



(12) 发明专利

(10) 授权公告号 CN 102326134 B

(45) 授权公告日 2015. 07. 29

(21) 申请号 201080009556. 7

(22) 申请日 2010. 02. 04

(30) 优先权数据

12/391, 188 2009. 02. 23 US

(85) PCT国际申请进入国家阶段日

2011. 08. 22

(86) PCT国际申请的申请数据

PCT/US2010/023253 2010. 02. 04

(87) PCT国际申请的公布数据

W02010/096283 EN 2010. 08. 26

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 N·贾殷 C·J·威廉姆斯

J·劳瑞斯 D·里德

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 杨洁

(51) Int. Cl.

G06F 1/32(2006. 01)

G06F 15/16(2006. 01)

(56) 对比文件

US 2006/0184287 A1, 2006. 08. 17,

US 2003/0196126 A1, 2003. 10. 16,

US 7134011 B2, 2006. 11. 07,

CN 101088072 A, 2007. 12. 12,

CN 1510570 A, 2004. 07. 07,

审查员 罗湘

权利要求书2页 说明书11页 附图5页

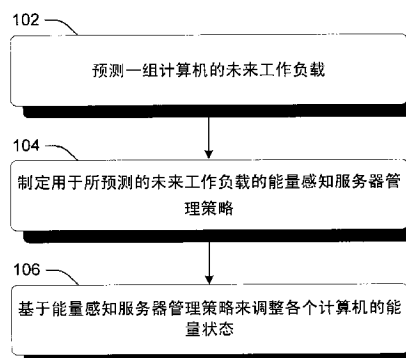
(54) 发明名称

能量感知服务器管理

(57) 摘要

所描述的实现涉及能量感知服务器管理。一个实现包括自适应控制单元, 该自适应控制单元被配置为通过在活动和不活动状态之间转变各个服务器并且以预定义级别维持服务器场的响应时间来在服务器场中管理能量使用。

方法 100



1. 一种用于能量感知服务器管理的方法 (100), 包括:

预测一组计算机的未来工作负载, 其中各个计算机具有至少两个能量状态 (102), 所预测的未来工作负载利用涉及过去数据中未表示的即将到来的事件的外部信息;

将对所述未来工作负载的响应时间建模为所述未来工作负载的量和活动服务器的数量的函数;

制定用于所预测的未来工作负载的能量感知服务器管理策略 (104); 以及

基于所述能量感知服务器管理策略调整各个计算机的能量状态 (106)。

2. 如权利要求 1 所述的方法, 其特征在于, 所述预测包括将所述未来工作负载估计为, 用于给定未来时间段的请求的预测数量与过去时间段内每个请求的平均数据量的乘积。

3. 如权利要求 1 所述的方法, 其特征在于, 所述制定包含对包括所述未来工作负载和服务级协定的函数求解。

4. 如权利要求 1 所述的方法, 其特征在于, 所述制定考虑用于与状态转变相关联的各个计算机的可能负面的期望寿命结果。

5. 如权利要求 1 所述的方法, 其特征在于, 在所预测的未来工作负载小于实际的未来工作负载的场景中, 所述制定考虑将不活动状态计算机转变到活动状态的转变时间。

6. 如权利要求 1 所述的方法, 其特征在于, 所述调整还包括发送标识各个计算机调整后的状态的信号。

7. 如权利要求 1 所述的方法, 其特征在于, 所述调整考虑在各个能量状态之间的转变时间。

8. 如权利要求 1 所述的方法, 其特征在于, 所述调整考虑与所预测的未来工作负载相关联的不确定性。

9. 如权利要求 1 所述的方法, 其特征在于, 所述调整留下未利用的计算机的缓冲区作为保障以维持未来请求的响应时间。

10. 一种用于能量感知服务器管理的系统 (200), 包括:

自适应控制单元 (204), 被配置为:

预测对服务器池的多个未来请求, 所述服务器池包括被配置具有活动状态和至少两个不同的不活动状态的各个服务器, 所述预测利用涉及过去数据中未表示的即将到来的事件的外部信息;

将未来请求的响应时间建模为未来请求的数量和活动服务器的数量的函数;

计算满足响应时间的运行于活动状态和不活动状态中的每一个的服务器的数量; 以及使得各个服务器在状态之间进行转变以满足所计算的数量。

11. 如权利要求 10 所述的系统, 其特征在于, 所述自适应控制单元包括被配置为预测所述服务器池的未来工作负载的工作负载预测模块。

12. 如权利要求 11 所述的系统, 其特征在于, 所述自适应控制单元包括策略模块, 所述策略模块被配置为通过在活动和不活动状态之间转变各个服务器且满足与所预测的工作负载相关联的性能标准来解决能量使用的方式处理所预测的工作负载。

13. 如权利要求 10 所述的系统, 其特征在于, 所述自适应控制单元包括被配置为根据能量感知服务器管理策略来调整各个服务器的状态的状态调整模块。

14. 如权利要求 10 所述的系统, 其特征在于, 还包括基于来自所述自适应控制单元的

指令为各个活动状态服务器分配工作负载的负载分派器。

能量感知服务器管理

背景技术

[0001] 数据中心通常包含大量的计算机（即，服务器）。这些服务器中的很多个暂时未被充分利用或甚至未被使用，但被维护以处理达到数据中心的请求的激增。例如，考虑在线购物站点的圣诞节高峰，或者通勤者交通信息站点下午早些时候的拥塞。数据中心中当前使用的服务器即使在其空闲时也消耗大量的能量（对于某些服务器，空闲占用全部能量使用的 70%）。关闭这些服务器可节省能量，但如果用户的激增到来且没有足够的开着机的服务器，则它可不利地影响 web 服务的响应时间。本发明概念可以以一种满足请求服务响应时间且考虑数据中心的能量使用的方式来管理数据中心。

发明内容

[0002] 所描述的实现涉及能量感知服务器管理。更具体地，本专利申请涉及用于请求 / 响应服务的能量感知供应和负载分派的工具。

[0003] 一个实现涉及自适应控制单元，它被配置为通过在活动和不活动状态之间转变服务器并且以预定义级别维持服务器场的响应时间来在服务器场中管理能量使用。例如，可在服务级协定（SLA）中定义预定义级别。

[0004] 在另一个实现中，工具预测一组计算机的未来工作负载。各个计算机具有诸如开机 / 活动、睡眠和关机的至少两种能量状态。该工具制定用于所预测的未来工作负载的能量感知服务器管理策略，并且基于该能量感知服务器管理策略来调整各个计算机的能量状态。

