



(12) 发明专利申请

(10) 申请公布号 CN 101853197 A

(43) 申请公布日 2010. 10. 06

(21) 申请号 201010158799. 9

(22) 申请日 2010. 03. 30

(30) 优先权数据

12/415369 2009. 03. 31 US

(71) 申请人 国际商业机器公司

地址 美国纽约

(72) 发明人 G·巴克内尔

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 于静 杨晓光

(51) Int. Cl.

G06F 11/22(2006. 01)

H04L 12/26(2006. 01)

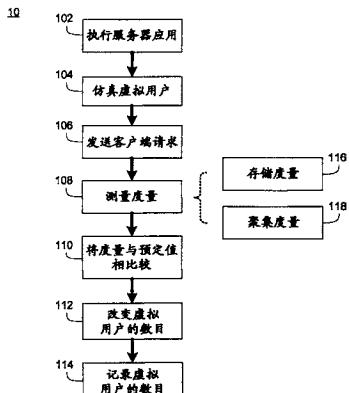
权利要求书 2 页 说明书 7 页 附图 2 页

(54) 发明名称

确定服务器负载能力

(57) 摘要

一种用于确定与服务器负载相应的用户数目的方法、计算机程序产品和系统，包括：在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户。从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用。测量表示所述服务器计算机系统上的负载的度量。将所述度量与预定值相比较；以及基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目，直到所述度量在所述预定值的预定范围内。



1. 一种用于确定与服务器负载相应的用户数目的方法,包括:
在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户;
从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用;
测量表示所述服务器计算机系统上的负载的度量;
将所述度量与预定值相比较;以及
基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目,直到所述度量在所述预定值的预定范围内。
2. 如权利要求1所述的方法,其中所述度量表示所述服务器计算机系统的一个或多个处理器上的负载。
3. 如权利要求1所述的方法,其中测量度量包括:周期性测量所述度量,并将每个测量存储在存储器中。
4. 如权利要求3所述的方法,其中测量度量包括以下步骤中的一个或多个:
计算所述度量的运行平均数;以及
对于所述度量应用低通滤波器。
5. 如权利要求1所述的方法,其中测量度量包括:测量由以下项目中的一个或多个提供的度量:
在所述服务器计算机系统上执行的Windows管理工具服务;
在所述服务器计算机上执行的rstat.d性能测量服务;以及
在所述服务器计算机上执行的自定义测量工具。
6. 如权利要求1所述的方法,其中改变所述一个或多个虚拟用户的数目包括:至少部分地基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。
7. 一种用于确定与服务器负载相应的用户数目的计算机程序产品,其位于计算机可读介质上,在上面存储有多个指令,在由处理器执行时其使得所述处理器执行以下操作,包括:
在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户;
从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用;
测量表示所述服务器计算机系统上的负载的度量;
将所述度量与预定值相比较;以及
基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目,直到所述度量在所述预定值的预定范围内。
8. 如权利要求7所述的计算机程序产品,其中所述度量表示所述服务器计算机系统的一个或多个处理器上的负载。
9. 如权利要求7所述的计算机程序产品,其中测量度量包括:周期性测量所述度量,并将每个测量存储在存储器中。
10. 如权利要求9所述的计算机程序产品,其中测量度量包括以下步骤中的一个或多个:
计算所述度量的运行平均数;以及

对于所述度量应用低通滤波器。

11. 如权利要求 7 所述的计算机程序产品,其中测量度量包括 :测量由以下项目中的一个或多个提供的度量 :

在所述服务器计算机系统上执行的 Windows 管理工具服务 ;

在所述服务器计算机上执行的 rstat. d 性能测量服务 ;以及

在所述服务器计算机上执行的自定义测量工具。

12. 如权利要求 7 所述的计算机程序产品,其中改变所述一个或多个虚拟用户的数目包括 :基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。

13. 一种用于确定与服务器负载相应的用户数目的系统,包括 :

一个或多个处理器 ;

一个或多个存储器 ;

在所述一个或多个存储器上存储的一个或多个软件指令集,在由所述一个或多个处理器执行时其执行以下操作,包括 :

在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户 ;

从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用 ;

