



(12) 发明专利申请

(10) 申请公布号 CN 103227809 A

(43) 申请公布日 2013. 07. 31

(21) 申请号 201210551788. 6

(22) 申请日 2012. 12. 18

(30) 优先权数据

11194246. 2 2011. 12. 19 EP

(71) 申请人 西门子公司

地址 德国慕尼黑

(72) 发明人 安德烈亚·博埃罗 达维德·科尔索

(74) 专利代理机构 北京集佳知识产权代理有限公司

11227

代理人 郎晓虹 李春晖

(51) Int. Cl.

H04L 29/08 (2006. 01)

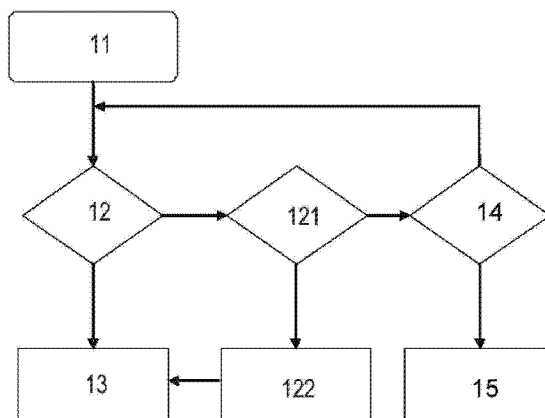
权利要求书2页 说明书5页 附图4页

(54) 发明名称

用于在不同客户端之间管理资源以用于专用的方法和系统

(57) 摘要

本申请公开了用于在不同客户端之间管理资源以用于专用的方法和系统。该系统包括：服务器，包括最大数量 m 个资源，其中，预定数量 w 个资源能够被初始化以准备用于客户端使用并且构成池；通信系统，用于与客户端通信；管理装置，用于管理资源且至少包括处理器和存储器，能够改变池的资源的状态、将新资源添加到池、终止池的资源，其特征在于，存储器包括如下指令：用于确定时刻 $T4$ 处池中包括的资源数量 p ，并且用于确定 p 个资源中在时刻 $T4$ 处为“使用中”的资源数量 p_u ；用于在且仅在时刻 $T4$ 处 $m \geq p > w$ 且 $p_u < w$ 时，终止池的 $p - p_u$ 个可用资源中的 $p - w$ 个资源；用于在时刻 $T4$ 处 $p \leq w$ 或 $p_u > w$ 的情况下，独立于时间地保持池的资源初始化。



1. 一种用于在不同客户端之间共享服务器的资源的方法,其中,所述服务器包括最大数量 m 个资源,在所述资源中,预定数量 w 个所述资源被初始化,并准备用于所述客户端使用,并且包括在池中,所述方法包括在时刻 T_4 释放(21) 服务器资源,包括将所释放的资源的状态从“使用中”改变(22) 为“可用”,所述方法特征在于:

c. 所述方法包括用于确定在所述时刻 T_4 处所述池的尺寸的池尺寸确定(23);并且

c1. 如果在所述时刻 T_4 处所述池包括 p 个资源,其中 $m \geq p > w$,则在且仅在所述时刻 T_4 处“使用中”资源的数量 p_u 低于所述预定数量 w 时,终止(24) 所述池的可用资源中的 $p-w$ 个资源;

c2. 否则,独立于时间地保持(25) 所述池的资源初始化。

2. 根据权利要求 1 所述的方法,包括向客户端分配(13) 资源,并且其中所述分配(13) 包括:

a. 所述服务器接收对于使用一个服务器资源的请求(11);

b. 确定(12) 所述池中对于所述客户端的可用资源数,并且

b1. 如果在所述池中至少一个服务器资源可用,则向所述客户端分配(13) 所述一个资源,并且将所分配的资源的从“可用”改变为“使用中”,否则,

b2. 如果所述池的每个资源均是“使用中”,则确定(121) 是否使用了所述最大数量 m 个资源,并且

b21. 如果没有使用所述服务器的所述最大数量 m 个资源,则初始化(122) 新资源,并且将所述新资源添加到所述池;

b22. 如果使用了所述服务器的所述最大数量 m 个资源并且如果没有到达超时(14),则返回到步骤 b;

b23. 如果使用了所述服务器的所述最大数量 m 个资源并且如果到达所述超时,则向客户端请求返回错误(15)。

3. 根据权利要求 1 或 2 之一所述的方法,包括根据所述客户端或请求的排序来优先化对所述客户端或请求的资源的分配(13)。

4. 根据权利要求 1 至 3 之一所述的方法,包括对客户端的资源分配的至少一个条件。

5. 一种用于在不同客户端之间共享服务器的资源的系统,其中,所述系统包括:

- 服务器,包括最大数量 m 个资源,在所述资源中,预定数量 w 个所述资源能够被初始化,以在池中准备用于所述客户端使用;

- 通信系统,用于与所述客户端通信;

- 管理装置,用于管理所述资源,其中,所述管理装置至少包括处理器和存储器,所述管理装置能够

• 根据时刻 T_4 处对所述客户端的资源的分配或资源的释放来改变所述池中的所述资源的状态,

• 将新资源添加到所述池,

• 终止所述池的资源;

其特征在于

- 所述存储器包括如下指令:

• 用于确定所述时刻 T_4 处所述池中包括的资源的数量 p , 并且用于确定所述 p 个资源

中、在所述时刻 T4 处为“使用中”的资源数量 p_u 的指令；

- 用于在且仅在所述时刻 T4 处 $m \geq p > w$ 且 $p_u < w$ 时终止所述池的 $p - p_u$ 个可用资源中的 $p - w$ 个资源的指令；

- 用于在所述时刻 T4 处 $p \leq w$ 或 $p_u > w$ 的情况下独立于时间地保持所述池的资源初始化的指令。

6. 根据权利要求 5 所述的系统, 其中, 所述服务器包括数据库, 所述数据库包括客户端和 / 或请求的优先化列表。

7. 根据权利要求 5 或 6 所述的系统, 其中, 所述服务器能够对请求的接收加时间戳。

用于在不同客户端之间管理资源以用于专用的方法和系统

技术领域

[0001] 本发明涉及共享资源在客户端之间的分配,具体地,涉及计算机科学领域。更精确地,本发明涉及如下方法和系统:其用于根据权利要求1和5在不同客户端之间管理或共享资源以用于专用,并且具体地用于在客户端-服务器通信网络中在不同客户端之间共享服务器资源。

背景技术

[0002] 随着时间的推移,工业自动化在范围和精细化方面已经提高。通常,工业自动化聚焦于包括多个相互作用的机器的连续处理。这具体涵盖西门子公司以其**SIMATIC®**产品族提供的制造执行系统的广泛范围。在这样的环境下,数据处理中的公共设置存在于分布式客户端工作机、以下为客户端中,其中客户端程序/应用通常从客户端-服务器系统、以下为服务器请求服务。因此,与技术和/或逻辑处理关系密切的至少一个客户端通过通信网络连接至少一个服务器。

[0003] 在客户端之间共享服务器的有限数量的资源以用于专用。用来实现一组共享资源的典型设计模式是对象池模式:初始化的服务器资源是“池”中准备用于使用的对象并且被分配给请求客户端。换言之,池的客户端可从池中请求对象(或资源),并且对所返回的对象(或资源)执行操作。当前分配给客户端的资源被标记为“使用中”并且不能分配给其它客户端。当使用资源的客户端不再需要它时,客户端将资源释放到池中并且资源被标记为“可用”。如果客户端在所有资源都在使用中时进行请求,则客户端通常等待,直到资源可用或到达超时为止。

[0004] 例如对于特定应用确定池中资源的正确的总量(已知为“池尺寸”)经常是关键选择:小的池尺寸将导致在高峰期期间更长的等待(和频繁的超时),而大的池尺寸可能超出系统能力。池的初始尺寸具体取决于客户端的数量和请求分布。可以通过使用例如经验“猜测和检验”迭代序列来根据经验地确定池的初始尺寸。因此,随时间推移池的资源的管理取决于所使用的技术和方法。例如,Oracle 数据库资源管理器能够在不考虑系统的负载和客户端的数量的情况下为一些客户端预分配最小量的资源,并且以例如一些任务优先于其它任务的方式分配可用资源。在对于资源的请求数量增加的情况下,其可以创建上至用于应对需求的最大值的共享资源,并且然后,如果在预定时间段内不再使用所创建的新共享资源,则其以时间为基础释放它们。