[0005] 例如术语“工具”可以指以上上下文和通篇文档中所准许的设备、系统、计算机可读指令（例如，一个或多个具有可执行指令的计算机可读存储介质）、组件、模块和 / 或方法。在各种实例中，可用硬件、软件、固件或其组合来实现工具。以上列出的示例旨在提供快速参考以帮助读者，并且不旨在限定此处所描述的概念的范围。

附图说明

[0006] 附图示出了本申请中传达的概念的实现。所示实现的特征可通过参考以下结合附图的描述来更容易地理解。只要可行，各附图中相同的附图标记用来指代相同的元素。此外，每一个如图标记的最左边的数字传达其中首次引入该附图标记的附图及相关联的讨论。

[0007] 图 1 示出用于根据本发明概念的某些实现来实施能量感知服务器管理概念的方法的示例。

[0008] 图 2 示出用于根据本发明概念的某些实现来获得能量感知服务器管理的系统的示例。

[0009] 图 3 示出用于根据本发明概念的某些实现来获得能量感知服务器管理的组件的示例。

[0010] 图 4-5 示出与根据本发明概念的某些实现的能量感知服务器管理方法相关联的

流程图。

具体实施方式

[0011] 概览

[0012] 本发明申请涉及能量感知服务器管理。简言之，本发明描述用于以预期的性能级别操作诸如服务器场的一组计算机并且降低能耗的概念。

[0013] 本发明概念可采用响应于用户请求或“请求”的服务器场中的多状态服务器。当被共同地浏览时，请求可被认为是服务器场的“工作负载”。多状态服务器可具有处理请求的活动状态和不处理请求的不活动状态。某些多状态计算机可具有多种不活动状态。例如，睡眠、休眠、和 / 或关机可以是不活动状态的示例。处于不活动状态的服务器可使用比活动状态的服务器少得多的能量。但是，可能存在将不活动状态的服务器转变到活动状态中涉及的例如几秒钟到几分钟的等待时间段。此等待时间段可在附加的服务器可用于（即活动的）处理请求之前产生不想要的处理延迟。

[0014] 因此，能量感知服务器管理可考虑或平衡若干参数，诸如降低能量使用且仍以及时的方式响应请求的等待时间和性能标准。图 1 以方法 100 的形式提供本发明概念中的某些的介绍性示例。在此情形中，在 102 预测或估计诸如服务器计算机的一组计算机的未来工作负载。该预测可涉及诸如几秒钟到几分钟的相对短的时间段。在此特定的情形中，各个计算机可具有两种或多种能量状态。例如，计算机可具有活动状态（即开机）和一种或多种不活动状态（即，关机、睡眠和休眠等）。

[0015] 下面在题为“工作负载预测示例”下描述工作负载预测方法的详细示例。简言之，在某些实例中，工作负载预测可利用相对长期建模、相对短期历史、当前趋势、和 / 或外部知识（例如，商业项目、已调度的事件）等作为工作负载预测过程中考虑的参数。

[0016] 长期建模可观察相对于一天中的时间、几周中的天数、一年中的时间、多年循环的诸如选择等的工作负载趋势。例如，每个工作日上午 8 点到 9 点之间可存在一致的工作负载需求，因为员工上班并在开始其一天时作出各种个人和 / 或工作相关的请求。

[0017] 短期历史可涉及过去几秒钟或几分钟的用户请求级别。例如，受欢迎的电视节目上对网站的提及可能驱动接下来几分钟的对该网站的点击。过去几秒钟或几分钟内的向上趋势的指示可用于预测接下来几秒钟或几分钟的未来的请求。

[0018] 当前趋势可涉及在工作负载预测过程期间接收到的用户请求。例如，当前趋势可随最近可用的时刻从请求的图形的形状中导出。该图形可以是不变的，或者可按时间序列分布（例如，线性、指数）变化。

[0019] 外部信息可被认为是，不是从当前或过去的请求中获取的信息。相反，外部信息可以例如涉及过去数据中未表示的即将到来的事件。例如，外部信息可传达奥运会即将到来，并且在奥运会持续期间可能经历异常高的请求数量，因为用户会尝试下载来自事件的视频。

[0020] 作为对请求数量的替换或补充，外部信息可涉及请求大小。例如，继续上面的示例，外部信息可指示在奥运会期间请求的数量可上升，并且每个请求的大小可能大于正常请求，因为请求往往涉及下载数据密集的视频。

[0021] 以上描述提供非限制性的工作负载预测示例。其他示例可预测请求大小。所预测

的请求大小可随后被用于工作负载预测中。总之,可能存在预测未来工作负载的很多方式。所预测的未来工作负载可随后用于以下描述的系统 / 方法等。

[0022] 在 104 可确定用于所预测的工作负载的能量感知服务器管理策略。简言之,能量感知服务器管理策略可包括容量供应。在某些实现中,容量供应可被认为是抢先的(即,抢先的容量供应),因为它可随着工作负载变化以在线、实时的方式进行。抢先的容量供应也可被认为是抢先的,因为它可被抢先地执行以屏蔽服务器在不同的电量状态之间转变所用的时间。这可与在事件已经发生“之后”执行决策和任何响应性动作的反应性决策制定形成对比。

[0023] 换言之,处于其当前状态的一组计算机具有处理给定工作负载的容量。如果处于其当前状态的计算机的容量不同于用于处理所预测的工作负载的容量,则可采用容量供应以调整计算机中的某些的状态,以将未来容量与所预测的工作负载相对齐。

[0024] 能量感知服务器管理策略可包括改变计算机中的一种或多种的状态(即,从活动状态转移到不活动状态,反之亦然)。能量感知服务器管理策略可平衡多个因素,包括但不限于该组计算机的能量使用、用于响应用户请求的及时性预期、预测 / 估计的置信度、和 / 或对与状态转变相关联的各个计算机的预期服务生命的影响。下面在题为“容量供应示例”下更详细地描述这些因素。

[0025] 在 106 可基于能量感知服务器管理策略调整各个计算机的能量状态。例如,可将某些计算机从活动转移到不活动状态以节省能量。能量感知服务器管理策略还可指定如何在活动状态计算机之间分配请求。

