

(12) 发明专利

(10) 授权公告号 CN 1955932 B

(45) 授权公告日 2010.06.09

(21) 申请号 200610139223.1

Queueing Network Models with Histogram-Based Parameters. Proceedings of IPDS '98. 1998, 142-151.

(22) 申请日 2006.09.18

审查员 徐春

(30) 优先权数据

11/258, 435 2005. 10. 25 US

(73) 专利权人 国际商业机器公司

地址 美国纽约阿芒克

(72) 发明人 默西·德瓦拉康达

尼斯雅·拉杰马尼

马德哈卡·斯里瓦特萨

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 冯谱

(51) Int. Cl.

G06F 9/46 (2006.01)

(56) 对比文件

US 2005/0228856 A1, 2005. 10. 13, 第 36-111 段。

Johannes Luthi et al.. Interval
Matrices for the Bottleneck Analysis of

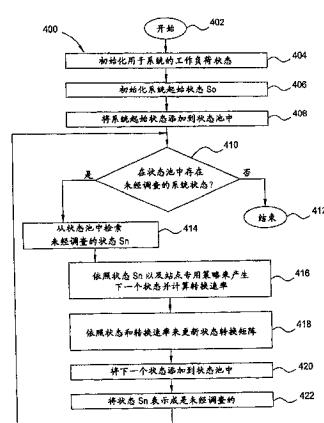
权利要求书 2 页 说明书 7 页 附图 4 页

(54) 发明名称

用于在分布式计算系统中的性能和策略分析的方法和装置

(57) 摘要

用于分布式计算系统中的性能和策略分析的方法和设备的一个实施例包括将分布式计算系统性能表示成一个状态转换模型。然后，在所述状态转换模型上叠加一个排队网络，并且根据所述排队网络的解来识别一个或多个策略施加于所述分布式计算系统的作用。



1. 一种用于对适合多个计算站点的一个或多个策略进行分析的方法,所述计算站点在分布式计算系统中对相应的工作负荷进行处理,所述方法包括:

将所述分布式计算系统表示成状态转换模型,其中将规定在计算站点之间共享资源的方式的策略建模成对状态以及状态转换的约束条件;

在所述状态转换模型上叠加一个排队网络模型;以及

根据所述排队网络模型的解来确定所述一个或多个策略施加于所述分布式计算系统的性能的作用。

2. 根据权利要求 1 所述的方法,其中所述表示包括:

根据表示所述多个计算站点的特性的至少一个模型来建造所述状态转换模型,其中所述至少一个模型包括下列模型中的至少一个:站点模型、工作负荷模型、工作负荷状态模型、站点状态模型、策略模型、事件模型或成本模型。

3. 根据权利要求 2 所述的方法,其中所述站点模型表示的是与相应的计算站点相关联的静态参数。

4. 根据权利要求 3 所述的方法,其中所述静态参数包括下列参数中的至少一个:与所述相应计算站点相关联的多个资源,一种所述的资源或一个或多个策略,其中所述策略规定的是所述相应计算站点用以与第二计算站点共享资源的方式。

5. 根据权利要求 2 所述的方法,其中所述工作负荷模型表示的是下列各项中的至少一项:由相应计算站点充当宿主的应用、所述相应计算站点提供的一个或多个服务等级协定、或转换概率矩阵,其中所述转换概率矩阵规定的是由所述相应计算站点处理的工作负荷如何在与处理所述工作负荷所需要的相应资源数量相关联的两个或多个等级之间转换。

6. 根据权利要求 2 所述的方法,其中所述工作负荷状态模型表示的是下列各项中的至少一项:与相应计算站点相关联的当前工作负荷等级、与当前服务于所述工作负荷的所述相应计算站点相关联的本地资源数量、或用于指示代表所述工作负荷而从第二计算站点借取的资源的阵列,其中所述等级指的是处理工作负荷所需要的资源数量。

7. 根据权利要求 2 所述的方法,其中所述站点状态模型表示的是与相应计算站点相关联的一个或多个时变参数。

8. 根据权利要求 7 所述的方法,其中所述一个或多个时变参数包括下列参数中的至少一个:与所述相应的计算站点相关联的工作负荷状态,由所述相应计算站点借出到第二计算站点的资源数量,或是与不处于维护模式的所述相应计算站点相关联的资源数量。

9. 根据权利要求 2 所述的方法,其中所述策略模型表示的是所述分布式计算系统对系统事件作出响应的方式。

10. 根据权利要求 9 所述的方法,其中所述策略模型描述的是所述分布式计算系统响应于特定外部事件而可以转换成的一个或多个可能的状态,其中所述外部事件在所述分布式计算系统处于给定的状态的时候发生。

11. 根据权利要求 2 所述的方法,其中所述事件模型表示的是在所述分布式计算系统中触发状态变化的一个或多个事件。

12. 根据权利要求 11 所述的方法,其中所述状态变化导致所述分布式计算系统在与一个或多个所述计算站点相关联的两个或多个工作负荷之间重新分配资源。

13. 根据权利要求 2 所述的方法,其中所述成本模型表示的是在扩展时段中用以操作

所述分布式计算系统的一个或多个成本。

14. 根据权利要求 13 所述的方法,其中所述一个或多个成本包括下列成本中的至少一个 :与服务等级协定的违规相关联的成本,其中所述协定与由所述分布式计算系统所处理的给定工作负荷相关联,与从远端计算站点借取资源来处理给定工作负荷相关联的成本,或是与为远端计算站点上的工作负荷的初始建立和供应相关联的成本。

15. 根据权利要求 2 所述的方法,其中将所述分布式计算系统表示成状态转换模型包括 :

在计算站点上将相关联的资源的可用性建模成状态 ;