测量表示所述服务器计算机系统上的负载的度量 ;

将所适度量与预定值相比较 ;以及

基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目,直到所述度量在所述预定值的预定范围内。

14. 如权利要求 13 所述的系统,其中所述度量表示所述服务器计算机系统的一个或多个处理器上的负载。

15. 如权利要求 13 所述的系统,其中测量度量包括 :周期性测量所述度量,并将每个测量存储在存储器中。

16. 如权利要求 15 所述的系统,其中测量度量包括以下步骤中的一个或多个 :

计算所述度量的运行平均数 ;以及

对于所述度量应用低通滤波器。

17. 如权利要求 13 所述的系统,其中测量度量包括 :测量由以下项目中的一个或多个提供的度量 :

在所述服务器计算机系统上执行的 Windows 管理工具服务 ;

在所述服务器计算机上执行的 rstat. d 性能测量服务 ;以及

在所述服务器计算机上执行的自定义测量工具。

18. 如权利要求 13 所述的系统,其中改变所述一个或多个虚拟用户的数目包括 :基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。

确定服务器负载能力

技术领域

[0001] 本发明涉及计算机测试系统,更具体地,涉及用于确定服务器负载的计算机测试系统。

背景技术

[0002] 通常,服务器计算机同时向许多用户提供服务。有时,同时尝试访问服务器的许多用户可能使服务器过载。这可使得服务器遭受服务延迟,或完全故障。避免这些故障的一个方法是知晓服务器的限制,并设计服务器或服务器的系统以处理期望的负载。

[0003] 测试服务器可包括向服务器发送由虚拟用户生成的测试业务和测试客户端请求。这些方法中的某些通过仿真用户业务和客户端请求、并测量被测试的服务器的响应来创建网络业务。然而,为了有效设计服务器或服务器的系统以处理期望的负载,设计者可需要知晓与特定操作点相关的服务器负载。例如,为了设计健壮的服务器系统,设计者可需要服务器不在特定能力(比如 80% 能力)以上运行。在这个情况下,期望知晓什么类型的负载将使得服务器以 80% 能力运行。

发明内容

[0004] 在一实施例中,一种用于确定与服务器负载相应的用户数目的方法,包括:在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户。从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用。测量表示所述服务器计算机系统上的负载的度量。将所述度量与预定值相比较;以及基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目,直到所述度量在所述预定值的预定范围内。

[0005] 可提供以下特征中的一个或多个。所述度量可表示所述服务器计算机系统的一个或多个处理器上的负载。测量度量可包括:周期性测量所述度量,并将每个测量存储在存储器中。该方法可计算所述度量的运行平均数和/或对于所述度量应用低通滤波器。所述度量可由在所述服务器计算机系统上执行的**Windows®**管理工具服务、和 rstat.d 性能测量服务、和/或自定义性能测量工具来提供(**Windows** 是微软公司在美国、其他国家、或美国和其他国家的商标)。此外,改变所述一个或多个虚拟用户的数目可包括:基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。

[0006] 在另一实施例中,一种用于确定与服务器负载相应的用户数目的计算机程序产品位于计算机可读介质上。计算机程序产品的计算机指令在由处理器执行时使得所述处理器执行以下操作,包括:在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户。从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用。测量表示所述服务器计算机系统上的负载的度量。将所述度量与预定值相比较;以及基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目,直到所述度量在所述预定值的预定范围内。

[0007] 可提供以下特征中的一个或多个。所述度量可表示所述服务器计算机系统的一个或多个处理器上的负载。测量度量可包括：周期性测量所述度量，并将每个测量存储在存储器中。所述操作可计算所述度量的运行平均数和 / 或对于所述度量应用低通滤波器。所述度量可由在所述服务器计算机系统上执行的 Windows 管理工具服务、和 rstat.d 性能测量服务、和 / 或自定义性能测量工具来提供。此外，改变所述一个或多个虚拟用户的数目可包括：基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。