[0005] 由于在高资源需求期间,服务器例如使用其最大能力,因此在高峰期资源的高效管理特别重要。因此,管理资源的方式将确定随时间推移服务器可能执行的请求数量,并且因此降低客户端请求的接收和其执行之间的平均等待时间。

[0006] 因此,本发明的目的是提供一种用于随时间推移在不同客户端之间高效地管理资源的方法和系统。

发明内容

[0007] 通过用于在不同客户端之间共享服务器的资源的方法、根据关于该方法的本发明来实现该目的,其中,服务器包括最大数量 m 个资源,在该资源中,预定数量 w 个所述资源被初始化,并准备用于所述客户端使用,并且包括在池中,其中,如果资源被分别分配给客户端或对于客户端可用,则池的每个资源由两个状态“使用中”或“可用”表征,该方法包括向客户端分配资源以及在时刻 $T4$ 释放服务器资源,向客户端分配资源具体可包括以下步骤:

[0008] a. 服务器接收对于使用一个服务器资源的所述请求,其中,所述请求由所述客户端之一例如经由通信网络生成和发送;

[0009] b. 如果在所述池中至少一个服务器资源可用,则向所述客户端分配所述一个资源,并且将所分配的资源的从“可用”改变为“使用中”,否则,如果所述池的每个资源均是“使用中”,则确定是否使用所述最大数量 m 个资源,并且

[0010] b1. 如果没有使用服务器的最大数量 m 个资源,则初始化新资源,并且将所述新资源添加到所述池;

[0011] b2. 如果使用服务器的最大数量 m 个资源并且如果没有到达超时,则返回到步骤 b;

[0012] b3. 如果使用服务器的最大数量 m 个资源并且如果到达所述超时,则向所述客户端返回错误,以及

[0013] 在时刻 $T4$ 释放服务器资源包括将所释放的资源的从“使用中”改变为“可用”,并且其特征在于:

[0014] c. 如果所述池在时刻 $T4$ 包括 p 个资源,其中 $m \geq p > w$,则在且仅在所述时刻 $T4$ 处“使用中”资源的数量 p_u 低于预定数量 w 即 $p_u < w$ 时,终止所述池的可用资源中的 $p-w$ 个资源,其中,终止资源指的是服务器去初始化(de-initialize)所述资源;

[0015] d. 否则,独立于时间地保持所述池的资源初始化。

[0016] 还通过用于在不同客户端之间共享服务器的资源的系统、根据关于该系统的本发明来实现该目的,其中,所述系统包括:

[0017] - 服务器,包括最大数量 m 个资源,在该资源中,预定数量 w 个所述资源能够被初始化,以准备用于所述客户端使用,并且从而构成池,其中,如果资源被分别分配给客户端或对于客户端可用,则所述池的每个资源由两个状态“使用中”或“可用”表征;

[0018] - 通信系统,用于与所述客户端通信;

[0019] - 管理装置,用于管理所述资源,其中,所述管理装置至少包括处理器和存储器,所述管理装置能够

[0020] • 根据对客户端的资源的分配或资源的释放即资源返回到池来改变池中的资源的状态,

[0021] • 将新资源添加到池,

[0022] • 终止池的资源、即去初始化资源从而从池中移除该资源;

[0023] 其特征在于

[0024] - 存储器包括如下指令:

[0025] ○用于确定在时刻 $T4$ 处池中包括的资源的数量 p 并且用于确定所述 p 个资源中在时刻 $T4$ 处为“使用中”的资源的数量 p_u 的指令;