[0026] 能量感知服务器管理策略的另一方面可指定特定的计算机应当处于何种状态。可执行此项以改善计算机的可靠性。例如,如果将某些类型的计算机在活动和不活动状态之间转变过多次,则它们可能经历负面的长期结果。因此,能量感知服务器管理策略可指示一个服务器在诸如一天的给定时间段内不应被转变超过给定数量的次数。系统可跟踪“机器 A”已被循环了多少次,并且如果机器 A 已被过于频繁地循环,则选择转变不同的机器。将在关于图 2 及在题为“状态转变示例”下详细描述能量感知服务器管理策略的各种方面。

[0027] 系统配置的示例

[0028] 图 2 示出可实现上述能量感知服务器管理的仅一个示例的系统 200。在此情形中,系统 200 包括负载分派器 202、自适应控制单元 204 以及一组四个服务器 206(1)、206(2)、206(3) 和 206(n)。负载分派器 202、自适应控制单元 204 和服务器 206(1)、206(2)、206(3) 和 206(n) 在各种实现中可以彼此搭配和 / 或分散。通常在 210 指示的用户可作出由负载分派器 202 接收的请求 212(1)-212(n)。负载分派器根据能量感知服务器管理策略 216 和活动机器列表 218 将各个请求 212(1)-212(n) 发送到如 214(1)-214(n) 所示的各个服务器 206(1)-206(n)。能量感知服务器管理策略 216 和活动机器列表 218 由自适应控制单元 204 生成并且发送到负载分派器 202。

[0029] 活动机器列表 218 反映能量管理控制决策 220,由自适应控制单元 204 传达给服务器 206(1)-206(n)。能量管理控制决策可被认为是通过控制各个服务器 206(1)-206(n) 的状态实现服务器级的能量感知服务器管理策略的控件。自适应控制单元 204 基于各种输入参数生成能量感知服务器管理策略 126 和活动机器列表 218。在此实例中,输入参数包括但不限于性能计数器 222、请求信息 224、能量使用 226、和 / 或输入请求工作负载(未具体示

出)。

[0030] 可将性能计数器 222 从服务器 206(1)-206(n) 发送到自适应控制单元 204。性能计数器可指示各个服务器和 / 或服务器的集体集合操作的容量。

[0031] 可从负载分派器 202 获取诸如响应时间的请求信息 224。可以若干方式表现请求信息, 诸如请求速率 (即每单位时间的请求)、请求大小、和 / 或请求响应时间等。

[0032] 可从各种源获取能量使用 226, 一个简单的源是服务器 206(1)-206(n) 的电源 (未示出) 上的能耗计量。尽管未具体示出, 但其他实现可单独或结合整体能量使用来为各个服务器提供能量使用。自适应控制单元还可接收用于未具体示出的源的信息。例如, 自适应控制单元可获取在服务级协定 (SLA) 中定义的性能信息。该性能信息可用于对能量感知服务器管理策略 216 的制定。此外, 自适应控制单元可仅将能量感知服务器管理策略 216 的特定部分发送到负载分派器 202。例如, 能量感知服务器管理策略可包括被发送到负载分派器以定义如何将请求分派到服务器的分派逻辑。

[0033] 在某些实例中, 负载分派器 202 可用作负载平衡器, 并将请求 212(1)-212(n) 均匀地分发给活动服务器 206(1)-206(n)。在其他实例中, 相对于活动服务器, 负载分派器可以以相对高的频率将请求策略地分发给活动服务器的子集, 从而将那些服务器保持在特定的容量, 例如全部容量的 70%。将会导致服务器的该子集超出特定容量的任何请求可被发送到活动服务器的第二子集。这样的配置可保持大多数活动服务器操作于特定的或期望的容量, 且在保留角色中操作第二子集, 从而改善能量效率。从一个角度来看, 活动服务器的该第二子集可被认为是针对意外的工作负载尖峰的预防。

[0034] 此外, 基于从自适应控制单元 204 获取的信息 (即, 能量感知服务器管理策略 216 和 / 或活动机器列表 218), 负载分派器 202 可对负载进行重新分发以远离正转变为不活动状态的任何服务器。

[0035] 总而言之, 在某些实现中, 负载分派器 202 可将请求仅在活动服务器之间进行分发并对负载进行重新分发以远离正转变为不活动状态的服务器。

[0036] 在系统 200 的所示配置中, 自适应控制单元 204 可预测未来的工作流 (即未来的请求)。自适应控制单元可制定以能量高效的方式处理未来工作负载的能量感知服务器管理策略。自适应控制单元在制定能量感知服务器管理策略中可考虑多个参数。这些参数可包括但不限于请求信息 224、能量使用 226、和 / 或性能计数器 222。

[0037] 总而言之, 自适应控制单元 204 可经由能量管理控制决策 220 根据能量感知服务器管理策略来控制服务器 206(1)-206(n) 的状态。在所示配置中, 自适应控制单元可运行于相对简单的闭环控制场景中, 因为它可经由基于所接收的输入生成的输出来控制系统功能。

[0038] 自适应控制单元通过接收与系统性能相关的输入数据来工作, 诸如仅举数例, 服务器中央处理单元 (CPU) 利用率、请求响应时间和能耗。可鉴于所预测的未来工作负载考虑输入数据, 以确定如何处理所预测的未来工作负载。例如, 调整处于每种能量状态的服务器的数量以平衡容量与所预测的工作负载, 并因此节省能量。换言之, 自适应控制单元可规定服务器 206(1)-206(n) 之间的能量状态从而以能量高效的方式处理所预测的工作负载。

[0039] 在某些情形中, 服务器 206(1)-206(n) 可采用被设计用于非服务器相关用途的处理器。例如, 本发明概念可允许将设计用于上网本和其他移动应用程序的处理器有利地应

用于服务器场景中。例如,可在当前配置的某些中利用 Intel 品牌的 Atom 处理器。Atom 处理器在活动状态中是相对能量高效的,并且具有相比活动状态带来约 90% 能量节省的不活动睡眠状态。

[0040] 不活动状态通常提供用于转变到活动状态的不同的等待时间。例如,真正的“关机”状态相比活动状态可带来高达 99% 的能量节省。但是,大量的能量节省是以相对长的转变等待时间为代价的。例如,从关机不活动状态到活动状态的转移等待时间可能是几分钟。作为对比,“睡眠”不活动状态可能带来例如 90% 的能量节省,但转变等待时间例如是 10-20 秒钟。这样长的转变的潜在的不利方面是如果工作负载预测实际上低估了未来工作负载,则关机的服务器将相对长时间地无法提供帮助。

