 参照下面的附图通过举例的方式描述本发明的某些实施例:

图 1 是根据实施例的包括电子设备和冷却管理系统的数据中心的简化透视图;

图 2 是根据实施例的基于区域的工作负荷管理和热管理的第一技术的流程图;

图 3 是根据另一实施例的数据中心内的被划分成多个区域的电子设备的示意图;

图 4 是根据另一实施例的基于区域的工作负荷管理和热管理的第二技术的流程图;

图 5 是根据又一实施例的基于区域的工作负荷管理和热管理的第三技术的流程图;以

及

图 6 是根据某些实施例的包括用于执行任务的软件模块的示例性计算机的框图。

具体实施方式

[0006] 根据某些实施例,提供一种机制或技术来根据包括电子设备的系统的环境条件在(可能位于数据中心或其它设施中的)的系统的电子设备上执行基于区域的工作负荷调度。

“工作负荷”指的是通常具有某个相当大的持续时间的请求、事务、批处理计算工作或者可以在电子设备中执行的任何活动。环境条件包括温度、湿度等等。所述系统可以是(包括计算机,例如诸如刀片服务器之类的计算机服务器的)处理系统、(包括存储设备的)存储系统、(包括 web 服务器的)基于 web 的系统等等。基于区域的工作负荷调度考虑到针对相对应的电子设备区域所计算的冷却效率。“电子设备”指的是可以针对工作负荷被调度的那些设备。

[0007] 在某些实施例中,所述系统的电子设备被划分成多个区域(其中每个区域都包括一个或多个电子设备)。举例来说,图 3 示出了把四个机架 302 — 308 中的计算机系统划分成一组区域 310 — 320 的实例。为电子设备计算把位于相对应的区域内的其它电子设备包括在内的基于区域的冷却效率。换句话说,对于每个特定的电子设备,为该特定电子设备计算的基于区域的冷却效率把来自包含该特定电子设备的区域中的其它电子设备的贡献包括在内。由工作负荷安排控制器根据所述基于区域的冷却效率来为电子设备安排工作负荷。

[0008] 电子设备的“冷却效率”指的是表现出会被用来把该电子设备维持在温度范围内(或者维持在温度设定点)的能量的量的度量。电子设备的冷却效率的指标可以是指数,诸如局部工作负荷安排指数(LWPI, local workload placement index)或供热指数(SHI, supply heat index),下面将对这些指数做进一步描述。可替换地,在某些实施方式中,冷却效率的指标可以仅仅是绝对温度,诸如在电子设备处测量的温度或者在电子设备机架的进口处测量的温度。在一个实例中,如果设定点温度在整个系统内是统一的,则与电子设备相关联的绝对温度会表现出为了冷却所述电子设备会消耗多少能量(换句话说,与更接近温度设定点的较高温度相关联的电子设备会表明在冷却这样的电子设备时可能会涉及更多能量)。

[0009] 可以为相对应的电子设备计算各单独冷却效率的指标。单独冷却效率是在不考虑来自相邻电子设备的影响的情况下为电子设备计算的冷却效率。这种单独冷却效率与上面所讨论的说明来自相邻电子设备的贡献的基于区域的冷却效率不同。针对每个区域内的电子设备的各单独冷却效率的指标被累计来形成该相应区域的累计指标。所述累计指标可以被用作电子设备的基于区域的冷却效率,或者可替换地,所述累计指标可以被用来导出电子设备的基于区域的冷却效率。所述工作负荷安排控制器随后根据所述基于区域的冷却效率来给电子设备指派工作负荷。

[0010] 此外,通过考虑区域内的条件而不是单独考虑电子设备的条件,热管理控制器可以检测出一个或更多区域是否已停顿(quiesce)(这样的区域内的电子设备或者被切断或者由于没有为这些电子设备安排工作负荷而处于空闲状态)。所述热管理控制器随后可以减少(多个)停顿的区域的冷却管理活动。减少冷却管理活动指的是以下各项中的任意一项或更多项:(1)关断与停顿的区域相关联的冷却设备;(2)诸如通过解除分配(多个)停顿区域的温度传感器而忽略来自所述(多个)停顿区域内的温度传感器的温度读数;或者(3)把(多个)停顿区域内的(多个)设定点温度配置成具有更大余量,这意味着在所述热管理控制器采取增加对所述(多个)停顿区域的冷却的步骤之前,允许如由所述(多个)停顿区域内的传感器所检测到的温度比所述(多个)设定点温度超出更大的余量。

[0011] 在某些实施例中,为相对应的单独电子设备计算的指标被称作局部工作负荷安排

指数(LWPI),其中 LWPI 表现出将电子设备维持在预定温度范围内的能量使用效率。在其它实施例中可以使用其它类型的指标。举例来说,所能使用的另一指标是供热指数(SHI),该供热指数(SHI)是电子设备的热管理和能量效率的指标。下面进一步描述 LWPI 和 SHI。在另外的实施例中可以使用其它指标。

[0012] 图 1 是数据中心 100 的部分的简化透视图。术语“数据中心”被用来表示能够发热的一个或更多电子设备可处于其中的房间或其它空间。

[0013] 数据中心 100 被描绘为具有多行机架 102a — 102n,其中“n”是大于 1 的整数。机架行 102a — 102n 通常被配置成容纳可以被调度来执行工作负荷并且能够发热/散热的电子设备 116,例如计算机、存储设备、交换机、路由器、显示器等等。电子设备 116 可以工作来执行各种电子功能,例如计算、存储、交换、路由、显示等等。

[0014] 机架行 102a — 102n 被描绘成被安置在架高的地板(或其它表面)110 上,这些机架行 102a — 102n 可以充当用于递送来自一个或更多流体移动设备(fluid moving device) 114a — 114n 的经过冷却的流体(诸如空气、制冷剂、水、水和乙二醇混合物等等)的集气室。如果所述流体是气体、诸如空气或气体制冷剂,则通过通风孔 118 向机架行 102a — 102n 递送所述流体。在其它情况下,如果所述流体是液体、诸如水、液体制冷剂、多态制冷剂等等,则可以通过一系列管道(未示出)向机架行 102a — 102n 递送所述流体。

[0015] 所述流体移动设备(或“冷却设备”)114a — 114n (其中“n”是大于 1 的整数)通常工作来向在架高的地板 110 之下的空间 112 供给流体流,并且冷却先前由电子设备加热的流体,(由箭头 128 指示)。此外,流体移动设备 114a — 114n 通常还工作来向被容纳在机架行 102a — 102n 中的电子设备 116 供给已经利用空调(AC)单元冷却过的流体流。举例来说,流体移动设备 114a — 114n 可以包括蒸气压缩类型的空调单元、冷水机组(chiller)类型的空调单元等等。更一般来说,“流体移动设备”指的是可以执行以下各项中的一项或全部两项的设备:(1)移动流体;以及(2)控制所述流体的温度。因此,流体移动设备可以仅仅是风扇或鼓风机,或者可替换地,流体移动设备可以是能够控制正被移动的流体的温度的 AC 单元。