[0026] ○用于在且仅在所述时刻 $T4$ 处 $m \geq p > w$ 且 $p_u < w$ 时终止所述池的 $p-p_u$ 个可用

资源中的 $p-w$ 个资源的指令；

[0027] ○用于在所述时刻 $T4$ 处 $p \leq w$ 或 $p_u > w$ 的情况下独立于时间地保持池的资源初始化的指令。

[0028] 所提出的发明可以应用于其中必须在不同客户端之间共享有限数量的资源以用于专用的任何情况,并且因此不仅限于服务器和通信网络,而是可以被应用于各种领域。因此,术语客户端可不仅指的是客户端工作机、而指的是任何资源用户,而术语服务器可指的是任何资源提供者。在资源初始化和去初始化(deinitialization)时间相关的情况、即当如所需一样长地保持分配给请求客户端的一组初始化资源是有利的时,本发明特别有用。根据本发明的池是准备用于客户端使用的所述一组初始化资源。

附图说明

[0029] 将通过以下附图更好地理解本发明：

[0030] 图 1 是本发明的第一功能的示意性表示。

[0031] 图 2 是本发明的第二功能的示意性表示。

[0032] 图 3 是根据传统 P1、Oracle P2、以及本发明的 P3 池的、总执行时间相对于客户端的数量的曲线图。

[0033] 图 4 是根据传统 P1、Oracle P2、以及本发明的 P3 池的、平均等待时间相对于客户端的数量的曲线图。

具体实施方式

[0034] 图 1 和图 2 示意性地表示基本池接口反映的两个功能,分别是“获得”功能(图 1)和“释放”功能(图 2),其中,“获得”功能涉及根据传统现有技术(图 1A)和根据本发明(图 1B)向客户端请求分配资源,而“释放”功能涉及根据现有技术(图 2A)和根据本发明(图 2B)释放所分配的资源。获得和释放功能涉及池客户端之间的池资源的共享。

[0035] 在图 1A 中示出的传统现有技术中,资源提供者或服务器通常包括最大数量 m 个资源,在该资源中,预定数量 w 个所述资源保持初始化以便准备用于所述客户端或资源用户使用。初始化的资源包括在服务器的池中,并且用于根据现有技术来向客户端分配池资源的步骤如下：

[0036] a. 服务器接收对于资源的客户端请求 11,所述请求 11 请求向所述客户端分配服务器资源；

[0037] b. 服务器确定 12 所述池中的可用资源的数量,并且

[0038] b1. 如果对于所述客户端至少一个资源可用,则向客户端分配 13 所述服务器资源；否则

[0039] b2. 如果所述池的每个资源均在使用中,则确定是否到达超时,其中,超时是可以作为参数记录在服务器中的预定时限,并且此后被称为超时确定 14 的、确定是否到达超时的步骤包括：确定请求的接收和超时确定 14 间隔的时间段是否大于或等于所述预定时限,并且

[0040] b21. 如果没有到达超时、即如果请求的接收和超时确定间隔的时间段不大于所述预定时限,则返回到步骤 b,否则

[0041] b22. 如果到达超时,则向客户端返回错误 15。

[0042] 根据本发明,向现有技术添加新步骤以用于提高获得功能的效率。实际上,用于向客户端分配服务器资源的步骤如图 1B 所示并且如下:

[0043] a. 例如在时刻 T1 处,服务器接收对于资源的客户端请求 11,其中,具体地,服务器可以对所述请求加时间戳以用于向请求 11 分配时刻 T1,所述请求 11 请求向所述客户端分配服务器资源,所述请求具体地由客户端生成和发送并由服务器的通信装置接收,其中,服务器和客户端可例如借助于通信网络连接,该通信网络被配置用于将每个客户端连接至服务器;

[0044] b. 服务器确定 12 所述池中对于客户端的可用资源数,所述确定 12 例如在时刻 $T2 > T1$ 处执行,并且

[0045] b1. 如果对于所述客户端池的至少一个资源可用,则向客户端分配 13 所述一个资源,具体地,所分配的资源的状态由服务器从“可用”改变为“使用中”,否则,

[0046] b2. 如果所述池的每个资源均是“使用中”、即如果所述池中没有任何资源可用,则确定 121 是否使用了最大数量 m 个资源,并且

[0047] b21. 如果没有使用服务器的最大数量 m 个资源,则初始化 122 新资源,并且将所述新资源添加到池,其中,所述新资源具体地被分配给发起请求 11 的客户端;