[0041] “休眠”不活动状态可带来介于睡眠与关机不活动状态之间的能量节省和转变等待时间。例如,与活动状态相比,休眠可带来 95% 的能量节省和 30 秒钟的等待时间转变。注意在某些表征中,休眠被表征为深度睡眠。关于上述 Intel 的 Atom 处理器,活动状态能耗是 28-34 瓦特,但在睡眠状态中其消耗 3-4 瓦特,且休眠消耗 1-2 瓦特,在空闲处理器的能耗中二者至少降低了十倍。出于讨论目的提供上述不活动状态示例。应当理解,本发明概念不绑定于可由特定的服务器计算机提供的不活动状态的特定集合。换言之,可将节能功能应用于可被降低功率和 / 或断电的任何系统、设备、和 / 或组件。

[0042] 往往在某些产品线中获得具有多种不活动状态的计算机,诸如笔记本电脑。提供多种不活动状态的产品线往往被设计用于独立的应用程序,诸如个人使用,而不是诸如服务器场的协调式应用程序。协调式应用程序中缺少这些特征的一个原因在于以高效和节能的方式控制各种计算机的复杂度 / 难度。

[0043] 在单个使用场景中平衡性能和能效的复杂度甚至是无法忍受的。例如,如果笔记本电脑超时 (即,如果未接收用户输入达到预定义时间段) 则简单地被降低功率至不活动状态。超时方案不以任何方式试图预测未来的使用并且不以一种以能量意识方式满足未来的使用或对用户高效的方式来操作计算机。这样,当用户返回到已转变到不活动状态的这样的计算机时,用户必须提供某些输入并在计算机转变回到活动状态时进行等待。

[0044] 能量感知服务器管理策略 216 可允许在诸如服务器场的集体设置中实现多状态计算机的潜在能效。简言之,能量感知服务器管理可确定何时在不活动状态和活动状态之间转变各个服务器 206 (1)-206 (n) 以服务当前和所预测的未来工作负载。

[0045] 图 3 示出用于实现能量感知服务器管理的自适应控制单元 204 的逻辑模块的示例。在此情形中,逻辑模块被表现为工作负载预测模块 304、策略或策略模块 306, 和状态调整模块 308。

[0046] 工作负载预测模块 304 可确定当前的工作负载并预测未来的工作负载。从一个角度,工作负载可被认为是将每单位时间的请求数量和请求的大小相乘。在涉及相对大量的请求的操作场景中,可以以相当高度准确度地计算平均请求大小。在那些情形中,工作负载可被简单地计算为将每单位时间的请求数量与平均请求大小相乘。工作负载预测模块可利用各种技术来预测未来的工作负载,上面介绍了各种技术中的一些,下面将更详细讨论其中的一些。简言之,在图 2 的系统中,工作负载预测模块 304 可利用请求信息 224 作为用于预测未来工作负载的输入数据。在其他实例中,工作负载预测模块可接收并利用外部信息。在某些实现中,工作负载预测模块可收集、建模并存储用于估计或预测未来工作负载的长

期历史趋势。

[0047] 策略模块 306 可制定用于满足由工作负载预测模块 304 提供的工作负载预测的能量感知服务器管理策略。例如,策略模块可将当前容量计算为活动服务器的数量和在规定时间内可由活动服务器处理的请求的数量。策略模块可计算用于所预测的工作负载的合适的容量。总而言之,能量感知服务器管理策略可包括处理所预测的工作负载的容量供应。

[0048] 策略模块 306 可制定用于将服务器从当前配置重新配置为用于所预测的工作负载的合适的容量的能量感知服务器管理策略。例如,考虑涉及具有 10 个服务器的服务器场的假设场景,每个活动服务器每秒钟可处理 10 个请求。当前的工作负载是正由 7 个活动服务器处理的每秒钟 67 个请求。工作负载预测指示在未来的时间 $t + \Delta t$ 时工作负载将是每秒钟 38 个请求。在那种情形中,策略模块可制定减少活动服务器数量的能量感知服务器管理策略。

[0049] 在制定能量感知服务器管理策略时,策略模块 306 可在成本效益分析中包括多个因素。例如,策略模块可考虑使各个服务器活动和不活动多少次从而优化服务器寿命。服务器的寿命往往与在活动和不活动能量状态之间转变的次数成反比。另一个因素可以是定义系统 200 的性能级别的 SLA。可被考虑的另一个因素是工作负载预测中的置信度级别。例如,在置信度级别为高的实例中,那么策略模块可推断四个活动服务器将足以用于所预测的工作负载,并且包括使另外三个服务器不活动已达到总共六个不活动服务器。如果置信度级别为低,则策略模块可制定维护五个活动服务器的能量感知服务器管理策略。

[0050] 在某些实现中,策略模块 306 可在制定能量感知服务器管理策略中采用可包括诸如 CPU、存储器、磁盘、网络等的不同参数或度量的性能学习模型。此外,为便于解释,以上讨论已对全部服务器进行相同的处理。但是,策略模块可制定识别和 / 或指出服务器之间的区别或非均匀性的能量感知服务器管理策略 216。区别可以是服务器内部和 / 或外部的。例如,能量感知服务器管理策略可反映出各个服务器具有不同的容量、不同的电量占用量 (footprint) 和在能量状态之间不同的转变时间。类似的,服务器可受诸如温度热点、修复和维护任务的外部因素的影响。能量感知服务器管理策略还可指定将各个服务器转变到哪些不活动状态。例如,由于睡眠不活动状态提供回到活动状态的更快的转变,能量感知服务器管理策略可指定应将某些服务器保持在睡眠状态而将其他服务器转变到带来更多能量节省的不活动状态。

[0051] 能量感知服务器管理策略可使用附加的因素来决定服务器转移。例如,对于诸如软件升级的修复 / 维护任务,将活动服务器转变到关机而将处于不活动状态的另一个服务器转变到活动从而不影响系统容量。

[0052] 状态供应模块 308 可使对各个服务器的状态改变与由策略模块 306 提供的能量感知服务器管理策略保持一致。此外,状态调整模块可传达哪些服务器对负载分派器 202 (图 2) 是活动的,以避免负载分派器向不活动服务器发送请求。