在所述计算站点上将相关联的工作负荷发生的变化以及所述相关联的资源的可用性发生的变化建模成状态转换。

16. 根据权利要求 15 所述的方法,其中在所述状态转换模型上叠加一个排队网络模型包括 :

将用于表示所述工作负荷变化的状态转换与取决于所述相关联的工作负荷的特性的第一概率相关联 ;

将用于表示所述资源可用性变化的状态转换与取决于所述相关联的资源故障和恢复的特性的第二概率相关联 ;

将所述第一概率以及所述第二概率与相应的策略相关联 ;

将所述相应策略中的每一个与第三概率相关联,其中所述第三概率取决于根据相应策略所采取的至少一个操作 ;以及

根据排队网络分析技术来推导所述分布式计算系统的至少一个状态的至少一个稳态概率。

17. 根据权利要求 16 所述的方法,其中根据所述排队网络模型的解来确定所述一个或多个策略施加于所述分布式计算系统的性能的作用包括 :

将至少一个成本模型应用于所述至少一个稳态概率 ;以及

作为所述至少一个稳态概率的至少一个函数,计算与所述分布式计算系统性能相关联的至少一个度量。

18. 根据权利要求 17 所述的方法,其中所述至少一个度量包括以下各项中的至少一项 :与所述分布式计算系统相关联的响应时间,与所述分布式计算系统相关联的吞吐量,或是与所述分布式计算系统相关联的资源可用性。

19. 一种用于对适合多个计算站点的一个或多个策略进行分析的设备,所述计算站点在分布式计算系统中对相应的工作负荷进行处理,所述设备包括 :

用于将所述分布式计算系统表示成状态转换模型的装置,其中将规定在计算站点之间共享资源的方式的策略建模成对状态以及状态转换的约束条件 ;

用于在所述状态转换模型上叠加一个排队网络模型的装置 ;以及

用于根据所述排队网络模型的解来确定所述一个或多个策略施加于所述分布式计算系统性能的作用的装置。

用于在分布式计算系统中的性能和策略分析的方法和装置

技术领域

[0001] 本发明主要涉及计算系统，尤其涉及用于其中多个计算站点共享资源的分布式计算机系统的策略分析。

背景技术

[0002] 图 1 是描述通常的分布式计算网络或系统 100 的示意图。系统 100 包括多个以通信方式相连的计算站点 102₁ ~ 102_n (在下文中将其统称为“站点 102”)，其中每一个计算站点都充当了一个或多个应用的宿主。每一个站点 102 都可以访问相应的多个本地资源 (例如服务器、处理器、存储器等等) 104₁ ~ 104_n (在下文中将其统称为“资源 104”)。此外，每一个站点 102 都会接收相应的工作负荷 106₁ ~ 106_n (在下文中将其统称为“工作负荷 106”)，其中所述工作负荷包含了对运行在站点 102 上的应用的请求。

[0003] 站点 102 使用其相应的本地资源 104 来满足其相应的工作负荷 106。此外，在诸如系统 100 之类的分布式计算系统中，以通信方式相连的站点 102 可以与其它站点 102 共享其相应的资源 104，由此，站点 102 可以从某个远端站点 102 借用资源 104，以便有效地处理其工作负荷 106，或者站点 102 也可以将其资源 104 借给某个远端站点 102，以便辅助该远端站点 102 进行工作负荷处理。每一个站点 102 都具有自己的策略集合，该策略集合支配着站点 102 如何以及何时出借 / 借用资源 104 可以进行管理。

[0004] 这些单独的策略极大地影响了整个系统 100 有效处理工作负荷 106 的能力。然而，由于这种策略是随着站点的不同而改变的，因此，这些策略对整个系统 100 及其处理工作负荷 106 的能力所产生的作用是很难量化的。

[0005] 由此，在本领域中需要一种用于在分布式计算系统中的性能和策略分析的方法和装置。

发明内容

[0006] 本发明的用于分布式计算系统中的性能和策略分析的方法和设备的一个实施例包括将分布式计算系统性能描绘成一个状态转换模型。然后，在所述状态转换模型上叠加一个排队网络，并且根据所述排队网络的解来识别一个或多个策略对于所述分布式计算系统产生的作用。

附图说明

[0007] 由此，通过参考附图中描述的实施例，可以得到关于上述简单综述的本发明的更具体的描述，并且可以得到用于详细理解和实现上述发明实施例的方式。然而应该注意的是，附图描述的仅仅是本发明的通常的实施例，因此不应将其视为是对的范围进行限制，因为本发明还可能包括其它效果等价的实施例。

[0008] 图 1 是描述通常的分布式计算网络或系统的示意图；

[0009] 图 2 是描述根据本发明的用于分析与分布式计算系统相关联的资源共享策略集

合的分析工具的一个实施例的示意图；

[0010] 图 3 是描述根据本发明的用于对图 2 所示的不同输入进行处理从而产生总成本的方法的一个实施例的流程图；

[0011] 图 4 是描述根据本发明的用于建造状态转换模型的方法的一个实施例的流程图；以及

[0012] 图 5 是使用通用计算设备所实现的策略分析方法的高级框图。

[0013] 为了便于理解，在这里尽可能使用了相同的参考数字来表示附图中共有的相同部件。

具体实施方式

[0014] 在一个实施例中，本发明是一种用于分布式计算系统中的性能和策略分析的方法和设备。本发明的实施例可以有效地分析适用于单个计算站点的各种资源共享策略对于包含计算站点的整个分布式计算系统的性能所产生的作用。

[0015] 图 2 是描述根据本发明而对关联于分布式计算系统的资源共享策略集合进行分析的分析工具 200 的一个实施例的示意图。如所示，分析工具 200 被适配成接收与分布式计算系统相关的多个输入，并且对这些输入进行处理，以便提供可供用户确定是否可以接受与分布式计算系统相关联的现有资源共享策略的信息。