[0048] b22. 如果使用服务器的最大数量 m 个资源,则如先前针对图 1A 所述执行超时确定 14,所述超时确定例如在时刻 T3 处执行并且被配置用于确定请求的接收和超时确定 14 间隔的时间段是否大于或等于预定时限 T_L 、即是否 $T3 - T1 \geq T_L$,并且

[0049] b221. 如果 $T3 - T1 < T_L$,则返回到步骤 b;否则

[0050] b222. 如果 $T3 - T1 \geq T_L$,则向客户端返回错误 15。

[0051] 优选地,服务器能够比各个其它请求或客户端优先化向某一请求或客户端的资源分配。为此,服务器可包括数据库,该数据库包括根据优先化方案对客户端或请求排序的优先化列表。有利地,根据本发明的方法具体包括根据所述客户端或请求的排序的、对客户端或请求的资源分配的优先化。例如,如果服务器在相同时刻 T1 处接收到两个请求,则较高排序的请求或较高排序的客户端的请求将比其它请求具有优先权,即,可用资源将优选地分配给较高排序的请求或较高排序的客户端的请求。通过相同方式,如果服务器接收到若干请求、而由于所有服务器资源在使用中的事实没有进行服务器资源的分配,并且如果尚未到达每个请求的超时,则根据本发明的系统具体地能够向在时间上第一接收的、较高排序的请求或较高排序的客户端的请求分配第一可用资源。

[0052] 当资源返回到客户端时,所述客户端然后可使用所述资源以用于执行一些操作。优选地,至少一个分配条件可与对客户的每个资源分配 13 相关联。然后,在向客户端分配资源的时间段期间控制每个分配条件,以确定是否遵守所述条件。分配条件的示例是将资源的分配时间限制于预定时间段。例如,如果客户端在大于所述预定时间段的时间段期间使用所述资源,则服务器自动释放资源。在所述客户端使用资源的时间段期间,所述资源对于其他请求、即对于其他客户端不可用。

[0053] 在对于资源的请求增加的情况下,传统池中包括的资源数可能不再足以高效地处理每个新进入的请求。有利地,为了向进入的请求提供高效的响应,根据本发明的系统能够将新资源添加到池,从而将新资源初始化为上至最大可用服务器资源。将新资源添加到池

使得能够增加单位时间处理的请求数。但是为了高峰期高效地处理请求,资源的释放也应是高效的。

[0054] 图 2A 示意性地示出了根据现有技术的释放功能。一旦客户端用完所述资源,则将资源返回到池。还具体地,一旦不遵守分配条件,则服务器释放资源并将资源返回到池。资源的标准释放包括两个主要步骤:释放 21 被返回到池的资源;并且然后将所释放的资源的状态从“使用中”改变 22 为可用。

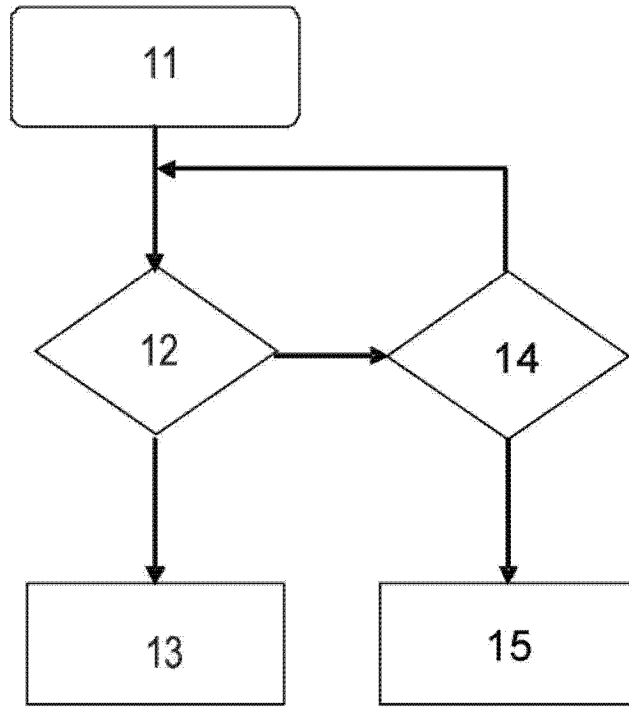
[0055] 图 2B 示出了根据本发明的用于释放资源的新方法。在时刻 T4 处释放 21 所分配的资源的情况下,资源返回到池并且方法还包括资源的状态从“使用中”改变 22 为“可用”,所述改变 22 由服务器执行,并且 c. 池尺寸确定 23 用于确定在时刻 T4 处池的尺寸、即在时刻 T4 处所述池中包括的资源数,并且