[0008] 在另一实施例中，一种用于确定与服务器负载相应的用户数目的系统，包括：一个或多个处理器；一个或多个存储器；在所述一个或多个存储器上存储的一个或多个软件指令集，在由所述一个或多个处理器执行时其执行以下操作，包括：在与网络耦合的测试计算机系统上仿真一个或多个虚拟用户。从所述一个或多个虚拟用户中的每个将一个或多个客户端请求发送至在与所述网络耦合的服务器计算机系统上执行的服务器应用。测量表示所述服务器计算机系统的负载的度量。将所述度量与预定值相比较；以及基于所述度量和所述预定值之间的差改变所述一个或多个虚拟用户的数目，直到所述度量在所述预定值的预定范围内。

[0009] 可提供以下特征中的一个或多个。所述度量可以是所述服务器计算机系统的一个或多个处理器上的负载的测量。测量度量可包括：周期性测量所述度量，并将每个测量存储在存储器中。所述操作可计算所述度量的运行平均数和 / 或对于所述度量应用低通滤波器。所述度量可由在所述服务器计算机系统上执行的 Windows 管理工具服务、和 rstat.d 性能测量服务、和 / 或自定义性能测量工具来提供。此外，改变所述一个或多个虚拟用户的数目可包括：基于在所述度量和所述预定值之间的百分数差改变所述一个或多个虚拟用户的数目。

附图说明

[0010] 图 1 是与分布式计算网络耦合的服务器测试处理的示意图。

[0011] 图 2 是图 1 的测试系统配置处理的流程图。

[0012] 在附图中的类似标号指示类似的元素。

具体实施方式

[0013] 图 1 示出服务器测试处理 10，其可位于测试计算机 12 上并且可由其执行，测试计算机 12 可连接至网络 14（例如因特网或局域网）。测试计算机 12 的实例可包括但不限于：个人计算机、服务器计算机、系列服务器或个人计算机、迷你计算机、和主机计算机。

[0014] 图 1 还示出服务器应用 16，其可位于服务器计算机 18 上并且可由其执行。服务器计算机 18 也可连接至网络 14，如图所示。服务器计算机 18 的实例也可包括但不限于：个人计算机、服务器计算机、系列服务器或个人计算机、迷你计算机、和主机计算机。

[0015] 服务器应用 16 可以是接受来自客户端应用的请求、输入、或其他通信的任意服务器应用。例如，服务器应用 16 可以是 web 服务器、应用服务器、ftp 服务器、或接受来自用户（例如用户 20）的请求或输入的任意服务器。服务器应用的实例包括 Microsoft® IIS 服务器和 IBM WebSphere® 服务器（Microsoft 是微软公司在美国、其他国家、或美国和其他国家的商标；IBM 和 WebSphere 是国际商业机器公司在美国、其他国家、或美国和其他

国家的商标或注册商标)。

[0016] 服务器计算机 18 可运行网络操作系统,其实例可包括但不限于:例如 Microsoft Windows XP 服务 器;Novell **Netware®**; 或 Redhat **Linux®** (Novell 和 NetWare 是 Novell 公司在美国、其他国家、或美国和其他国家的注册商标;Redhat 是 Red Hat 公司在美国、其他国家、或美国和其他国家的注册商标;Linux 是 Linus Torvalds 在美国、其他国家、或美国和其他国家的注册商标)。

[0017] 如以下将更详细讨论地,服务器测试处理 10 可仿真一个或多个虚拟用户,并且可从一个或多个虚拟用户将一个或多个客户端请求发送至服务器应用 16。服务器测试处理 10 还测量用以代表在服务器计算机系统 18 上的负载的度量,并将所述度量与预定值相比较。服务器测试处理 10 可改变被仿真的虚拟用户的数目,直到该度量在预定值的预定范围内。被仿真的虚拟用户的数目的改变可基于在度量和预定值之间的差。

[0018] 可由集成到测试计算机 12 中的一个或多个处理器(未示出)和一个或多个存储器架构(未示出)来执行服务器测试处理 10 的指令集和例程,其可包括一个或多个软件模块和一个或多个软件指令集,并且可存储在耦合至测试计算机 12 的存储设备 22 上。类似地,可由集成到服务器计算机 18 中的一个或多个处理器(未示出)和一个或多个存储器架构(未示出)来执行服务器应用 16 的指令集和例程,其可包括一个或多个软件模块和一个或多个软件指令集,并且可存储在耦合至服务器计算机 18 的存储设备 24 上。存储器设备 22、24 可包括但不限于:硬盘驱动器;固态驱动器;磁带驱动器;光盘驱动器;RAID 阵列;随机存取存储器(RAM);和只读存储器(ROM)。