[0016] 下文描述对流体移动设备的调节,这可以指调节流体流的速度和/或调节空调设备的温度设定点。

[0017] 流体移动设备 114a — 114n 可以包括相应的致动器(未示出),这些致动器被配置来操纵被供给到机架行 102a — 102n 的经过冷却的流体流的特性,诸如流体流温度和供给率。因此,所述致动器例如包括用于操纵所述经过冷却的流体的流体流温度的设备(诸如冷水机组、热交换器等等),并且还包括用于操纵所述经过冷却的流体的供给流率的设备(诸如由可变频率驱动器操作的鼓风机等等)。所述致动器可以由热管理控制器控制来执行热管理,以便防止电子设备过热。

[0018] 由箭头 124 指示的经过冷却的流体通过位于某些或所有机架行 102a — 102n 之间的流体递送设备 118 从空间 112 被递送到所述机架行 102a — 102n。流体递送设备 118 例如可以包括通风瓦、可变流体流量设备等等,并且被示为位于机架行 102a 与 102b 之间以及机架行 102c 与 102n 之间。虽然流体递送设备 118 和空间 112 已被描绘成位于数据中心 100 的地板上,但是应当理解的是,在可替换的实施方式中,流体递送设备 118 和空间 112 也可以被安置在数据中心 100 的天花板或墙壁上。

[0019] 包含在空间 112 中的经过冷却的流体可以包括由一个或更多流体移动设备 114a — 114n 供给的经过冷却的流体,并且在某些情况下还包括再循环到空间 112 中的流体流。因此,所述经过冷却的流体的特性(诸如温度、压力、湿度、流率等等)可能会受到多个所述流体移动设备 114a — 114n 的操作的显著影响。在这方面,处于空间 112 中的不同区处的经过冷却的流体和被供给到机架行 102a — 102n 的经过冷却的流体的特性例如可能会由于经过冷却的流体的混合而变化。换句话说,被供给到数据中心 100 中的特定位置的经过冷却的流体的特性可能不同于由单个流体移动设备 114a 供给的经过冷却的流体的特性。此外,通过流体递送设备 118 供给的经过冷却的流体的特性会受到流经其它流体递送设备 118 的流体的特性的影响。

[0020] 在图 1 中还示出了传感器 120a — 120n,其中“n”是大于 1 的整数,所述传感器 120a — 120n 被配置来检测在这些传感器的相应位置处的温度。(被表示为菱形的)传感器 120a — 120n 被安置来检测在机架行 102a — 102n 的进口附近的不同位置处的温度。虽然没有示出,但是传感器 120a — 120n 可以包括与所述电子设备 116 中的一个或更多相关联或者与之集成制造的传感器。然而,可替换地,传感器 120a — 120n 也可以包括分开安装的传感器 120a — 120n。

[0021] 传感器 120a — 120n 可以通过网络连接到控制子系统 130。如下所述,控制子系统 130 可以采用由传感器 120a — 120n 在不同的流体移动设备 114a — 114n 设置下所检测到的条件来确定每个流体移动设备 114a — 114n 对每个传感器 120a — 120n 附近的区的影响程度。所检测到的条件例如可以包括温度、压力、流体流量、湿度等等。此外,通过确定电子设备 116 中的哪个电子设备位于传感器 120a — 120n 中的哪个传感器附近,控制子系统 130 还可以确定每个流体移动设备 114a — 114n 对每个电子设备 116 的影响程度。

[0022] 流体移动设备 114a — 114n 可以由控制子系统 130 控制。虽然控制子系统 130 在图 1 中被图示为与容纳在机架行 102a — 102n 中的电子设备 116 分开,但是控制子系统 130 可以被实施在电子设备 116 中的一个或更多中,或者在可替换的实施方式中可以被实施在数据中心 100 外部的电子设备中。

[0023] 虽然后缀“n”被用来表示机架行(102a — 102n)、流体移动设备(114a — 114n)和传感器(120a — 120n)的数目,但是应当注意的是,可以存在不同数目的机架行、流体移动设备和传感器。

[0024] 控制子系统 130 可以包括上面讨论的工作负荷安排控制器和热管理控制器,以及可能还有用于执行其它任务的其它模块。例如可以利用一个或更多计算机来实施控制子系统 130。

[0025] 根据某些实施例,控制子系统 130 能够计算数据中心 100 中的电子设备的各单独冷却效率的指标。此外,控制子系统 130 还能够接收所述系统中的区域的定义,其中每个区域都包括相应的一个或更多电子设备的子集。换句话说,数据中心 100 的电子设备被划分成多个区域,其中这种区域的定义被提供给控制子系统 130。控制子系统 130 能够累计相对应的区域中的电子设备的各单独冷却效率的指标,以形成相应区域的累计指标。如上所述,所述累计指标可以被用作电子设备的基于区域的冷却效率,或者可替换地,所述累计指标可以被用来导出电子设备的基于区域的冷却效率。利用所述基于区域的冷却效率来对电子设备进行排序,并且根据所述排序来给电子设备指派工作负荷。通过说明区域条件提高

经由执行更加高效的工作负荷分配而节省更多能量的潜能。

[0026] 可以使用数种替换技术来执行基于区域的工作负荷安排和热管理,所述替换技术可以由安排控制子系统 130 来执行。根据实施例,图 2 示出了第一技术(“技术 1”)。利用该技术,为数据中心 100 中的相对应的电子设备计算(202) LWPI 形式的各单独冷却效率的指标。可替换地,各单独冷却效率的指标可以是另一类型的指标,诸如 SHI。

[0027] 控制子系统 130 还接收(204)针对每个电子设备 i 的区域的定义。在该第一技术中,为任何给定电子设备 i 定义的区域包括该电子设备 i 和附近的电子设备。所述“附近的”电子设备可以包括与电子设备 i 紧邻的电子设备,或者可替换地,所述“附近的”电子设备可以包括从电子设备 i 向外延伸某一数目的电子设备(例如延伸 m 个电子设备,其中 m 可以是 1、2、3 等等)的电子设备。可替换地,电子设备 i 的“附近的”电子设备可以是处于与电子设备 i 相距预定义距离内的电子设备。

[0028] 因此,每个电子设备 i 与附近的电子设备的相对应的区域相关联,并且根据电子设备 i 以及该区域内的附近电子设备的单独的 $LWPI_i$ 来计算基于区域的 $LWPI_i$ (被称为 Z_LWPI_i)。换句话说,电子设备 i 的区域是包含电子设备 i 以及附近的电子设备的区域。应当注意的是,根据该技术,区域的数目与电子设备的数目一样多,以致对于相对应的电子设备 i 将有相对应的数目的 Z_LWPI_i 值。

[0029] 接下来计算(206)每个区域的累计的冷却效率指标。在该例中,针对电子设备 i,累计指标是可以根据下式来计算的 Z_LWPI_i :

$$Z_LWPI_i = \frac{LWPI_i + \sum_j LWPI_j}{J+1} \quad (\text{等式 1}).$$

[0030] 根据等式 1,在电子设备 i 的区域内有 J 个附近的电子设备。所述 J 个附近的电子设备的 LWPI ($LWPI_j$, j=1 到 J)与电子设备 i 的 $LWPI_i$ 相加以形成 Z_LWPI_i 。可替换地,不是如等式 1 中那样对 LWPI 求平均,而是可以执行不同的累计,诸如确定中值、最大值、最小值、基于距离的加权平均值或者某种其它类型的累计。