[0016] 在所描述的实施例中，分析工具 200 接收到的输入包括与分布式计算系统 / 站点相关的多个模型 202。在一个实施例中，对分布式计算系统中的每一个计算站点而言，这些模型至少包括如下模型：站点模型、工作负荷模型、工作负荷状态模型、站点状态模型、事件模型或成本模型。此外，对包含了计算站点的整个分布式计算系统来说，应用于该分布式计算系统的策略模型也被提供了分析工具 202。分析工具 200 则对这些输入进行处理，以便产生与分布式计算系统性能相关联的一个或多个量度 206。在一个实施例中，这些量度包括下文中将更详细描述的用于实施特定策略的成本，此外还包括一个或多个系统性能量度（例如反映了分布式计算系统的响应时间、吞吐量、资源可用性等等的量度）。

[0017] 站点模型描述的是站点的静态（例如不依赖于工作负荷状态而变化）参数，在一个实施例中， $\langle ns, pt \rangle$ 数组阵列描绘一个站点，其中 ns 指的是池类型 pt 资源的数量。本领域技术人员将会了解，虽然站点模型所描述的参数相对于工作负荷状态变化而言可以被视为是静态的，但在其所相关联的站点获得或失去资源的时候，这些参数也可以有所改变。在一个实施例中，每一个池都包含了多个资源。属于共同的池的所有资源基本上都是同类的，这是因为这其中的任何一个资源都可以运行给定的应用。举例来说，在一个实施例中，资源是基于一个或多个标准而被分组到池中的，这些标准包括下列标准中的至少一个：服务器硬件、操作系统以及软件栈。此外，站点模型还包括一个策略集合 p，该集合描述的是与站点相关联的特定策略（例如资源共享策略）。

[0018] 工作负荷模型不但描绘了由某个站点充当宿主的应用，而且还描绘了服务等级协定 (SLA)，该协定明确规定了该站点为其所服务的客户机（例如提供工作负荷的用户）提供的保证。每一个工作负荷都是从至少一个池类型 pt 资源提取其资源的。在任何给定的时间，每一个工作负荷都是处于 n 个等级中的某一个等级的。其中的每一个等级转而将会映射到特定的资源（例如特定数量的服务器），而这些资源则是根据 SLA 来对相关联的工作负

荷满意地进行处理所必需的。

[0019] 此外,工作负荷模型还描绘了一个 $n \times n$ 的转换概率矩阵 tpm , 该矩阵规定了相关联的工作负荷如何在等级之间进行转换, 其中 $tpm(i, j)$ 规定的是工作负荷从等级 i 转换到等级 j 的概率 ($1 \leq i, j \leq n$)。假设某个给定工作负荷保持在等级 i 上的时间量分布是未知的。在一个实施例中,这种分布是指数分布或帕累托 (Pareto) 分布。

[0020] 在一个实施例中,工作负荷状态模型被描绘 $\langle 1v, n1, nb \rangle$, 其中 $1v$ 指的是当前的工作负荷等级 (例如上文所述的工作负荷模型所描述的), $n1$ 指的是当前为该工作负荷提供服务的本地资源 (例如服务器数量), nb 则是一个阵列 (每一个远端站点为其中一个元素), 它指的是代表该工作负荷而从其它站点借取的资源 (例如服务器数量)。在一个实施例中,无论是本地资源还是远端资源,当前为工作负荷提供服务的所有资源都是属于该工作负荷所需要的一个共有的池类型 pt 。

[0021] 站点状态模型描绘的是相关联的站点的当前状态, 其中站点“状态”定义了该站点的本地资源可用性。因此,与描述关联于该站点的静态参数的站点模型相反,相关联的站点状态模型描绘的是与站点相关联的时变参数。在一个实施例中,站点状态被描绘为 $\langle ws, nd, as \rangle$, 其中 ws 是本地工作负荷状态阵列, nd 是表示该站点为处于另一个站点的远端工作负荷所贡献的资源 (例如服务器数量) 的阵列 (每一个远端工作负荷为其中一个元素), as 则是表示处于有效模式的本地资源的有效资源 (例如服务器) 的阵列 (每一个池类型 pt 为其中一个元素), 其中处于有效模式的资源是那些当前可以用于为工作负荷提供服务的资源。

[0022] 策略模型描绘的是分布式计算系统对系统事件做出响应的方式。在一个实施例中,策略模型被描绘为 $P(S, e)$, 其中 P 是应用于系统 (例如借助系统中包含的单个站点) 的策略集或策略集合, S 表示的是分布式计算系统的当前状态 (例如资源可用性), e 则是调用了集合 P 中的一个或多个策略的外部事件 (例如下文中更详细描述的工作负荷事件或服务器事件)。当系统处于状态 S 时,如果出现这种外部事件 e ,则通过应用策略集 P 而将系统指引到可能存在的新状态的集合,其中所有这些状态都是符合该策略集的。例如,在发生外部事件 e 的时候,如果当前系统状态是 S ,那么可以应用策略集 P ,从而形成 $\{(S1, p1), (S2, p2), \dots, (Sn, pn)\}$, 其中 $Si (1 \leq i \leq n)$ 是根据策略集 P 的有效的下一个状态, pi 则是系统被建议为转换到状态 Si 的概率,并且 $\sum_i pi = 1$ 。

[0023] 应该指出的是,如果 $n = 1$,那么与策略集 P 相符合的下一个状态只有一个。可选择的是,如果 $n > 1$,那么系统将会从集合 $\{Si : 1 \leq i \leq n\}$ 中以概率统计的方式选择下一个状态。策略模型是一般性的,它可以容纳大范围的用于分布式计算的策略。