[0019] 服务器测试处理 10 和服务器应用 16 的指令集和例程可通过使用编译器生成,例如 IBM XL 编译器、Visual **Studio®** 编译器、或汇编程序(VisualStudio 是微软公司在美国、其他国家、或美国和其他国家的商标)。指令集和例程还可作为脚本生成,例如 **JavaScript®** 脚本、或其他类型软件代码(Java 和所有基于 Java 的商标和 logo 是 Sun Microsystems 公司在美国、其他国家、或美国和其他国家的商标)。一般地且本领域已知地,服务器测试处理 10 和服务器应用 16 的指令集和例程可经由可由处理器执行的生成指令集和例程的任意方法来生成。

[0020] 各种客户端电子设备可耦合至网络 14。客户端电子设备的实例可包括但不限于:测试计算机 12、服务器计算机 18、膝上型电脑 26、个人数字助理(未示出)、数据使能的蜂窝式电话(未示出)、和专用网络设备(未示出)。一个或多个用户(例如用户 20)可访问各种工具和应用,例如来自于连接至网络 14 的任意设备或直接来自于上面有工具或应用的设备(例如,直接来自于测试计算机 12 和 / 或服务器计算机 18)的服务器测试处理 10 或服务器应用 16。客户端电子设备可直接耦合至网络 14,或可通过网络 28 间接耦合至网络 14。例如,测试计算机 12 可通过二级网络 28(如幻影链路线 30 所示)间接耦合至网络 14。

[0021] 存储设备 22,24 可包括但不限于:硬盘驱动器;固态驱动器;磁带驱动器;光盘驱动器;RAID 阵列;随机存取存储器(RAM);只读存储器(ROM);压缩闪速(CF)存储设备;安全数字(SD)存储设备;和存储棒存储设备。

[0022] 各个客户端电子设备可直接地或间接地耦合至网络 14(或网络 28)。例如,示出测试计算机 12 经由硬连线网络连接直接耦合至网络 14。示出膝上型电脑 26 经由在膝上型电

脑 26 和无线接入点 (即 WAP) 34 之间建立的无线通信信道 32 无线地耦合至网络 14, 示出为其直接地耦合至网络 14。WAP 34 可以是例如 IEEE 802.11a、802.11b、802.11g、Wi-Fi、和 / 或能够在膝上型电脑 26 和 WAP 34 之间建立无线通信信道 32 的蓝牙设备。

[0023] 本领域已知地,所有 IEEE 802.11x 规范可使用以太网协议和具有避免碰撞的载波检测多址接入 (即 CSMA/CA) 用于共享。各个 IEEE 802.11x 规范可使用例如相移键控 (即 PSK) 调制或补码键控 (即 CCK) 调制。本领域已知地,蓝牙是允许例如移动电话、计算机和个人数字助理使用短距离无线连接互联的电信业规范。

[0024] 参照图 2,服务器计算机 18 可执行 102 服务器应用 16。服务器测试处理 10 可仿真 104 一个或多个虚拟用户,并从一个或多个虚拟用户中的每个将一个或多个客户端请求发送至 106 服务器应用 16。服务器测试处理 10 还可测量 108 用以代表在服务器计算机系统 18 上的负载的度量,并将所述度量与预定值相比较 110。服务器测试处理 10 还可改变被仿真的虚拟用户的数目,直到所述度量在预定值的预定范围内。被仿真的虚拟用户的数目的改变可至少部分地基于在度量和预定值之间的差。当被仿真的虚拟用户的数目的改变达到稳定状态时,例如,度量和预定值之间的差较小或为 0,并且虚拟用户的数目不再改变,服务器测试处理 10 可记录 114 用以代表与所述预定值相关的用户数目的虚拟用户数目。