[0031] 随后根据 Z_LWPI_i 值来指派(208)工作负荷。举例来说,可以根据电子设备的相应的 Z_LWPI_i 值来对这些电子设备进行排序。控制子系统 130 随后可以选择具有最高排序(Z_LWPI_i 值)的电子设备来执行所选择的工作负荷。按照这种方式,为具有(如由 Z_LWPI_i 值表示的)最高冷却效率的电子设备指派工作负荷。

[0032] 接下来,控制子系统 130 识别出(210)任何(多个)停顿区域,所述(多个)停顿区域是其中所有其电子设备都已被关断或处于空闲或者可以被置于较低功率状态(例如休眠状态)的(多个)区域。停顿区域可以被定义为其中温度传感器的温度读数被允许爬升到高于标准温度阈值水平的区域。举例来说,假设所有活跃电子设备的进气温度都要被保持低于 25°C,但是在停顿区域内,所述进气温度可以被允许升高到 30°C,或者交替地升高到任意其它温度。作为在 208 处执行的工作负荷指派的结果,当数据中心的利用率低于 100% 时,可能有一个或更多整个区域将处于停顿。

[0033] 在(多个)停顿区域内,冷却管理活动减少。冷却管理活动的减少可以涉及与该停顿区域相关联的传感器的解除分配。解除分配传感器意味着控制子系统 130 的热管理控制器在执行热管理时并不考虑来自这种被解除分配的传感器的测量数据。可替换地,可以为

停顿区域内的传感器的设定点温度给出更大余量,从而在由所述热管理控制器采取任何行动之前允许来自所述传感器的温度读数比所述设定点温度超出比通常所允许的余量更大的余量。

[0034] 下面描述计算各种电子设备的单独的 LWPI 的一种实施方式。在于 2006 年 10 月 31 日提交的标题为“Method for Assessing Electronic Devices (用于评估电子设备的方法)”(代理人案卷号为 200600611-1)的美国序列号为 11/590,525 中提供了关于计算单独的 LWPI 的进一步细节。

[0035] 可以如下计算每个电子设备 i 的单独的 $LWPI_i$:

$$LWPI_i = \frac{(T_{set} - T_{in} + [(T_{SAT} - T_{SAT,min})_j \cdot TCI_j]_i}{(T_{in} - T_{SAT})_i} \quad (\text{等式 2}).$$

[0036] 在该例中,控制子系统 130 基于每个电子设备 116 处的热管理余量、流体移动设备余量、TCI (热相关系数) 和再循环水平中的一个或更多计算所示电子设备 116 中的每个的 $LWPI_i$ 。对于给定的电子设备 116,所述热管理余量可以被定义为该电子设备 116 的设定点温度 (T_{set} , 其是针对该电子设备 116 的最高允许温度) 与进口温度 (T_{in}) 之间的差。因此,所述热管理余量可以定义可用于该电子设备 116 的附加冷却余量。举例来说,如果服务器处于 16°C (T_{in}) 并且所述设定点温度 (T_{set}) 是 25°C , 则附加的冷却余量是 9°C , 这是可得到的温度增加量。

[0037] 此外,所述流体移动设备余量可以被定义为供给流体温度 (T_{sat}) 与最低供给流体温度 ($T_{sat,min}$) 之间的差,同时关于位置 i 把针对特定流体移动设备 114 的 TCI 作为因素计入。换句话说,所述流体移动设备余量通常表明特定流体移动设备 114 的影响特定电子设备 116 的给定地区 (region) 的温度可以被降低多少。举例来说,如果流体移动设备 114a 的供给流体温度是 20°C , 并且流体移动设备 114a 的最低供给流体温度是 14°C , 则将该信息与所述 TCI 相组合,以便确定来自任意给定流体移动设备 114a - 114n 的多少温度降低可用于任意给定电子设备 116。如果流体移动设备 114a 正工作在非常低的温度下并且因此正消耗大量能量,则在确定冷却效率时将这一事实纳入考虑。

[0038] 所述供给流体温度 (T_{sat}) 和最低供给流体温度 ($T_{sat,min}$) 可以是关于特定电子设备 116 具有预定 TCI 水平的所有流体移动设备 114a - 114n 的平均温度。附加地或可替换地,在确定 T_{sat} 和 $T_{sat,min}$ 时,控制子系统 130 可以根据不同流体移动设备 114a - 114n 关于所述特定电子设备 116 的相应的 TCI 水平对这些流体移动设备 114a - 114n 所供给的温度进行不同的加权。因此,例如与由具有相对较低 TCI 的流体移动设备 114b 所供给的流体温度相比,为由具有显著更高 TCI 的流体移动设备 114 所供给的流体温度将给出更高加权。

[0039] 每个电子设备 116 处的再循环水平通常可以被定义为被再循环回到该电子设备 116 中的热流体的数量。更具体来说,对于特定的电子设备 116,所述再循环水平可以被定义为该电子设备 116 的进口温度 (T_{in}) 与被确定为对该电子设备 116 具有预定影响水平的至少一个流体移动设备 114a 的供给流体温度 (T_{sat}) 之间的差。到电子设备 116 中的热流体再循环的水平越高,为该电子设备 116 安排工作负荷的效率就越低。

[0040] 所述 TCI 是流体移动设备 114a - 114n 对传感器 120a - 120n 所具有的相对影响

水平的函数。因此,例如,流体移动设备 114a 的传感器 120a 的 TCI 值越高,流体移动设备 114a 对该传感器 120a 的影响就越大。

[0041] 虽然可以通过任意数目的适当相关算法来确定所述 TCI,但是可以采用下面的算法来计算传感器 120a - 120n (i) 的 TCI。

$$[0042] \quad TCI_i = \frac{(T_1 - T_2)}{N} \quad (\text{等式 3}).$$

[0043] 在等式 3 中, T_1 是在第一步记录的温度测量,而 T_2 是在第二步记录的温度测量。此外, N 是流体移动设备 114 在第一步与第二步之间的供给流体条件变化的数量。换句话说,在将给定的流体移动设备调节所述供给流体条件变化 (N) 之前,对于给定传感器记录 T_1 。在所述供给流体条件变化 (N) 之后并且在预定时间段之后,执行第二步并且测量 T_2 。因此, T_2 代表在对给定流体移动设备做了变化 (N) 之后的给定传感器的温度。

[0044] 举例来说,如果在第一步记录的传感器 120a 的温度 (T_1) 是 20°C,在第二步记录的传感器 120a 的温度 (T_2) 是 25°C,并且供给流体的温度变化 (N) 是 +10°C,则传感器 120a (i) 与流体移动设备 114a 之间的 TCI_i 是 0.5。作为另一例子,如果在第一步记录的传感器 120b 的温度 (T_1) 是 20°C,在第二步记录的传感器 120b 的温度 (T_2) 是 21°C,并且供给流体的温度变化 (N) 是 +10°C,则传感器 120b 与流体移动设备 114a 之间的 TCI_i 是 0.10。因此,传感器 120a 对流体移动设备 114a 的 TCI 高于传感器 120b。因此,如与传感器 120b 相比,对来自流体移动设备 114a 的供给的流体流所做的变化可能对传感器 120a 处的条件具有更大影响。