[0024] 事件模型描述的是触发状态变化的事件或是外部变化,由此需要系统在系统工作负荷之间重新分配或是重新分发系统资源,以便满足系统业务目标。在一个实施例中,事件模型描述的是以下两种普通类型事件中的至少一种:(1) 工作负荷事件,它是在一个或多个站点的工作负荷需要从第一等级移动到第二等级时候发生;以及(2) 资源事件,它是在在资源 (例如服务器) 出现故障、转换到待用模式时发生,或是从故障中恢复以及维护结束后重新恢复到系统之中的时候发生。通常,对给定的状态 S 而言存在一个事件集合 E ,当系统处于状态 S 时,有可能发生这些事件。对每一个工作负荷事件 $e \in E$ 来说,其变换到下一个事件 e 的时间的概率分布是从上述工作负荷模型中获取的。对每一个资源事件来说,假

设资源是独立发生故障（以及恢复）的，并且发生故障（和恢复）的时间是指数分布的。

[0025] 成本模型描述的是在扩展时段中在分布式计算系统内部操作多个站点的成本。由此，成本模型被用于评估各种站点专用策略对于整个分布式计算系统的作用。在一个实施例中，成本模型描述了至少三种主要成本：(1) 违规成本 VC ，它表示的是违反给定工作负荷 SLA 的成本；(2) 远端资源成本 RRC ，它表示的是使用远端资源处理给定工作负荷的成本；以及 (3) 重新分配成本 RC ，它表示的是用于给定工作负荷的初始建立和供应成本。

[0026] 在另外的实施例中，该成本模型还描述了至少三个较低等级成本函数。第一个较低等级成本函数 $\alpha(S, w)$ 描述的是在系统处于状态 S 时被用于工作负荷 w 的违规成本。这个违规成本与 $num_deficit_servers(S, w)$ 是成比例的，所述 $num_deficit_servers(S, w)$ 描述的则是处理工作负荷 w 所需要的资源（例如服务器数量）与实际分配给工作负荷 w 的资源（本地和远端）之间的差别。由此，在工作负荷 SLA 中，违规成本 $\alpha(S, w)$ 被表述成是单位时间内的每个亏损资源的处罚。

[0027] 第二个较低等级成本函数 $\gamma(S, w)$ 描述的是在系统处于状态 S 时使用远端资源来处理工作负荷 w 的远端资源成本。这个远端资源成本与 $num_borrowed_servers(S, w)$ 成比例，所述 $num_borrowed_servers(S, w)$ 描述的则是在系统处于状态 S 时代表工作负荷 w 所借取的远端资源（例如服务器数量）。在这种情况下，在站点策略集合中，远端资源成本 $\gamma(S, w)$ 被表述成是单位时间内每一个远端资源的处罚。在一个实施例中，远端资源成本 $\gamma(S, w)$ 对从不同站点借取的资源进行了区分。在这种情况下，用于某一个被借取的资源的远端资源成本 $\gamma(S, w)$ 至少部分取决于用以出借资源的站点。

[0028] 第三个较低等级成本函数 $\beta(S, w)$ 描述的是在系统从状态 S 转换到状态 S' 时用于工作负荷 w 的重新分配成本。重新分配成本 $\beta(S, w)$ 至少部分取决于重新分配的资源在被重新分配时处于空闲还是在运行某些工作负荷，并且至少部分取决于重新分配的资源相对于工作负荷而言处于本地还是远端，此外还至少部分取决于提供和建立工作负荷的成本。

[0029] 图 3 是描述根据本发明来处理图 2 所示的各种输入 202，以便产生总体成本 206 的方法 300 的一个实施例的流程图。其中举例来说，方法 300 是可以在分析工具 200 中实现的。

[0030] 方法 300 在步骤 302 开始并且进行到步骤 304，在步骤 304 中，方法 300 使用状态转换模型来描绘分布式计算系统的分析。特别地，方法 300 将分布式计算系统在给定时间的状况描绘成是状态转换模型中的状态。在一个实施例中，状态转换模型是从与包括在分布式计算系统中的站点相关联的各种模型（如上所述）中建造的。在一个实施例，这包括将每一个站点上的资源可用性建模成一个状态，将站点上的资源可用性的变化（例如由于资源故障、恢复、借入或借出）模拟成状态转换模型，以及将站点工作负荷变化建模成状态转换模型。此外，应用于站点的任何策略都建模为站点状态和站点转换的约束条件。在以下更详细描述的图 4 中说明了根据与站点相关联的模型来建造状态转换模型的方法的一个实施例。在一个实施例中，状态转换模型的建造包括将单个站点策略编码到上述函数 $P(S, e)$ 中。

[0031] 一旦建造了状态转换模型，那么方法 300 进行到步骤 306，并且会在状态转换模型上叠加一个排队网络。在一个实施例中，在状态转换模型上叠加排队网络的处理包括识别

用于分布式计算网络的有效状态以及状态转换。然后,这些有效状态转换将会使用其发生的概率(举例来说,该概率取决于相关联的工作负荷的特性以及资源故障/恢复特性)以及与策略相关联的概率(举例来说,该概率可以是与策略相关联的操作的概率,其中该操作可以包括将分布式计算网络转换到一个或多个不同状态)来加以注释。

[0032] 在一个实施例中,方法 300 将一个排队网络叠加在状态转换模型之上,以便使用与之相关联的概率分布函数来注释状态转换 T。根据步骤 306,转换 $T:S \rightarrow eS'$ 是用数组 $\langle fe, pr \rangle$ 来标记的,其中函数 fe 描述的是导致系统从状态 S 转换到状态 S' 的事件 e 的概率分布。在一个实施例中,函数 fe 是基于希望的精确度等级来选择的恰当的统计分布(例如指数分布或帕累托分布)。概率 pr 描述的是系统响应于事件 e 而从状态 S 转换到状态 S' 的概率。