[0053] 工作负载预测示例

[0054] 在一个实现中,可利用加权线性回归来实现工作负载预测。在此情形中,加权线性回归可被认为是一种形式的回归分析,其中使用一个或多个独立变量的数学函数来表示观察。例如,观察可以是诸如每秒钟的请求的从属变量。独立变量的示例可以是诸如时间的输入。对于加权线性回归,函数可被表达为回归系数的线性组合。加权线性回归的目标可

以是降低（并且可能地最小化）观察和由函数所预测的值之间的平方的加权和。权重用于向不同的观察分配不同的重要性值。总之，此特定的工作负载预测方法将回归分析与历史趋势、长期模型、外部知识等相结合来预测未来工作负载。

[0055] 容量供应示例

[0056] 在某些特定的示例中，可利用性能学习模型中的机器学习来实现容量供应。在一个这样的情形中，容量供应采用分位数回归的机器学习方法，在分位数回归中，响应时间的条件分布的分位数被表达为输入工作负载（每秒钟的请求）和活动服务器的数量的函数。

[0057] 此分位数回归方法考虑了如下事实，当服务器利用率较低并且随着输入工作负载接近服务器的处理能力而快速增长（例如，可能指数地）时，响应时间几乎线性地增长。换言之，容量供应方法可通过改变输入工作负载参数和活动服务器的数量以离线的方式获知系统的响应时间。随后，对于给定的 SLA 要求和对所预测的工作负载的估计，它计算满足所指定的 SLA 中的工作负载所需的服务器的数量。

[0058] 状态转变示例

[0059] 图 4 提供用于根据能量感知服务器管理策略实现状态转变的方法 400 的流程图。方法 400 始于框 402。在框 404，方法对服务器进行排序（即，服务器场的服务器或者另一组服务器）。框 404 可根据一个或多个参数对服务器进行排序。例如，可根据能量状态（即，活动或不活动）对服务器进行排序。此外，可在反映其相对能量节省的已排列组中对不活动服务器进行排序。例如，可列出活动服务器，随后跟着处于睡眠状态的不活动服务器，随后是处于休眠状态的不活动服务器，并且最后是关机状态。

[0060] 框 406 对能量状态进行排序。在一个情形中，可基于在不活动和活动状态之间的转变等待时间对能量状态进行排序。另选地或另外地，可基于其能量使用或占用量对能量状态进行排序或排列。因此例如，可从耗能最多（即活动）到耗能最少（即关机）列出能量状态。

[0061] 框 408 计算在未来时间 (t_f) 的所预测的服务器需求（即工作负载）与在时间 t_f 的可用容量之间的差值或 Δ 。以上描述了用于预测工作负载的技术。总而言之，框 408 用于识别在某些未来时间可用的服务器容量是否正好、过高或过低。

[0062] 如果在框 408，差值等于零（即所预测的工作负载等于未来的容量），则方法继续至框 410。框 410 此时停止控制服务器的过程，因为不需要容量调整。在诸如 30 秒钟的某些增量时间之后，方法随后返回至开始 402 并且重复该方法。

[0063] 如果在框 408，存在差值（即差值不为零），则方法继续至框 412。

[0064] 框 412（从框 404 引入的全部服务器中）识别在当前时间 t 和未来时间 t_f 之间的过渡期间转移状态的一组服务器。换言之，框 412 标识正在状态之间进行转变和 / 或已被调度如此进行的服务器。在框 408 的差值是正数（即 $\Delta > 0$ ）的示例中，那么方法继续至框 414，否则，如果差值为负（即 $\Delta < 0$ ），则方法继续至框 416。

[0065] 在框 414，方法通过标识转变到活动状态的不活动服务器来增加用于时间 t_f 的服务器容量。以上在框 404 列出各个服务器的状态。考虑用于标识的不活动服务器不属于在 412 识别的集合。换言之，所标识的服务器是未在过渡期间进行转变或被调度转变的不活动服务器。

[0066] 可利用各种其他参数来选择所标识的服务器。例如，可以考虑转变等待时间。例

如,如果未来时间 t_f 是 30 秒钟之后,则方法可考虑转变等待时间作为转变服务器的考虑。例如,如果从睡眠到活动的转变等待时间是 20 秒钟并且从关机到活动的转变等待时间是 1 分钟,则方法可从睡眠状态中选择服务器从而在未来时间 t_f (即 30 秒钟) 完成转变。

[0067] 方法还可考虑各个服务器在给定时间段内被转变的次数的数量。例如,如果睡眠状态的服务器之一已经转变了阈值数量的次数,则方法可选择转变低于阈值的另一个睡眠状态的服务器。在成本效益分析中,可在标识过程中利用其他参数。例如,成本可包括由频繁的状态转变等导致的增加的响应时间和降低的服务器寿命的风险。效益可以是降低的能量使用等。

[0068] 框 416 解决在未来时间 t_f 服务器容量大于所预测的工作负载的情形。框 416 通过对转变到不活动状态的活动服务器进行排列来降低用于时间 t_f 的服务器容量。被考虑的活动服务器不包括可能已经正在转变的在框 412 集合 $\{ts\}$ 中所标识的那些。可基于利用成本效益分析考虑的一个或多个参数来选择活动服务器。以上关于框 414 描述了参数的示例。简言之,如果给定服务器已经转变了阈值数量的次数,则可对成本收益分析进行加权以防止再次转变那个服务器。作为替代,选择转变另一个服务器。

[0069] 可被考虑的另一个参数涉及预防不足的容量。例如,假设每次转变服务器花费 20 秒钟并且一旦启动过程就不能将其停止。这样,一旦转变开始使活动服务器(称为服务器 1) 重新被激活要花费 40 秒钟(即,从活动到不活动要 20 秒钟并且从不活动到活动还要 20 秒钟)。出于讨论目的假设 t_f 等于未来 30 秒钟。现在如果另一个不活动服务器(称为服务器 2) 可用于被转变到活动状态,则对于前 10 秒钟与转移服务器 1 相关联的风险是相当低的,因为只要时间 t_f 是未来至少 20 秒钟,那么就可以及时地激活服务器 2。因此例如,如果作出使服务器 1 不活动的决策并且 8 秒钟后检测到涉及时间 t_f 的低估,则可及时地激活服务器 2 以覆盖附加的容量要求(即,8 秒钟加上 20 秒钟转移在 28 秒钟或者容量可能不足前 2 秒钟时将服务器 2 置于活动状态)。这样,服务器 2 的可用性可使对于使服务器 1 不活动的成本效益分析发生倾斜。