[0045] 可替换地,不是使用 LWPI 作为冷却效率的指标,而是可以使用不同的指标,诸如供热指数 (SHI)。正如上面所述的那样,SHI 充当每个电子设备的熱管理和能量效率的指标。

[0046] 可以如下确定每个电子设备的 SHI (ϕ):

$$\phi = \left[\frac{\delta Q}{Q + \delta Q} \right] = \left[\frac{T_{in} - T_{ref}}{T_{out} - T_{ref}} \right]. \quad (\text{等式 4})$$

其中, δQ 是由于热空气渗透而引起的热负荷,而 Q 是实际热负荷。此外, T_{in} 表示电子设备的进口温度,而 T_{out} 表示电子设备的出口温度。 T_{ref} 可以表示一个或更多流体移动设备的供给空气温度。可替换地, T_{ref} 可以表示一个或更多流体移动设备的温度。可替换地, T_{ref} 可以表示通过一个或更多通风铤 (vent file) 118 提供的气流的温度。

[0047] 在公开号为 2006/0047808 的美国专利申请中描述了关于如何可以将 SHI 用于工作负荷安排管理的进一步细节。

[0048] 不是使用第一技术来执行基于区域的工作负荷安排和热管理,而是可以使用第二替换技术 (“技术 2”)。在图 3 和 4 的情境中描述了执行基于区域的工作负荷安排和热管理的第二技术。

[0049] 图 3 示出了示例性机架行 102 (图 1 的机架行 102a - 102n 之一)。图 3 的机架行 102 包括机架 302、304、306 和 308,每个机架都包括相应的电子设备集合。图 3 的某些电子

设备具有不同的轮廓,以表明这种电子设备可能是不同类型的电子设备。但是,在其它实施方式中,包含在所述机架中的电子设备可以全部是相同类型的。

[0050] 如图 3 中所描绘的那样,所述电子设备被划分成多个区域 310、312、314、316、318 和 320。区域 310 延伸跨越四个机架 302、304、306 和 308 的底部。区域 312 延伸跨越两个机架 302 和 304,而区域 314 包括仅仅一个机架内的电子设备。所述区域可以具有任意形状。数据中心 100 中的其它机架行可以类似地被分成相对应的区域。

[0051] 与用在第一技术中的相对应的电子设备的单独区域对比,第二技术的区域被视为预先划分的区域。

[0052] 利用第二技术,如在图 4 中所描绘的那样,相对应的电子设备的单独的 LWPI 被计算(402)。可替换地,可以计算其它类型的单独指标。接下来,接收(404)所述预先划分的区域的定义,并且为每个区域计算(406) Z_LWPI_i 。在一个例子中,可以基于用户对数据中心的理解来手动执行区域的预先划分。计算每个区域的 Z_LWPI_i 可以涉及到取得该区域(区域 310 — 320 之一)内的电子设备的单独的 LWPI 的累计。计算单独的 LWPI 的累计包括:计算单独 LWPI 的平均值,确定单独的 LWPI 的最大值,确定单独的 LWPI 的最小值,计算单独的 LWPI 的中值,或者基于所述区域内的电子设备的单独的 LWPI 计算某种其它累计。

[0053] 在一个实施例中,给区域内的所有电子设备指派相同的相对应的 Z_LWPI 值,这确保该区域内的电子设备都被排序在相同级别。控制子系统 130 的工作负荷安排控制器将通过在开始填充另一区域之前首先填充一个区域的方式来指派工作负荷。

[0054] 在不同的实施例中,可以给区域内的每个电子设备指派不同的指标值,该指标值等于 Z_LWPI 加上基于该电子设备的单独的 LWPI 的另一值。举例来说,电子设备 i 的指标值可以是(针对电子设备 i 所在的区域的) Z_LWPI 与乘以电子设备 i 的单独的 LWPI (这被称作经过调节的 Z_LWPI 值)的某一常数 k (例如 0.0001)之和。电子设备的不同指标值接着可以被用作在每个区域内给电子设备指派工作负荷的平局决胜机制。

[0055] 如在图 4 中进一步描绘的那样,根据 Z_LWPI 值(或者经过调节的 Z_LWPI 值)来指派(408)工作负荷。接下来,与图 2 中的任务 208 和 210 类似,识别出(410)(多个)停顿区域并且减少针对这样的(多个)停顿区域的冷却管理活动。

[0056] 不是利用上面描述的第一或第二技术,而是可以使用第三技术(“技术 3”)来执行基于区域的工作负荷指派和热管理。图 5 示出了该第三技术的示例性流程图。

[0057] 输入(502)关于完整传感器栅格的信息。对于每个流体移动设备执行以下任务 504 — 508。识别出(504)“瓶颈”传感器。对于每个流体移动设备 114,存在安排该流体移动设备的当前操作点的下界的单个瓶颈传感器。该瓶颈传感器是距所期望的设定点(参考温度)具有最小热余量的传感器(温度与设定点最接近的传感器)。可替换地,如果所有参考温度都是统一的,则所述瓶颈传感器可以是最热的传感器。在具有 N 个流体移动设备的数据中心中,可以是在 1 到 N 个瓶颈传感器之间的任何地方。

[0058] 所述系统可以包括驱动流体移动设备的操作的传感器(换句话说,所述流体移动设备基于这样的传感器的测量来调节其操作)。这些传感器可以被称作“主”传感器。瓶颈传感器是防止流体移动设备进一步减少其容量的“主”传感器。应当注意的是,所有瓶颈传感器都是主传感器,但并非所有主传感器都是瓶颈传感器。

[0059] 通过热区域内的温度变化来确定主传感器是否是瓶颈传感器,或者更确切地说,

通过传感器与其设定点之间的差的变化来确定主传感器是否是瓶颈传感器。如果该变化大于预定义的数量(比如说例如 3°C),则该传感器充当高效操作的障碍或瓶颈,并且因此被标识为瓶颈传感器。另一方面,如果所述变化小,则该传感器不被视为瓶颈传感器。

[0060] 接下来,如果流体移动设备的瓶颈传感器不予考虑,则可以由控制子系统 130 执行(506)软件仿真,以确定该流体移动设备的(不同于该流体移动设备的当前操作点的)新操作点。可替换地,可以利用启发法来确定去除所述瓶颈传感器将是否值得。作为另一替换方案,可以执行软件仿真以将传感器分类为瓶颈传感器(在这种情况下,在这样的替换实施例中可以把任务 506 移到任务 504 之前)。

[0061] 可以按照多种方式执行软件仿真。举例来说,在第一实例中,可以使用与所述热管理控制器分开的仿真软件,除了所述仿真软件采用与所述热管理控制器的软件相同的算法以外。所述仿真软件可以把在假设所述瓶颈传感器被去除或处于停顿状态的情况下所得到的新操作条件与较高的设定点进行比较。