[0056] c1. 如果在时刻 T4 处池包括 p 个资源,其中 $m \geq p > w$,则该方法包括:在且仅在所述时刻 T4 处“使用中”资源的数量 p_u 低于预定数量 w (即,如果 $p_u < w$) 时,终止 24 所述池的可用资源中的 $p-w$ 个资源;

[0057] c2. 否则,该方法包括独立于时间地保持 25 所述池的资源初始化。

[0058] 在图 3 和 4 中强调了本发明与现有技术相比的优点。这些图呈现了传统池管理 P1、类 Oracle 池管理 P2、以及根据本发明的池管理 P3 之间的比较。图 3 示出了以客户端的数量为函数的总执行时间,而图 4 示出了以客户端的数量为函数的平均等待时间。传统池包括五个资源,即,服务器的资源的最大数量 m_{P1} 为 5 并且对应于初始化的资源的数量 w_{P1} 。类 Oracle 池包括 5 个预分配的资源,即 $w_{P2}=5$,其中服务器资源的最大数量 m_{P2} 等于 10 : $m_{P2}=10$ 。根据本发明的池包括数量为 5 的初始化资源,即, $w_{P3}=5$,其中服务器资源的最大数量 m_{P3} 等于 10 : $m_{P3}=10$ 。用于比较根据传统的、类 Oracle 的、以及本发明技术的池的管理,客户端根据以下统计数据向池请求资源:6 个客户端持续请求资源并且长时段使用资源以便总是使至少 5 个资源在使用中,并且所有剩下的客户端以“突发模式”请求资源、即短时段使用资源,其中,在两个连续突发之间存在大的间隙,并且类 Oracle 技术的超时略微短于两个连续突发之间的平均间隙。根据这种分布,图 3 和 4 中呈现的曲线图清楚地示出:主要归因于根据本发明资源的释放独立于时间的事实,对于相同数量的客户端,本发明允许较短的请求执行时间、以及在所请求的资源可用之前客户端经历的较短的平均等待时间。

A



B

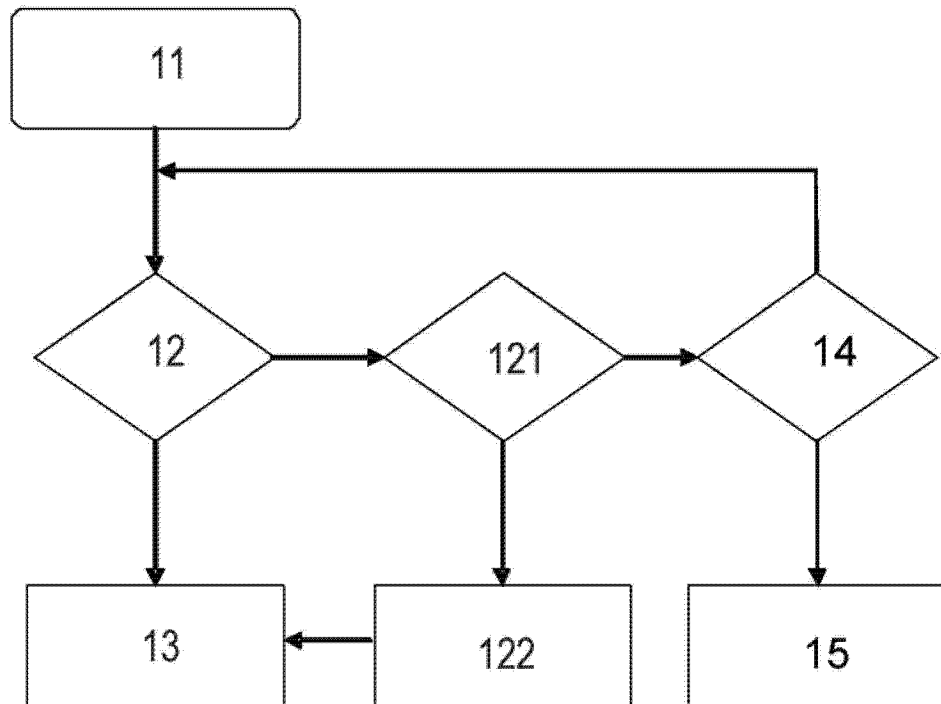
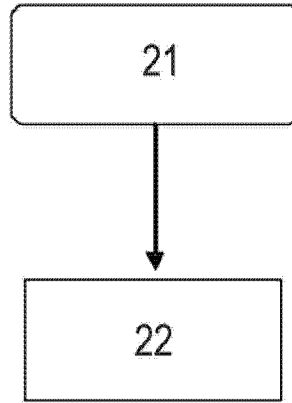