[0033] 在步骤 308 中,方法 300 通过求解排队网络(例如通过求解马尔可夫(Markov)链)来推导出至少一个成本度量。根据本发明,排队网络模型的解为分布式计算系统的不同状态给出了稳态概率。在一个实施例中,排队网络的解为所有状态 S 给出了 $pr(S)$,并且为所有转换 T 给出了 $rate(T)$,其中 $pr(S)$ 描述的是系统处于状态 S 的概率(在某个扩展时段中), $rate(T)$ 描述的则是系统执行转换 T 的速率(在某个扩展时段中)。在一个实施例中,与工作负荷相关联的成本是根据下式来推导的(用单位时间的平均成本单位来表示):

$$[0034] VC(w) = \sum S \sum w \alpha(S, w) * pr(S) \quad (\text{等式 1})$$

$$[0035] RRC(w) = \sum S \sum w \gamma(S, w) * pr(S) \quad (\text{等式 2})$$

$$[0036] RC(w) = \sum T:S \rightarrow S' \sum w \beta(S, S', w) * rate(T) \quad (\text{等式 3})$$

[0037] 项 $pr(S)$ 和 $rate(T)$ 可以采用如下方式而从排队网络模型中确定。在一个实施例中,其中工作负荷事件 e 是指数分布的,那么,稳定的概率分布是通过使用标准的分析技术来求解马尔可夫链而被计算得到的。这个稳定的概率分布为所有的状态 S 提供了 $pr(S)$ 。对每一个转换 $T:S \rightarrow eS'$ 来说,

$$[0038] rate(T) = pr(S) * rate(fe) \quad (\text{等式 4})$$

[0039] 其中 $rate(fe)$ 描述的是指数分布 fe 的速率。

[0040] 在一个替代的实施例中,工作负荷事件 e 遵循的是一个重尾帕累托分布,其中可以通过应用一个离散事件模拟来求解排队网络模型。在一个实施例中,该模拟是在一个很长的时段 t_{sim} 中执行的。并且在一个实施例中, t_{sim} 表示的是在分布式计算系统上运行的应用达到稳态所需要的时间量(例如对某些应用而言,该时间量可以是大约 8,000 秒)。在模拟过程中,对处于任何状态 S 中的系统所消耗的时间量通过 $t(S)$ 进行了测量,并且使用了这个消耗时间来计算 $pr(S)$,由此

$$[0041] pr(S) = t(S) / T_{sim} \quad (\text{等式 5})$$

[0042] 同样,系统从状态 S 转换到状态 S' 所耗费的时间量 $n(T)$ 是使用转换 T 通过 $n(T)$ 来测量的,由此可以将 $rate(T)$ 估计成:

$$[0043] rate(T) = n(T) / t_{sim} \quad (\text{等式 6})$$

[0044] 在为所有的状态 S 给出了 $pr(S)$ 并且为所有的转换 T 给出了 $rate(T)$ 的情况下,工作负荷成本可以参考等式 1、2、3 并以上文所述的方式来估计。

[0045] 对分布式计算性能而言,其附加度量(例如响应时间、吞吐量、资源可用性等等)也可以作为排队网络解的函数来计算。由此,用户可以对方法 300 产生的度量进行检查,以

便确定用于分布式计算系统的当前策略集 P 是否能使分布式计算系统以一种令人满意的方式来处理工作负载。然后，方法 300 在步骤 310 结束。

[0046] 由此，方法 300 有助于对适合单个站点的各种资源共享策略对于包含了站点的整个分布式计算系统所产生的作用进行有效的分析。通过使用状态转换模型，以及随后通过应用给定策略来识别有效的分布式计算状态和状态转换，可以快速确定给定所述策略对分布式计算系统的成本和性能所产生的影响，并且采用一种易于分析的形式来显示所述影响。

[0047] 图 4 是描述根据本发明并用于基于上述与站点相关联的模型来建造状态转换模型的方法 400 的一个实施例的流程图。其中举例来说，方法 400 是可以根据方法 300 中的步骤 304 来加以实施的。

[0048] 方法 400 始于步骤 402 并且进行到步骤 404，在步骤 404 中，方法 400 将会根据所有事件的集合 E 以及应用于分布式计算系统中的站点的所有站点策略的共同策略集合 P 而对用于分布式计算系统的工作负载状态进行初始化处理。

[0049] 在步骤 406，方法 400 初始化系统起始状态 S_0 。然后，方法 400 会将系统起始状态 S_0 添加到状态池中。该状态池包含了一个或多个由方法 400 推导得到的分布式计算系统的状态。

[0050] 在步骤 410，方法 400 确定状态池中是否存在未经调查的系统状态。在一个实施例中，对一个系统状态而言，如果不存在下一个潜在状态并且已经为该系统状态确定了相应的转换速率，那么这个系统状态是未经调查的。如果方法 400 确定状态池中并未保留未经调查的系统状态，那么方法 400 在步骤 421 终止。方法 400 的终止表明已经建立了状态转换模型，并且可以结合上文中参考图 3 所描述的方法 300 来实施该状态转换模型。

[0051] 作为选择，如果方法 400 确定在状态池仍旧存在一个或多个未经调查的系统状态，那么方法 400 前进到步骤 414，并且从状态池中检索未经调查的状态 S_n 。然后，方法 400 前进到步骤 416b，如果存在与检索到的状态 S_n 相关联的下一个状态，那么方法 400 产生下一个状态以及与检索到的状态 S_n 相关联的相应的转换速率。在一个实施例中，生成下一个状态以及相应的转换速率的处理是根据状态 S_n 以及策略集 P 中的站点专用策略来执行的。