[0025] 服务器计算机 18 可执行 102 服务器应用 16。服务器应用 16 可以是从用户接受输入或为用户提供输出的任意类型服务器应用。这里,从用户接受输入和 / 或为用户提供输出可包括 (例如响应于向客户端应用的用户输入) 从客户端应用接受输入和 / 或 (例如经由客户端应用) 提供可由用户接收的输出。在一个实施例中,服务器应用 16 可以是从用户 (例如经由膝上型电脑 26 执行的 web 浏览器应用 (未示出) 从用户 20) 接收 web 页面请求的 web 服务器,并且响应地可通过利用网络 14 发送来向用户 20 提供 web 页面和 / 或 web 内容。然而,服务器应用 16 可以是任意类型服务器应用,包括但不限于:电子邮件服务器、web 服务器、ftp 服务器、搜索引擎、定制服务器、和任意其他服务器应用。

[0026] 典型地,服务器计算机 18 可执行服务器应用 16。换句话说,服务器计算机 18 的处理器可执行与服务器应用 16 相关的任意指令。执行服务器应用 16 可消耗服务器计算机 18 的处理器周期和 / 或其他处理器资源。同样,服务器应用 16 可使用例如与服务器计算机 18 相关的其他硬件资源。服务器应用 16 例如可选地可利用网络连接、I/O 端口、硬驱动器、存储器、或与服务器计算机 18 相关的任意其他硬件资源。服务器应用 16 还可利用例如在服务器计算机 18 上执行的软件应用的软件资源。例如,服务器计算机 18 可提供软件加密服务。如果服务器应用 16 需要加密数据,例如,服务器应用 16 可使用在服务器计算机 18 上运行的加密服务。

[0027] 服务器应用 16 可利用服务器计算机 18 的硬件和软件资源的程度可取决于在服务器应用 16 上存在有多大负载。例如,如果服务器应用 16 正在忙于执行复杂处理作业,则服务器应用 16 可利用服务器计算机 18 的多个 CPU 周期。同样,如果服务器应用 16 正在存储或发送大量数据,则服务器应用 16 可利用服务器计算机 18 的大部分硬盘驱动器或网络带宽。另一方面,如果服务器应用 16 不忙,则其可以以更低程度使用服务器计算机 18 的资源。例如,如果服务器应用 16 没有正在存储或发送数据,则其可以以小程度使用服务器计算机 18 的硬盘驱动器和 / 或网络带宽 (如果有)。

[0028] 如果服务器应用 16 接受来自用户的输入或向用户提供输出,然后访问服务器应

用 16 的用户量可影响服务器应用 16 上的负载，并随后影响服务器应用 16 利用服务器计算机 18 的硬件和软件资源的程度。例如，如果许多用户正在访问和 / 或请求服务器应用 16，然后服务器应用 16 可需要服务器计算机 18 的多个 CPU 周期来服务于所述许多用户。同样，服务器应用 16 可需要多个硬件驱动器或网络带宽（或任意其他硬件或软件资源）来服务于所述许多用户。作为推理，如果仅几个用户正在访问服务器应用 18，然后服务器应用 18 可需要更少的硬件和软件资源来服务于用户。例如，如果服务器应用 16 是 web 服务器，并且如果许多用户尝试同时访问 web 服务器，然后服务器应用 16 可不得不使用服务器计算机 18 的多个 CPU 周期来处理请求，使用服务器计算机 18 的多个硬件驱动器带宽来提取所存储的 web 页面，以及使用服务器计算机 18 的多个网络带宽来向许多用户发送所请求的 web 页面。

[0029] 然而，如果存在太多用户访问或请求服务器应用 16，然后服务器应用 16 和服务器计算机 18 可能无法有效服务于所有用户。例如，服务器计算机 18 可用完 CPU 周期，并且可变得太忙以至于无法处理所有用户请求。备选地，某些用户请求可能被服务器应用 16 丢掉。这两种情况可使得期望访问正在快速和有效运行的服务器的用户无法满足。如果给定服务器计算机的软件和硬件资源，则可能存在服务器应用 16 可服务于用户的最佳、最大、和 / 或最小操作点。