[0062] 最热传感器与次最热传感器之间的温度差反映出在从与最热传感器相关联的电子设备去除工作负荷的情况下总体温度所能升高的程度。接着,可以利用所述仿真软件、诸如基于相对应的 TCI 值,对会导致这种温度升高的流体移动设备的操作变化(降低)进行仿真。正如上面所述的那样,TCI 是流体移动设备对传感器的相对影响水平的函数。所确定的流体移动设备的操作变化接着提供该流体移动设备的新操作点。

[0063] 可替换地,可以通过以下方式来执行仿真 506:利用数值仿真软件(诸如基于计算流体动力学的数值仿真)来对去除与瓶颈传感器相关联的电子设备处的负荷的效果进行仿真,并且随后把所述流体移动设备的操作降低到所述新操作点。

[0064] 接下来确定(508)在流体移动设备的当前操作点与新操作点之间所能实现的能量节省。所述确定可以基于特定流体移动设备的功率消耗模型。

[0065] 选择(510)坚持具有最大能量节省的瓶颈传感器。随后给不得被关断的所有电子设备指派最差的可用的 Z_LWPI 值,以忽略所选择的瓶颈传感器。这种不得被关断以忽略瓶颈传感器的电子设备被视为瓶颈区域的部分,其具有被指派给这种瓶颈区域的基于区域的冷却效率指标(Z_LWPI)。“瓶颈区域”指的是与特定瓶颈传感器相关联的电子设备区域。可以预先定义 Z_LWPI 值的列表,其中把 Z_LWPI 值从最差到最好进行排序。可以从该列表中选择可用的最差 Z_LWPI 值,以指派所述瓶颈区域内的电子设备。为了确保一致性,应注意的是,根据该第三技术的算法被指派给电子设备的所有 Z_LWPI 值都来自于所述列表。应注意的是,由于图 5 的任务 504 — 510 被迭代地执行,所以下一次迭代会选择接下来可用的最差 Z_LWPI 值,以指派电子设备。该第三技术的 Z_LWPI 值与第一和第二技术的不同之处在于:所述 Z_LWPI 值是预先定义的,而不是基于单独的电子设备的条件计算的。此外,利用第三种技术定义的区域是基于识别瓶颈传感器的区域。

[0066] 接下来,对于不予考虑的针对先前选择的流体移动设备的瓶颈传感器,重复(512)任务 504、506、508 和 510。任务 504、506、508 和 510 的下一代迭代将导致把接下来最差 Z_LWPI 值指派给电子设备。图 5 的过程被迭代执行,直到所有电子设备都已被指派 Z_LWPI 值为止。

[0067] 在可替换的实施例中,不是针对每个流体移动设备执行任务 504、506 和 508,而是可以针对已经被识别出的所述 1 到 N 个瓶颈传感器中的每个瓶颈传感器执行所述任务。在

这样的可替换的实施例中,对于每个瓶颈传感器执行图 5 的任务 506 和 508,其中任务 506 的仿真同时考虑所有流体移动设备。在 510 选择提供最大能量节省的瓶颈传感器。

[0068] 作为上述过程的变型,不是给针对要被忽略的给定流体移动设备的瓶颈传感器的每组要被关断的电子设备指派相同的 Z_LWPI ,所指派的目标值可以是 Z_LWPI 加上乘以每个相应电子设备的单独的 $LWPI$ 的 k (常数,诸如 0.00001)。

[0069] 上面讨论的技术 1 和 2 可以在数据中心中的工作负荷的斜升期间被使用,或者在工作负荷的斜降(合并(consolidation))期间被使用。可以实时地计算 Z_LWPI 值,或者可替换地,可以使用历史的 Z_LWPI 值。技术 3 可以在合并期间被使用。在数据中心的正常操作期间,可以使用技术 1、2 和 3 中的任意一种(或某种组合)。

[0070] 图 6 是示例性控制子系统 130 的方框图,该控制子系统 130 可以是计算机(或多个计算机)。控制子系统 130 包括连接到存储装置 604 的一个或多个中央处理单元(CPU)602。作为软件模块的工作负荷安排控制模块 606、热管理模块 608 和仿真模块 610 可以在所述(多个)CPU 602 上运行。控制子系统 130 还包括网络接口 612,以使得网络接口 130 能够通过网络诸如与传感器和流体移动设备进行通信。

[0071] 上面进一步提到的工作负荷安排控制器包括工作负荷安排控制模块 606 与(多个)CPU 602 的组合。类似地,所述热管理控制器包括热管理控制模块 608 与(多个)CPU 602 的组合。仿真模块 610 是在上面描述的技术 3 中使用的仿真软件。

[0072] 各种软件模块的指令被加载以便在处理器(诸如(多个)CPU 602)上运行。所述处理器包括微处理器、微控制器、处理器模块或(包括一个或更多微处理器或微控制器的)子系统或者其它控制或计算设备。“处理器”可以指代单个组件或指代多个组件。

[0073] (所述软件的)数据和指令被存储在相应的存储设备中,这些存储设备被实施为一个或更多计算机可读的或计算机可用的存储介质。所述存储介质包括不同形式的存储器,这些存储器包括:半导体存储器设备,诸如动态或静态随机存取存储器(DRAM 或 SRAM)、可擦写和可编程只读存储器(EPROM)、电可擦写和可编程只读存储器(EEPROM)以及闪速存储器;磁盘,诸如硬盘、软盘和可移动盘;包括磁带在内的其它磁介质;以及光学介质,诸如光盘(CD)或数字视频光盘(DVD)。应注意的是,上面讨论的软件的指令可以被提供在一个计算机可读的或计算机可用的存储介质上,或者可替换地可以被提供在分布于可能具有多个节点的大系统中的多个计算机可读的或计算机可用的存储介质上。这样的或更多计算机可读的或计算机可用的存储介质被视为产品(或制造产品)的部分。产品或制造产品可以指代任何所制造的单个组件或多个组件。

[0074] 在前面的描述中阐述了许多细节以便提供对于本发明的理解。但是,本领域技术人员将理解的是,可以在没有这些细节的情况下实践本发明。虽然关于有限数目的实施例已公开了本发明,但是本领域技术人员将会想到许多修改及其变型。所附权利要求书意图如涵盖落在本发明的精神和范围内的这样的修改和变型。