[0052] 在步骤 418 中，方法 400 将会根据步骤 416 中产生的下一个状态以及相应的转换速率来为分布式计算系统更新状态转换矩阵。然后，在步骤 420 中，所述这些下一个状态被添加到状态池中。

[0053] 在步骤 422 中，方法 400 将检索到的状态 S_n 表示成是未经调查的。然后，方法 400 返回步骤 410 并以上述方式继续进行，以便调查状态池中任何剩余的未经调查的状态。步骤 410 ~ 422 可被重复执行所需要的次数，直至调查了状态池中的所有状态以及无法将新状态添加到状态池中为止。本领域技术人员将会了解，一旦没有将新的状态添加到状态池中，则不重复执行步骤 412 并且该步骤只会出现一次。

[0054] 图 5 是使用通用计算设备 500 所实现的策略分析方法的高级框图。在一个实施例中，通用计算设备 500 包含了处理器 502、内存 504、策略分析模块 505 以及例如显示器的各种输入 / 输出 (I/O) 设备 506、键盘、鼠标、调制解调器等等。在一个实施例中，至少有一个 I/O 设备是存储设备（例如磁盘驱动器、光盘驱动器、软盘驱动器）。应该理解的是，策略分析模块 505 可以实现为物理设备，也可以实现为经由通信信道而与处理器相连的子系统。

[0055] 作为选择,策略分析模块 505 也可以用一个或多个软件应用(甚至是软件和硬件的组合来表示,其中举例来说,该组合可以用专用集成电路(ASIC)来实现)来表示,其中软件是从存储介质(例如 I/O 设备 506)加载并在通用计算设备 500 的内存 504 中由处理器 502 来操作的。由此,在一个实施例中,对在这里参考先前附图所描述的用于分析与分布式计算系统相关联的性能和策略的策略分析模块 505 可以存储在计算机可读介质或载体上(例如 RAM、磁或光驱动器或磁盘等等)。

[0056] 由此,本发明描述了分布式计算系统分析领域中的一个显著进步。在这里提供的方法和设备可以有效地分析适合单个站点的各种资源共享策略施加对包含站点的整个分布式计算系统产生的作用。通过将状态转换建模为状态转换模型,以及随后应用给定策略来识别有效的分布式计算状态和状态转换,可以快速识别给定所述策略对分布式计算系统的成本和性能所产生的影响,并且采用排队网络分析来显示所述影响。