[0030] 由此，服务器测试处理 10 可仿真 104 服务器应用 16 的虚拟用户。每个虚拟用户可以是仿真服务器应用 16 的用户的脚本、程序、或程序的一部分。服务器测试处理 10 可生成网络业务和 / 或服务器请求，并将他们发送至 106 服务器应用 16。对于每个虚拟用户，例如，服务器测试处理 10 可生成服务器应用 16 的用户典型的网络业务和 / 或服务器请求。例如，如果服务器应用 16 是 web 服务器，则测试处理 10 所仿真的虚拟用户可以是典型的 web 用户 - 他们可生成浏览 web 和访问 web 服务器的用户典型的 web 业务、web 请求和 web 下载，即服务器应用 16。

[0031] 服务器测试处理 10 可仿真 104 任意数目个用户，并且将他们的业务和请求发送至 106 服务器应用 16。例如，服务器测试处理 10 可仿真 104 许多用户（例如，对于特定情形的 100000 个用户），或几个用户（例如，对于特定情形的 1 个、或甚至 0 个用户）。这些被仿真的虚拟用户中的每个可将网络业务和请求发送至 106 服务器应用 16。这样，服务器测试处理 10 可仿真真实世界的服务器使用情形。例如，如果服务器应用 16（当安装在场内时）期望从 X 个数目的用户接收客户端请求时，然后服务器测试处理 10 可仿真 104X 个数目的虚拟用户，并将来自 X 个数目的被仿真的用户的客户端请求发送至 106 服务器应用 16。

[0032] 在一个使用情形下，开发者可期望通过确定多少个用户给服务器 18 到特定程度的负载将来最大化服务器计算机 18 的效率。换句话说，如果特定服务器的目标 CPU 负载被指定为 80%，则开发者可希望找出需要多少个访问服务器应用 16 的用户来创建 80% 的 CPU 负载。由此，服务器测试处理 10 可测量 108 服务器应用 16 的度量（例如 CPU 负载），将测量的度量与预订值相比较 110（例如 80%），并且如果测量的度量和预定值不同，则改变 112 被仿真的虚拟用户的数目。

[0033] 服务器测试处理 10 可改变 112 或调节虚拟用户的数目，直到被测量的度量与预定值的预设差额 (margin) 相同或在其内。在某些情况下，增加虚拟用户的数目可增加测量的度量的值。例如，服务器测试处理 10 可仿真 X 个数目的虚拟用户。如果目标 CPU 负载为

80%，并且如果服务器测试处理 10 测量 108CPU 负载在 60%，则服务器测试处理 10 可增加 X 个数目的虚拟用户，直到 CPU 负载到达 80% 或落入 80% 的预设差额（例如 +/-2%，或某些其他预设差额）。同样，如果测量的 CPU 负载大于目标 80%，则服务器测试处理 10 可减少 X 个数目的虚拟用户，直到 CPU 负载到达 80% 或落入 80% 的预设差额。

[0034] 在其他情况下，增加虚拟用户的数目可实际减少测量的度量的值。例如，采用测量的度量是 CPU 空闲时间的情况。如果目标 CPU 空闲时间为 20%，并且测量的 CPU 空闲时间为 30%，则服务器测试处理 10 可增加 X 个数目的虚拟用户，由此使得服务器计算机 18 的 CPU 更忙并减少空闲时间。在该情况下，服务器测试处理 10 可增加 X 个数目的虚拟用户，直到测量的 CPU 空闲时间减少到 20% 或落入 20% 的差额。同样，如果测量的 CPU 空闲时间为 10%，则服务器测试处理 10 可减少 X 个数目的虚拟用户，直到测量的 CPU 空闲时间增加到 20% 或落入 20% 的差额。

[0035] 测量的度量和预定值可以是服务器计算机 18 和 / 或服务器应用 16 的性能的任意测量或指示符。度量和值可以是 CPU 负载、CPU 空闲时间、硬盘驱动器带宽、硬盘驱动器空闲空间、网络使用、软件线程数、或表示服务器计算机 18 或服务器应用 16 的性能的任意其他度量。在一个实施例中，服务器测试处理 10 可通过访问在服务器计算机 18 上运行的 Windows 管理工具 (WMI) 或 rstat.d 接口来测量 108 度量。WMI 和 / 或 rstat.d 服务可以是提供服务器计算机 18 的性能度量的统计和测量的工具。服务器测试处理 10 可访问和测量 108 这些 WMI 和 / 或 rstat.d 度量。本领域普通技术人员将认识到，可使用提供性能度量的任意工具，包括但不限于：例如，WMI、rstat.d、定制测量工具、可测量服务器计算机 18 的性能的任意其他工具。定制测量工具可以是用于测量服务器计算机 18 的度量的整体或部分开发的软件程序和 / 或硬件工具。