图 1

A



B

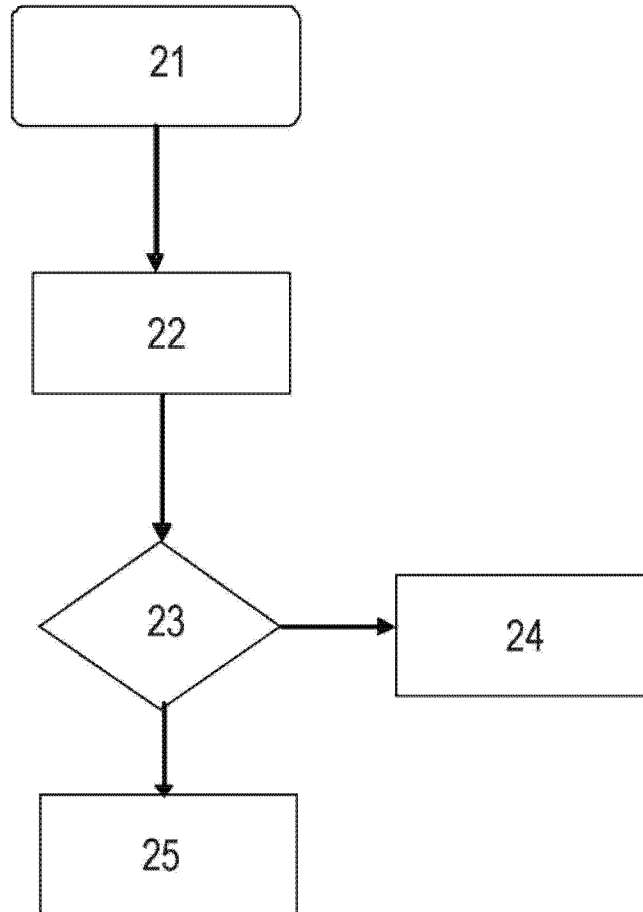


图 2