[0057] 虽然上文涉及的是本发明的优选实施例,但在不脱离本发明的基本范围的情况下,可以设想本发明的其它实施例,本发明的范围是由下列权利要求确定的。

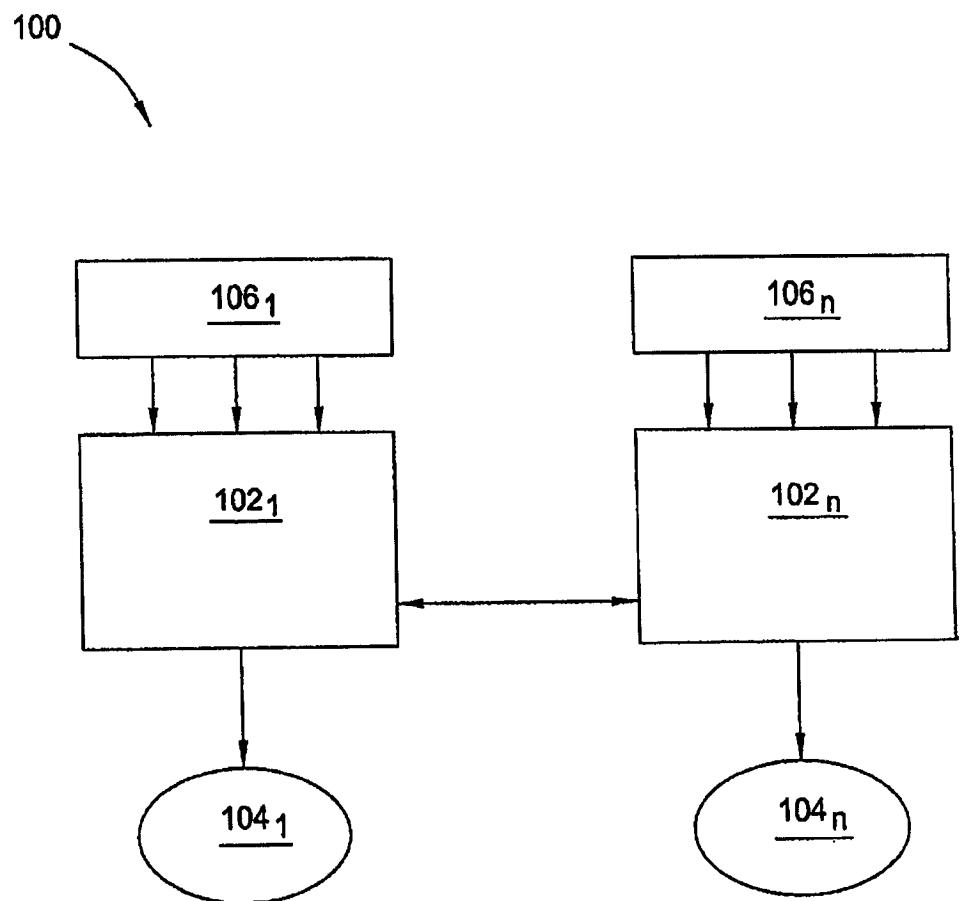


图 1

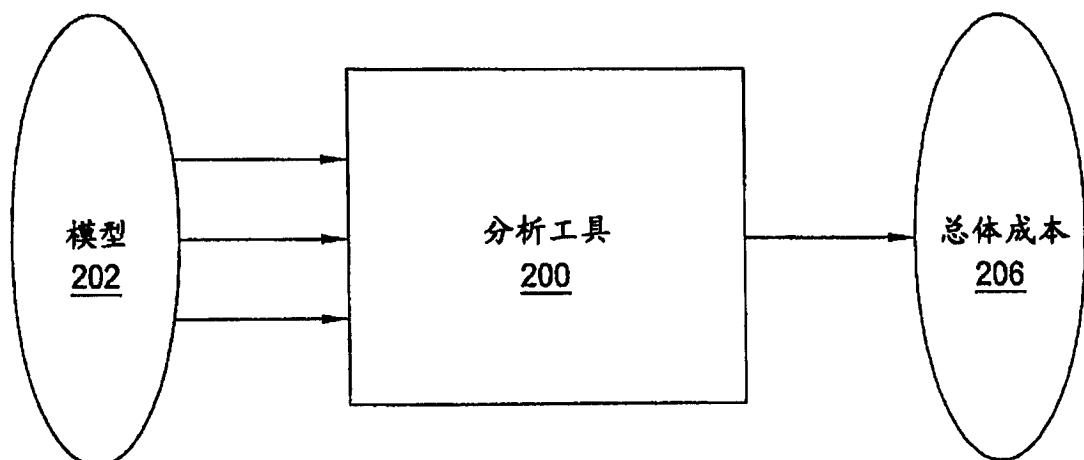
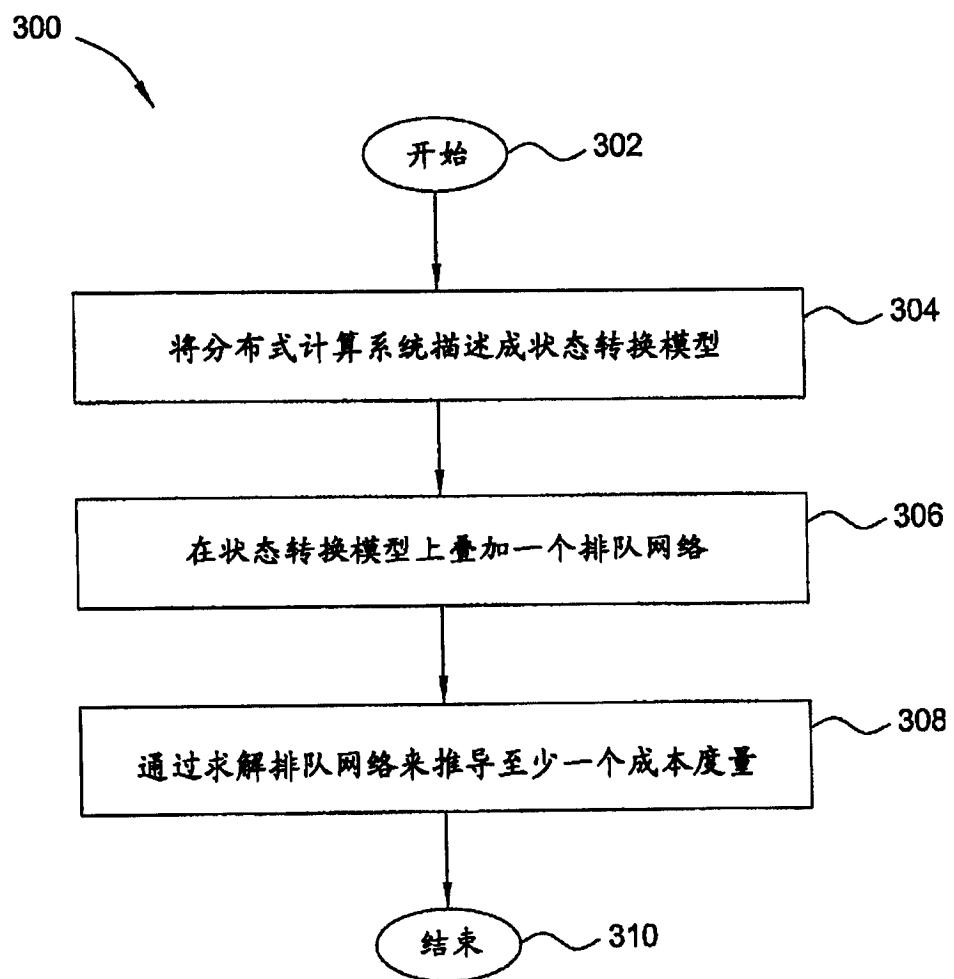