[0070] 另一个类似的参数涉及服务器 1 本身覆盖任何低估的能力。例如,继续以上示例,假设对于 40 秒钟的活动到活动的转变,服务器 1 每次花费 20 秒钟来转变。考虑未来时间 t_f 是 30 秒钟之后的第一情形。在这样的情形中,如果检测到低估,则服务器 1 无法在未来时间 t_f 覆盖附加的要求。在那种情形中,此参数可对不转变服务器 1 进行加权。考虑未来时间 t_f 是 60 秒钟以后的另一情形。在这样的情形中,可使服务器 1 不活动并且一旦检测到低估就在时间 t_f 之前(即在 40 秒钟以内)使之重新活动。在此情形中,此参数可对使服务器 1 不活动进行加权。技术人员应该认识到此成本效益分析可考虑提供预期的服务级别且节省能量的很多参数。

[0071] 对于框 410,在诸如 30 秒钟的某些增量时间之后,方法随后返回至开始 402 并且重复该方法。

[0072] 算法 1 示出采用成本效益分析来生成能量感知服务器管理策略的的部分的特定的状态转变方法。

[0073] 算法 1

[0074] 步骤 0. 根据处于活动状态的服务器被排序在处于不活动状态的服务器之前的序列对服务器进行排序—称为 S 序列。

[0075] 按照从该能量状态到活动状态的转移等待时间递增的顺序对能量状态进行排序—称为 E 序列。

[0076] 按照能量占用量递增的顺序对能量状态进行排序—称为 P 序列。

[0077] 步骤 1. 计算在所预测的服务器需求（称为 P 服务器）和在预先确定的未来时间步的可用的服务器数量（称为 F 服务器）之间的差值。

[0078] （在预先确定的未来时间步可用的服务器的数量依赖于当前活动的服务器的数量、当前处于转变的服务器的数量以及在预先确定的时间步之前将处于转变的服务器的数量。）

[0079] 步骤 2. 如果差值为零, 停止（称差值为 $D = P - F$ ）。

[0080] 步骤 3. 从 S 序列中取出前 T 个服务器并将其转变到 $T = \text{MINIMUM}(P, F)$ 的活动状态（如果不是已经活动或正转变到活动）。对于 T 个服务器中的每一个 i, 设置 $\text{NoChangeUntilTime}[i]$ 为预先确定的未来时间步。

[0081] 步骤 4. If D 是正的（即, 添加更多容量）

[0082]

then

按照 E 序列递增的顺序从 S 序列中取出接下来（相对于步骤 3）的 D 个服务器

For 每个这样的服务器（如 s）

设置 $\text{NoChangeUntilTime}[s]$ 为预先确定的未来时间步

if（未来时间步—当前时间>从该能量状态转变到活动）（即仍剩余某些时间）

在未来时间（未来时间步—当前时间）调度服务器 s 到活动状态的转变

else

现在调度服务器 s 到活动状态的转移

end if

end for

else

[0083]

If D 是负的（即, 更低的容量）

then

[0084] 从 S 序列中取出接下来（相对于步骤 3）的处于活动状态的 D 个服务器并且对于每个这样的服务器（如 j), $\text{NoTransitionUntilTime}[j] \leq \text{当前时间}$

[0085] 步骤 4. A. For P 序列中的每个能量状态

[0086]

if (未来时间步 - 当前时间 $\geq$ 从该能量状态转移到活动 + 从活动转移到该能量状态)

现在调度 D 进入该能量状态; 停止

else if (未来时间步 - 当前时间 $\geq$ 从该能量状态转移到活动)

AND (存在处于该能量状态的 D 个数量的服务器)

then

现在调度 D 进入该能量状态; 停止

else

现在调度 (D - 处于该能量状态的服务器的数量) 进入该能量状态

D = D - 处于该能量状态的服务器的数量

GO TO STEP (转到步骤) 4.A (并考虑那个顺序中的下一个能量状态)

end if

end for

end if

end if

[0087] 总而言之, 容量供应方法采取抢先的方法将服务器转变到不同的能量状态从而屏蔽这些转变和延迟转变的等待时间, 以便降低能量使用且仍然满足负载需求。

[0088] 第二, 实现应用转变且考虑了诸如每个服务器的转变数量的历史、热点、修复 / 维护动作、失效等的附加因素。例如, 为了平衡跨节点的转移的数量, 它可将某些活动服务器置于低能量状态, 反之亦然。

[0089] 方法示例

[0090] 图 5 示出了与本发明概念的至少某些实现一致的方法或技术 500 的流程图。描述方法 500 的次序并不旨在解释为限制, 并且任何数量的所述框都可以按任何次序组合以实现本方法或实现替换方法。此外, 该方法还可以用任何合适的硬件、软件、固件或其组合来实现, 以使得计算设备可实现该方法。在一种情况下, 该方法作为指令集被存储在计算机可读存储介质上, 以使得计算设备的执行使得该计算设备执行该方法。

[0091] 在框 502, 方法预测对服务器池的多个未来请求, 该服务器池包括被配置具有活动状态和至少两个不同的不活动状态的各个服务器。

[0092] 在框 504, 方法将未来请求的响应时间建模为未来请求的数量和活动服务器的数量的函数。

[0093] 在框 506, 方法计算满足响应时间的运行于活动状态和不活动状态中的每一个的服务器的数量。

[0094] 在框 508, 方法使得各个服务器在状态之间进行转变以满足所计算的数量。

[0095] 以上所述方法可满足用户的期望（即，诸如通过满足 SLA 的条款）且降低服务器的能量使用。

[0096] 结论

[0097] 尽管已用对结构特征和 / 或方法动作专用的语言描述了涉及能量感知服务器管理策略的技术、方法、设备、系统等，但可以理解，所附权利要求书中定义的主题不必限于所述具体特征或动作。相反，上述具体特征和动作是作为实现所要求保护的方法、设备、系统等示例性形式而公开的。