[0036] 某些度量的值可改变或波动。这些波动可引起错误的或不精确的测量结果。感兴趣的测量可以是波动测量的总数或平均数，但是度量值的改变和波动可能独立于服务器测试处理 10。例如，如果 CPU 负载是波动度量，则服务器测试处理 10 所进行的顺序测量可能波动。也就是说，例如，目标 CPU 值为 80%，并且在服务器计算机 18 上的实际、中间 CPU 负载为 80%，但是由于测量错误或 CPU 负载测量的特征，服务器计算机 18 的 CPU 负载的任意个体测量可在 70% 和 90% 之间改变。在该情况下，服务器计算机 18 在期望负载下运行，但是任意给定测量可能落入 70% 和 90% 之间的任意比例。如果特定测量为 70%，则服务器测试处理可错误地增加虚拟用户的数目。如果特定测量为 90%，则服务器测试处理可错误地减少虚拟用户的数目。如果度量随着每个测量波动，则服务器测试处理 10 可随着每个元素改变 112 虚拟用户的数目，并且可能无法达到稳定状态调节。

[0037] 为了避免这样的波动，服务器测试处理 10 可多次测量 108 度量，并且可在存储器或计算机可读存储设备（例如存储设备 22）中存储 116 每个测量。如果服务器测试处理 10 存储 116 一系列测量，则服务器测试处理 10 可聚集 118 测量，以去除测量波动。服务器测试处理 10 可通过采用移动平均数 - 通过存储最后几个测量和使用最后几个测量的平均数来聚集 118 测量，以确定是否改变 112 虚拟用户的数目。服务器测试处理 10 也可对最后几个存储的测量采用低通滤波器，例如低通无限输入响应 (IIR) 或有限输入响应 (FIR) 滤波器，以去除测量中的波动。一般地，服务器测试处理 10 可从测量去除波动和差异的任意方式来聚集 118 测量。

[0038] 服务器测试处理 10 还可记录 114 与特定度量和 / 或特定运行点相关的用户数目。例如,假设设计者期望知晓需要多少个用户在服务器计算机 18 的 CPU 上创建 80% 的负载。在该情况下,在测试服务器计算机 18 的同时,服务器测试处理 10 可测量服务器计算机 18 的 CPU 负载,并改变虚拟用户的数目,直到服务器计算机 18 的 CPU 负载达到 80% (或到达 80% 的预设差额中的数目)。服务器测试处理 10 可发现其必须仿真 10000 个虚拟用户,用于服务器计算机 18 的 CPU 达到 80%。服务器测试处理 10 然后可记录 114 其发现的用户数目,即 10000,作为需要在服务器计算机 18 的 CPU 上施加 80% 负载的用户数目。服务器测试处理 10 可在存储器、存储设备、文件、数据库中、或用于存储数据的任意其他介质中记录 114 虚拟用户的数目。

[0039] 根据一个实施例,服务器测试处理 10 可用于确定达到服务器上的特定目标负载所需的用户数目。在另一简化实例中,如果服务器计算机 18 上的目标 CPU 负载为 50%,则服务器测试处理可仿真 10410000 个用户。服务器测试处理 10 可测量 108 服务器计算机 18 上的 CPU 负载,并发现在具有 10000 个虚拟用户时,服务器计算机 18 在 100% 下运行。为了达到 50%,服务器测试处理 10 可基于测量值 (100%) 和预定值 (50%) 的百分数或比例来改变 112 虚拟用户数目。例如,由于预定值 (50%) 是目标值 (100%) 的一半,则服务器测试处理 10 可将用户数目减少一半。由此,服务器测试处理 10 可将虚拟用户数目从 10000 减少到 5000 个虚拟用户。服务器测试处理 10 可重复该处理,增加和减少虚拟用户的数目,直到测量的度量等于预定值或落入预定值的范围内。