图 2



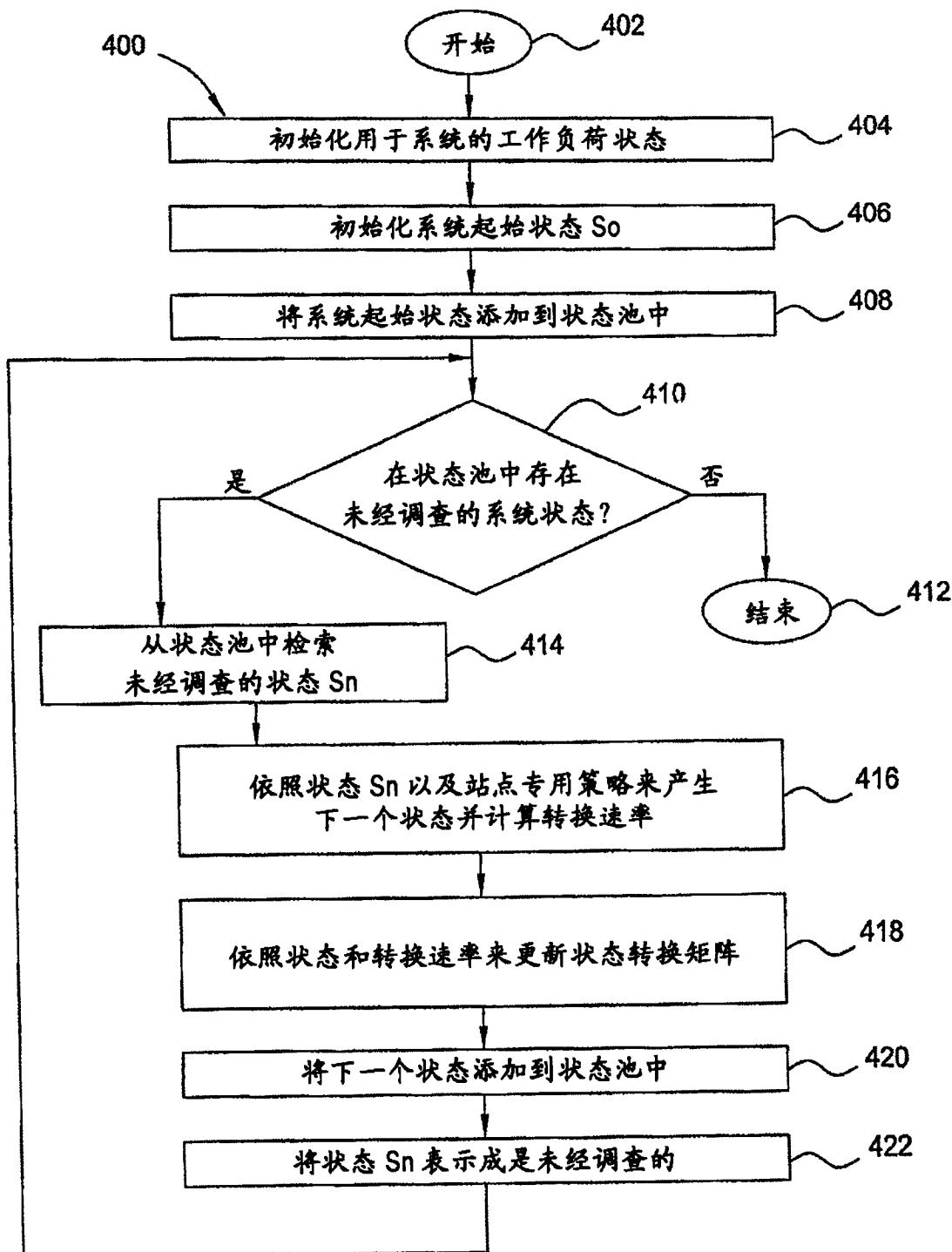


图 4

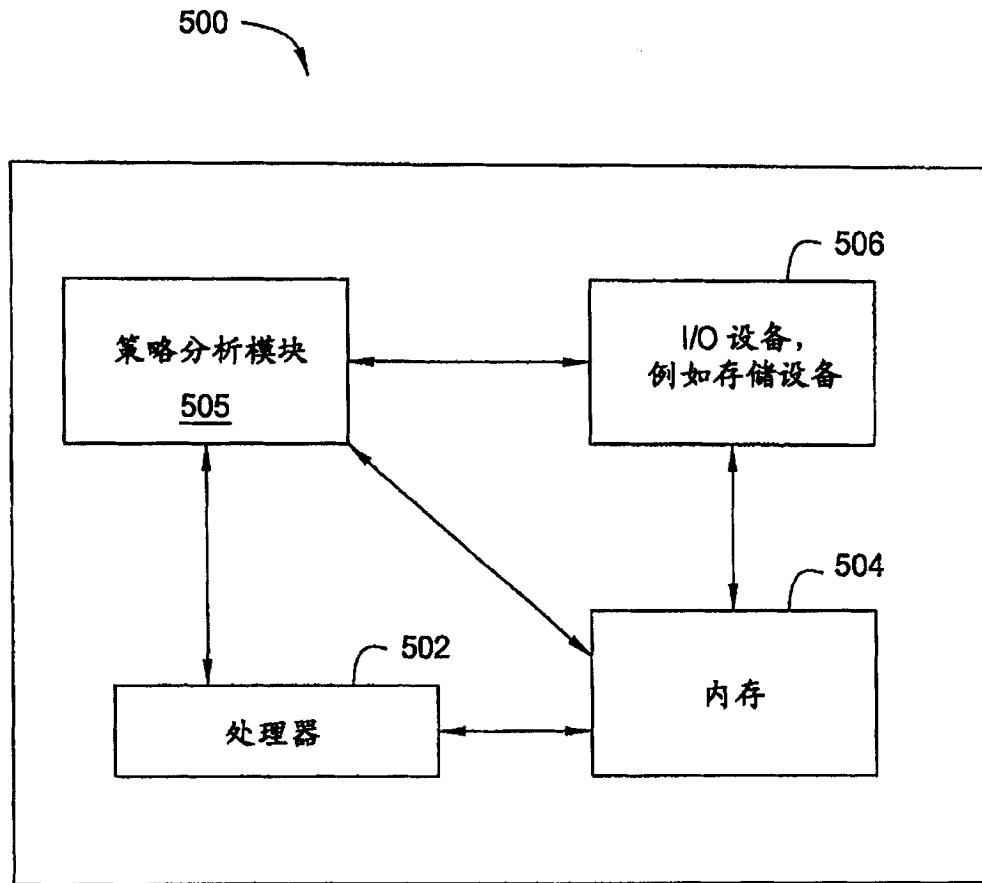


图 5