方法 100

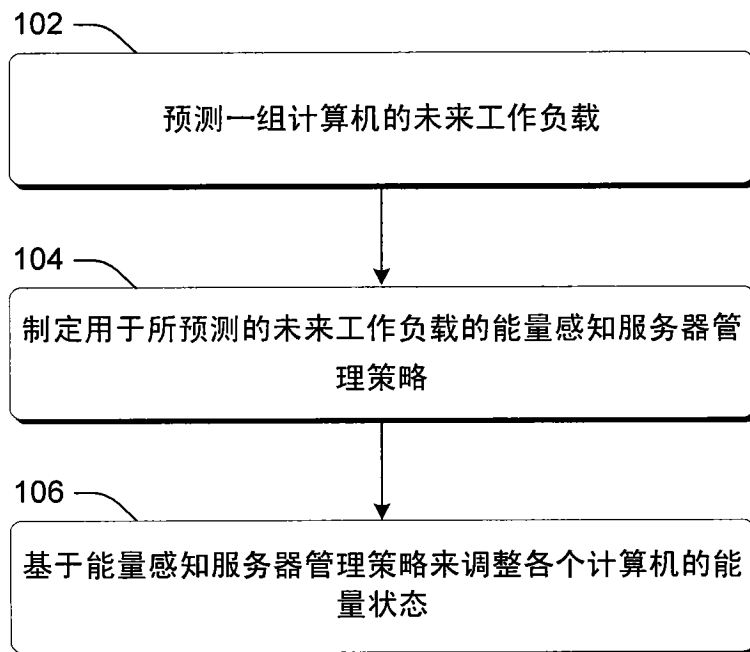


图 1

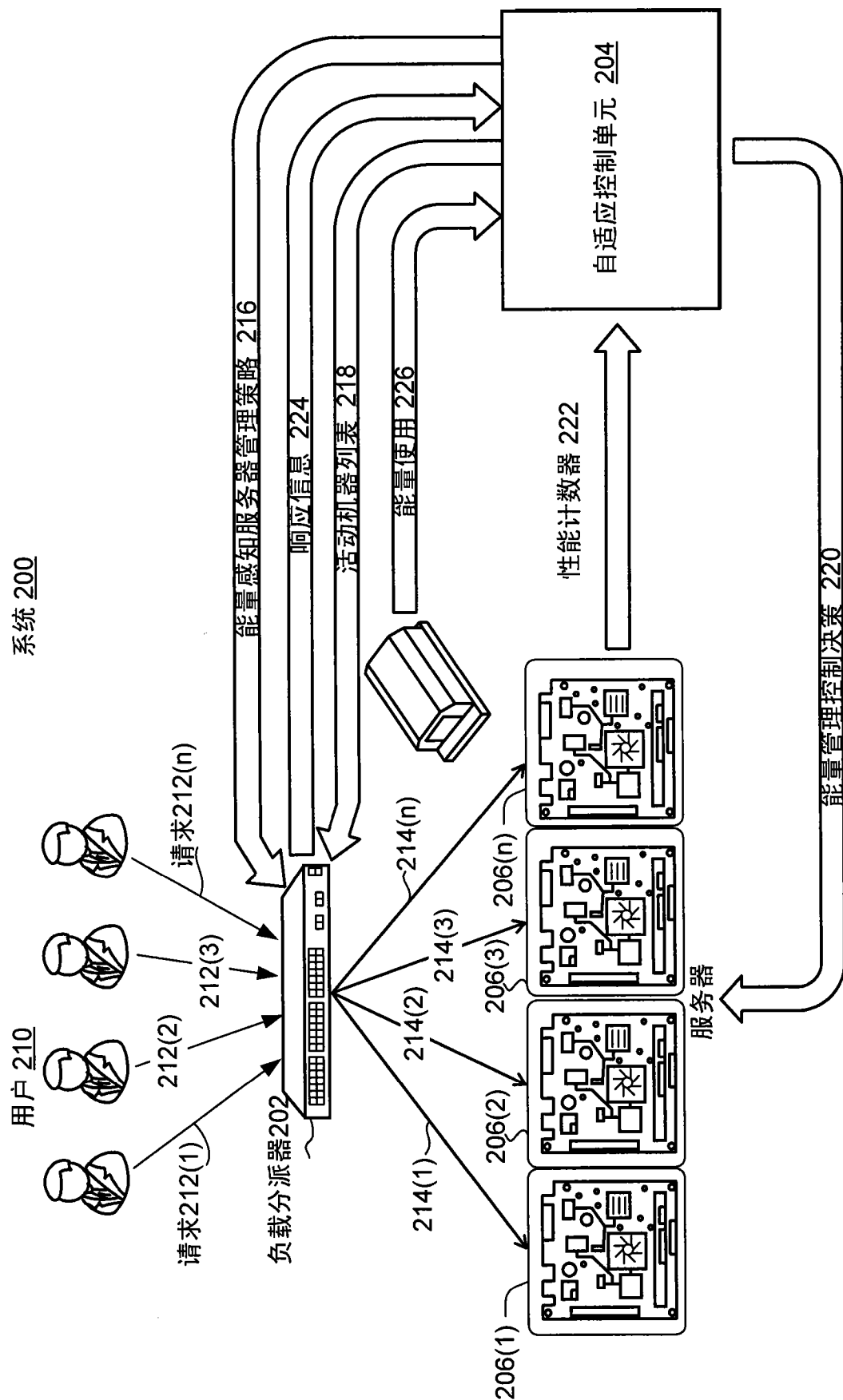


图 2

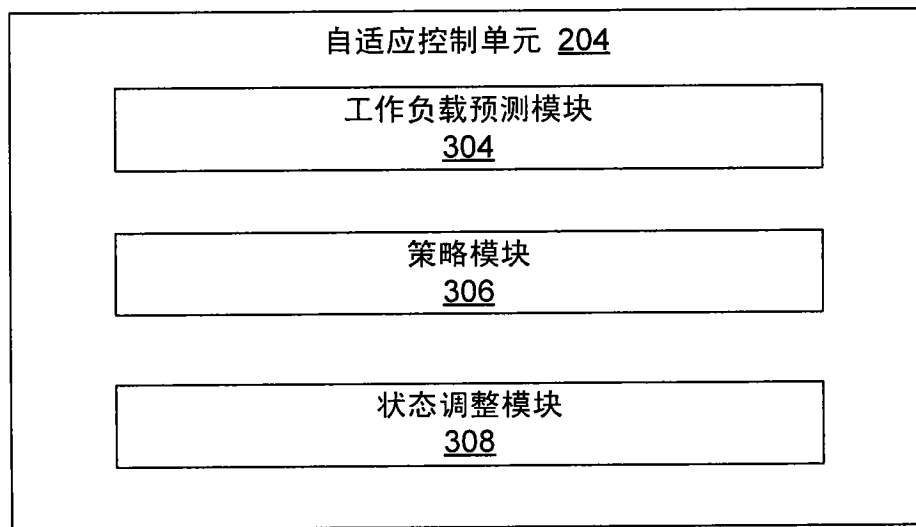