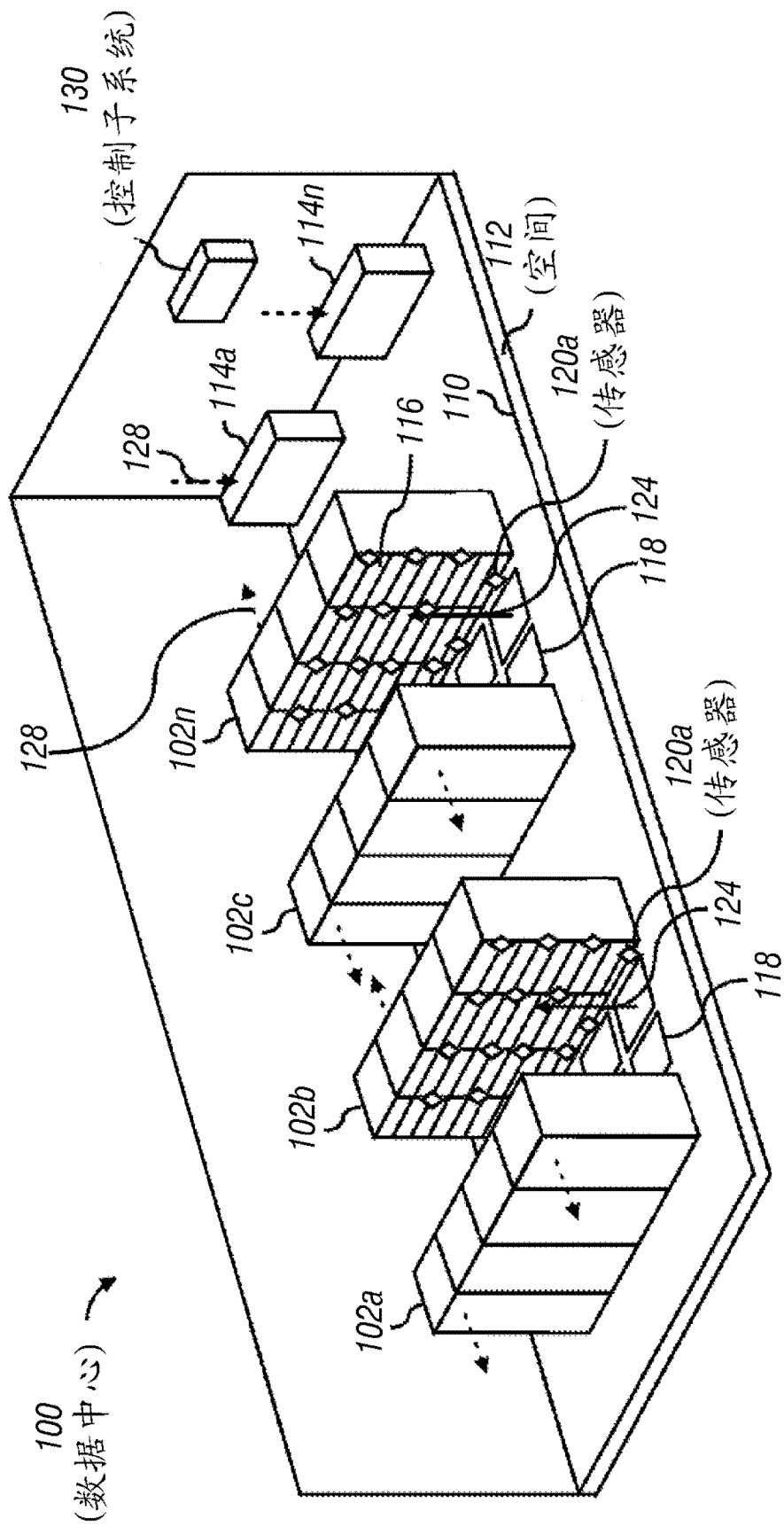


图 1

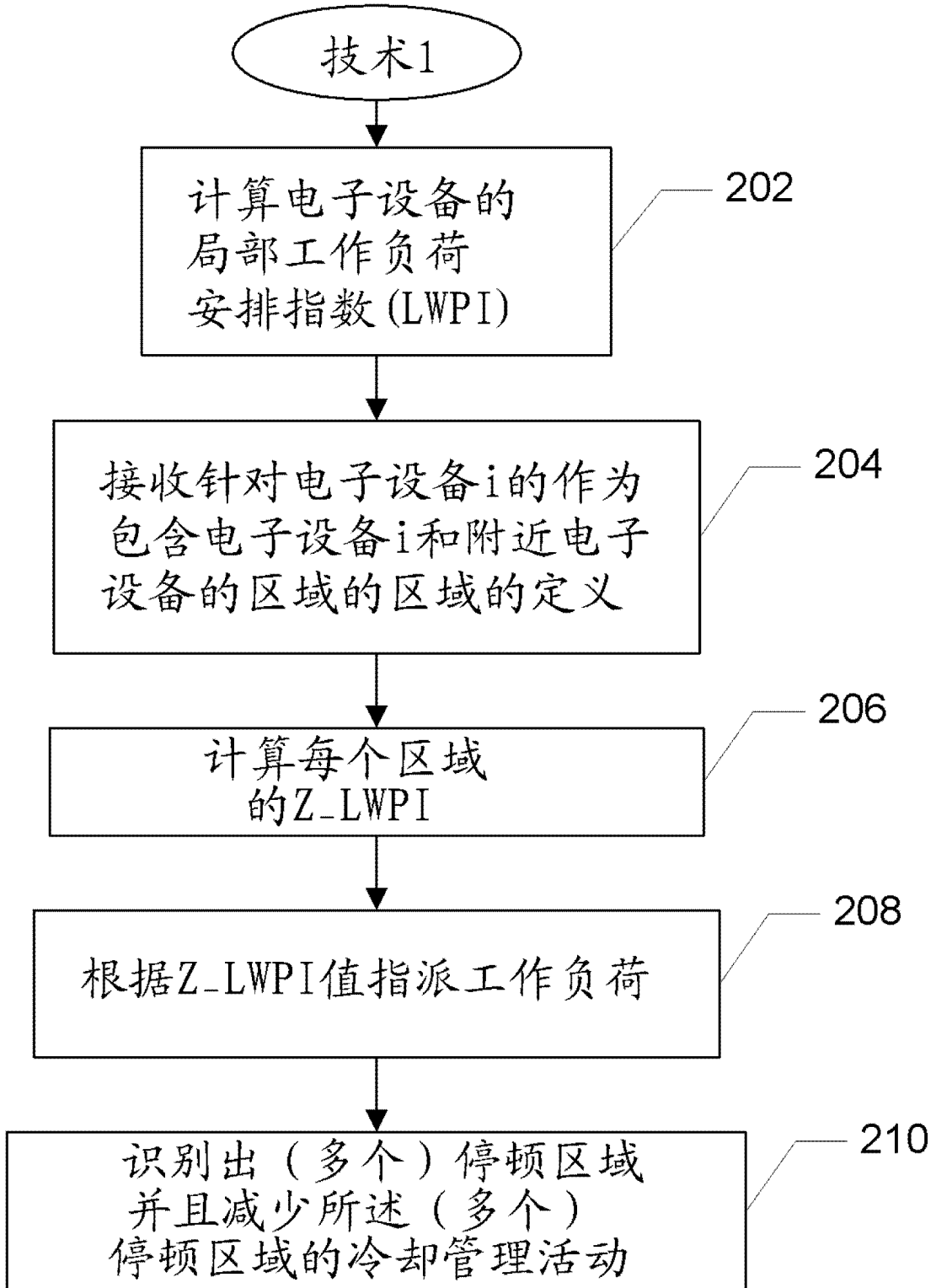


图 2