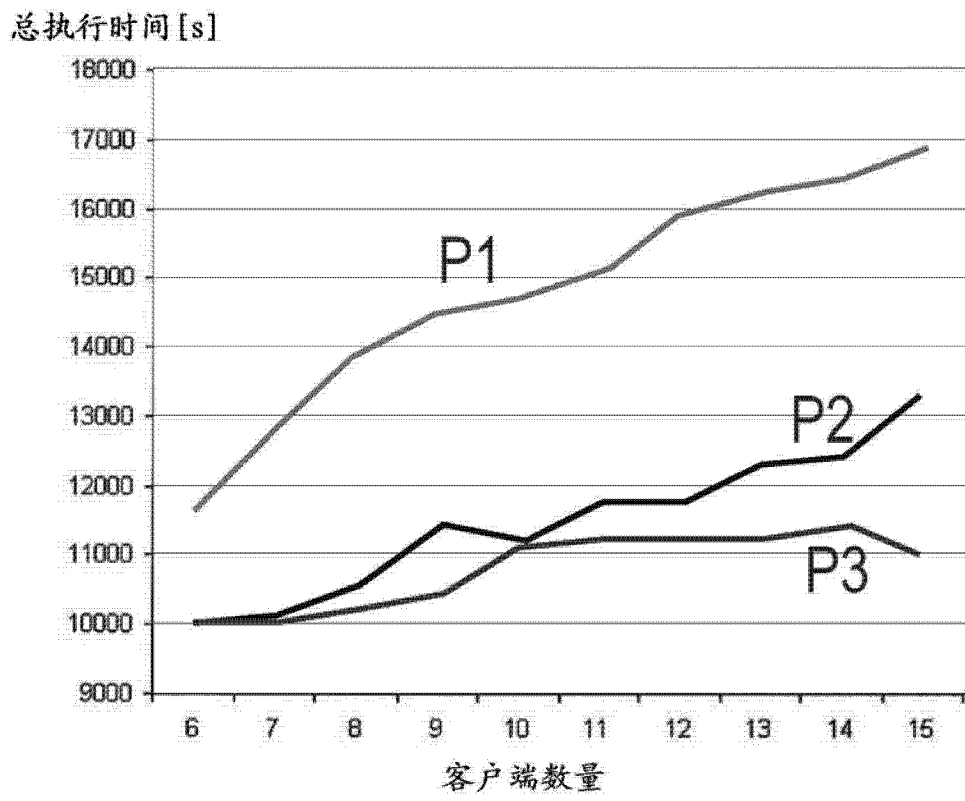


图 3

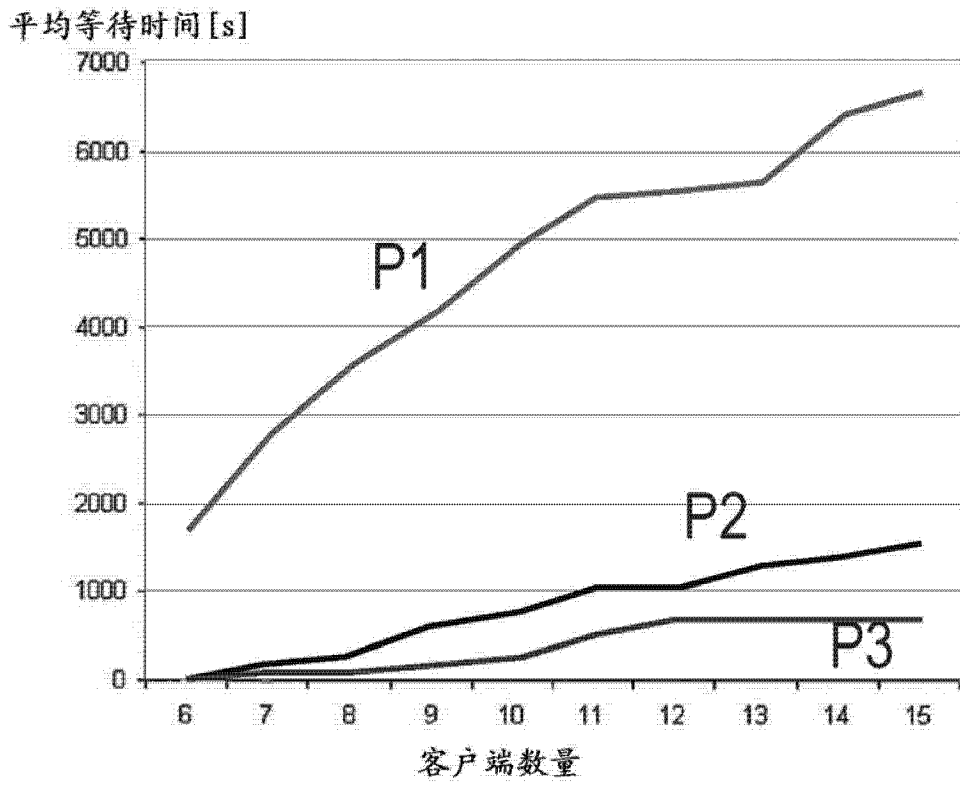


图 4