图 3

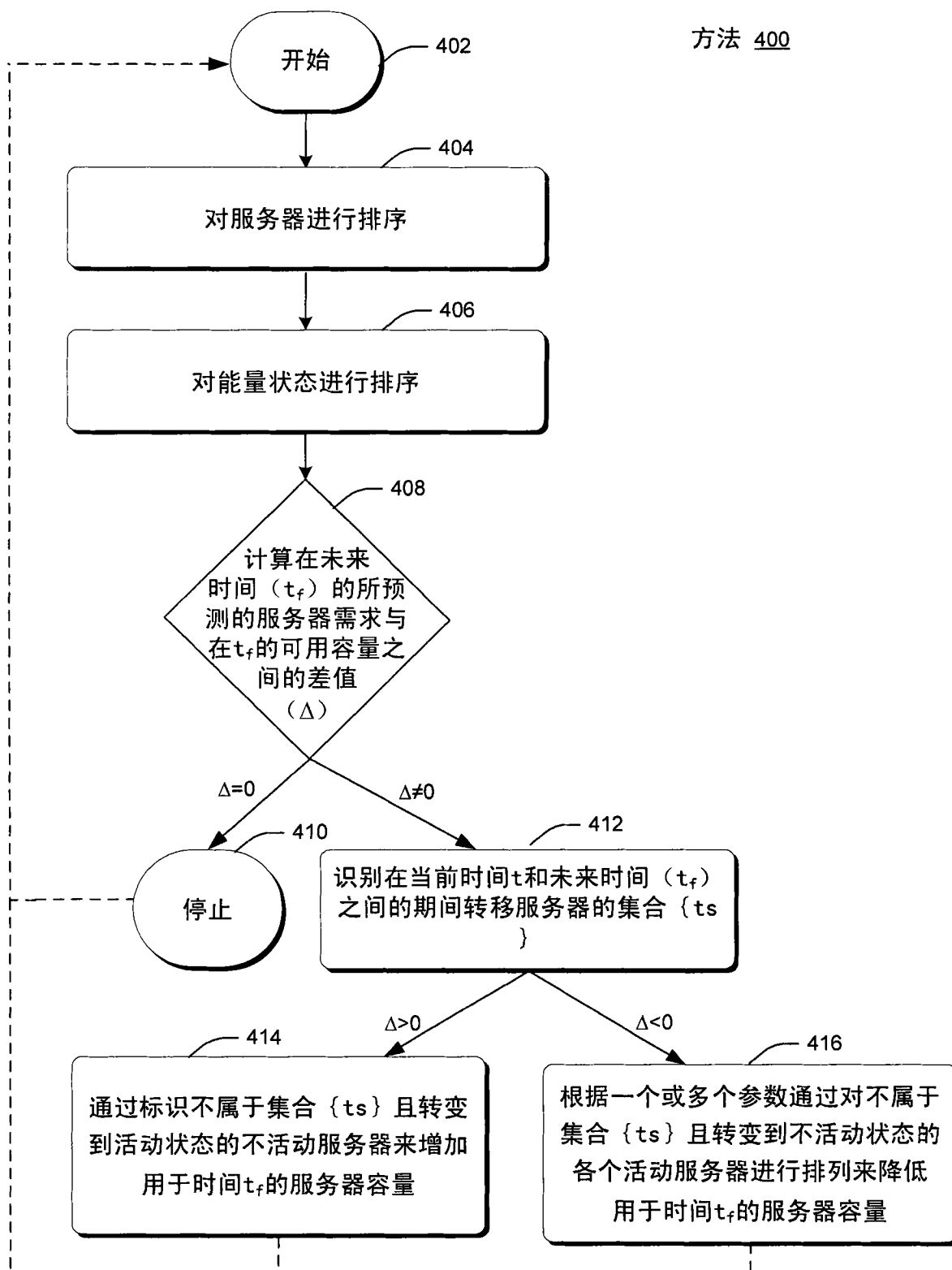


图 4

方法 500

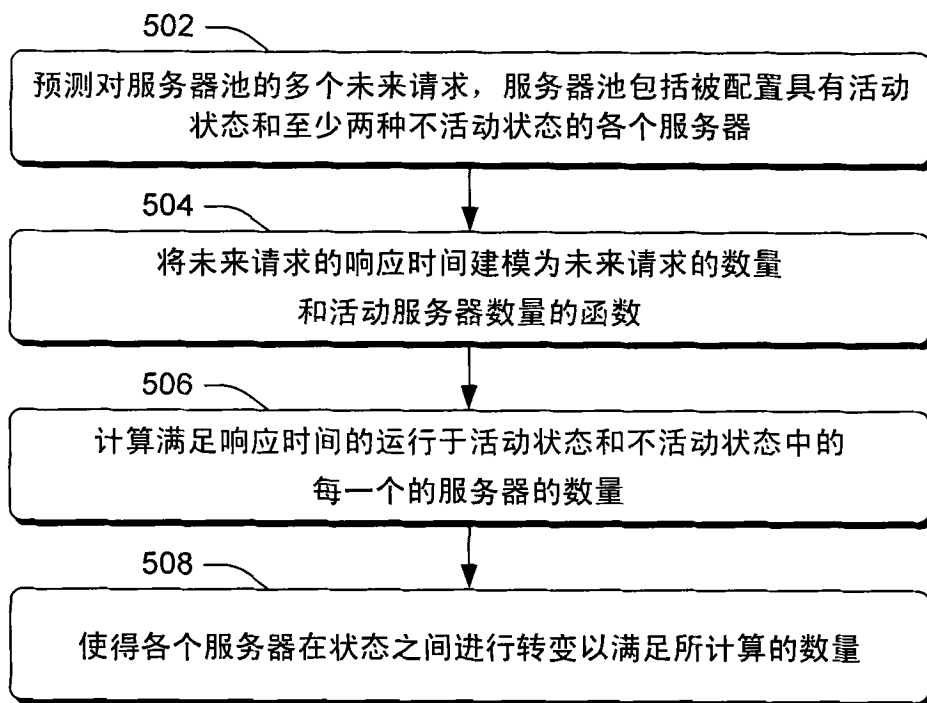


图 5