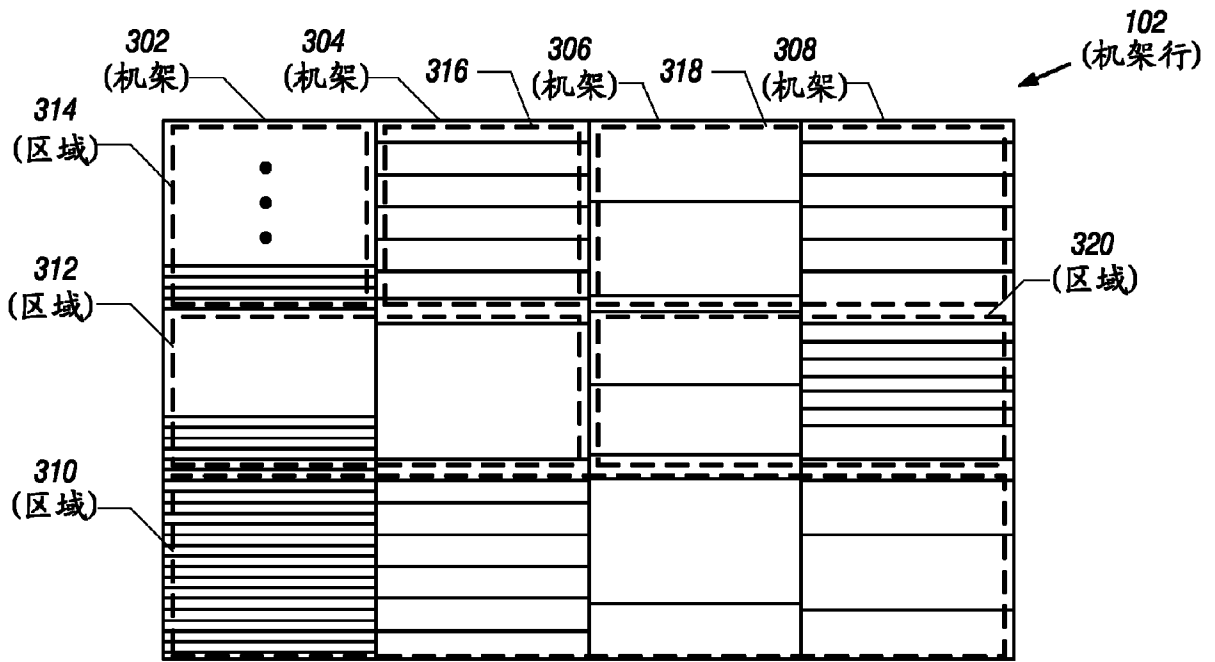


图 3

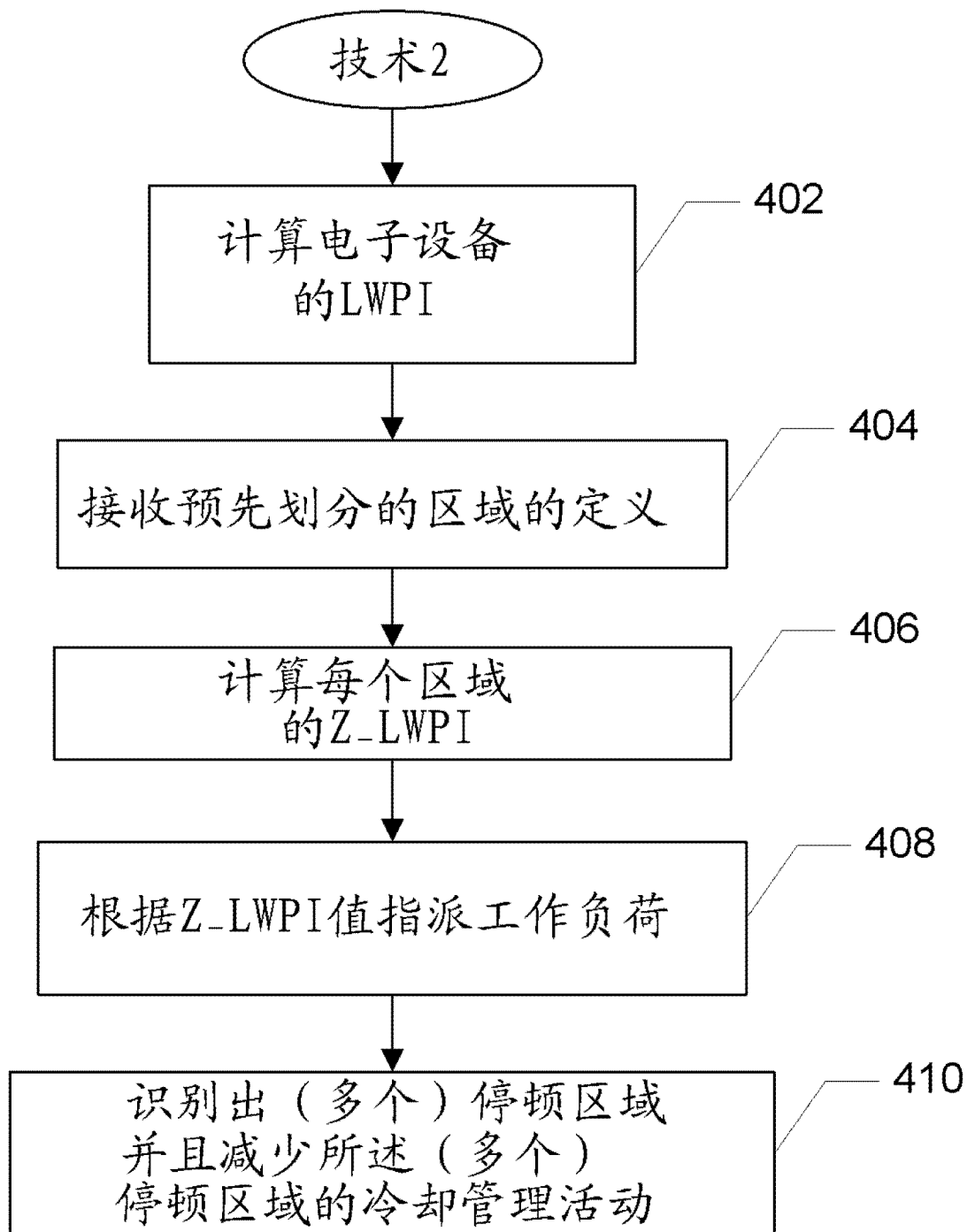


图 4

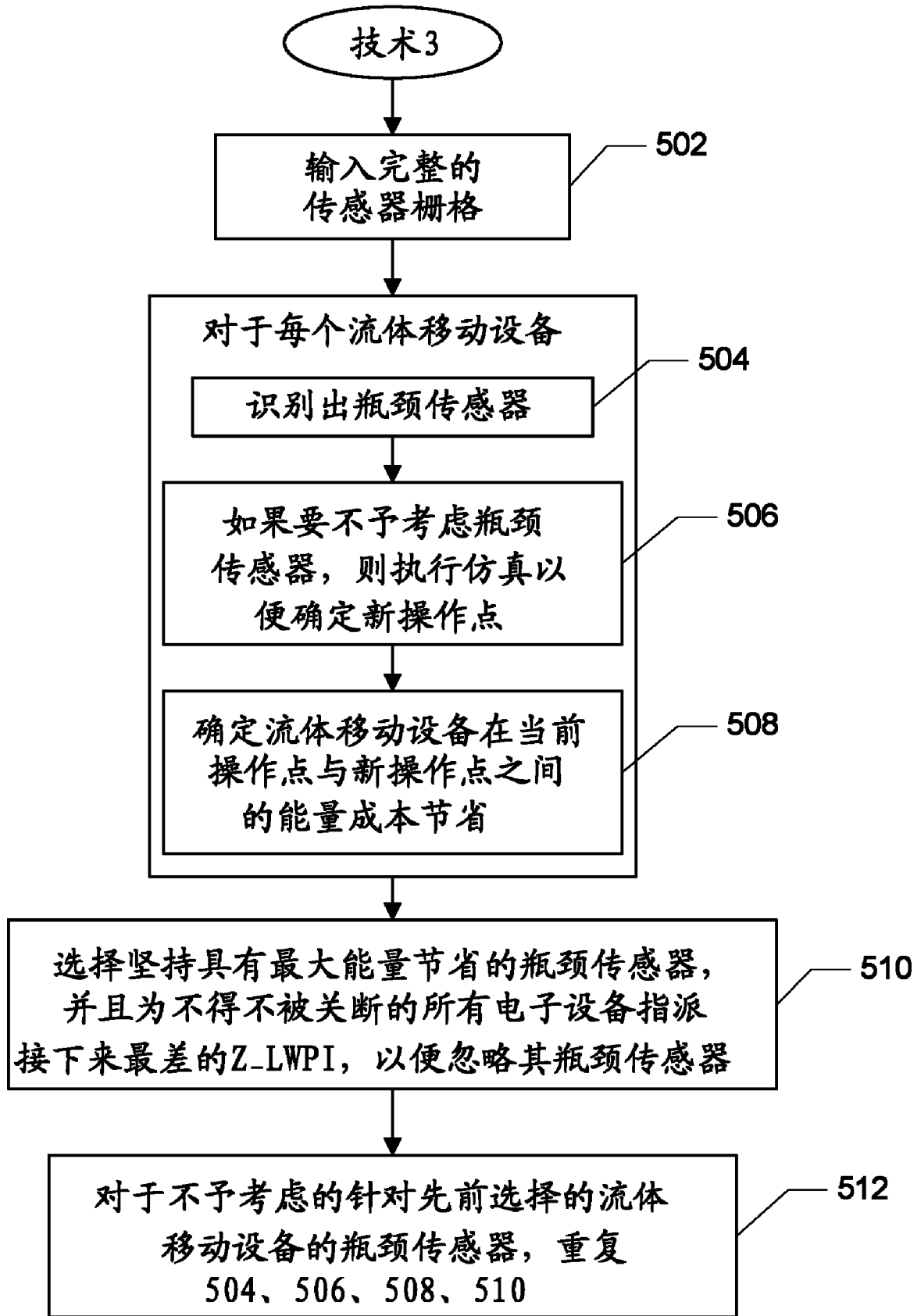


图 5

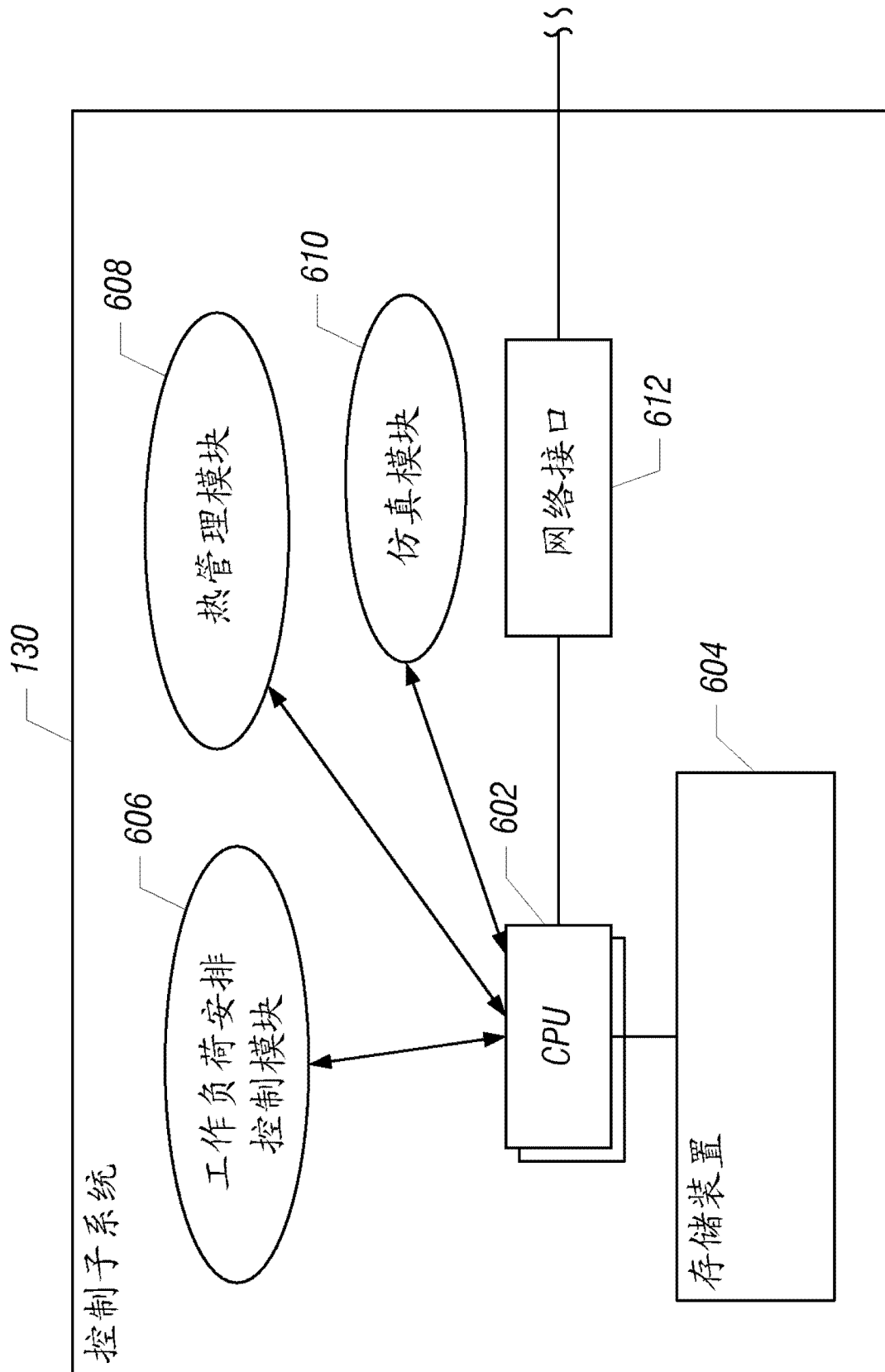


图 6