[0040] 描述了多个实施例。然而,以下权利要求限定了本发明的范围。

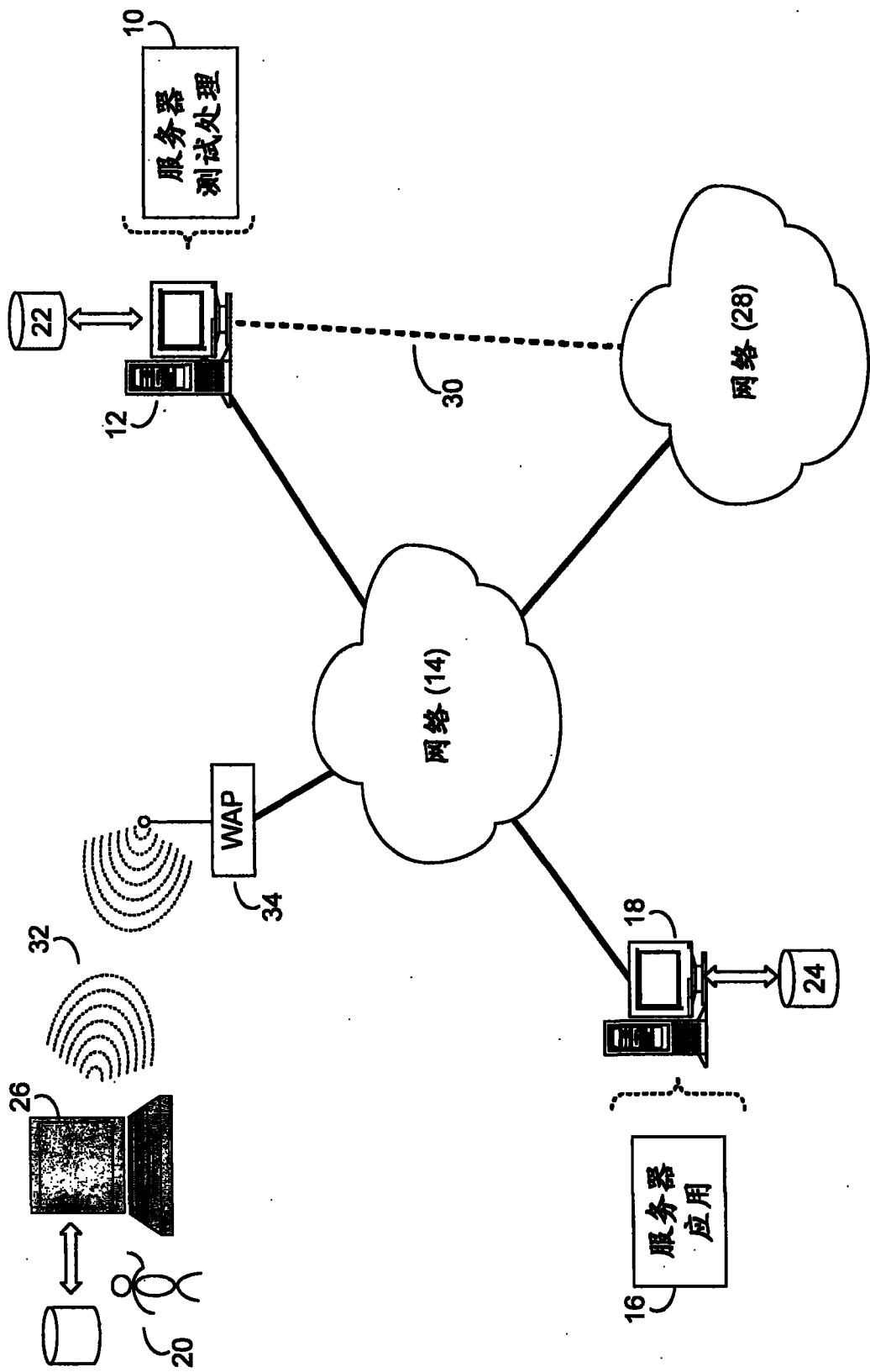


图 1

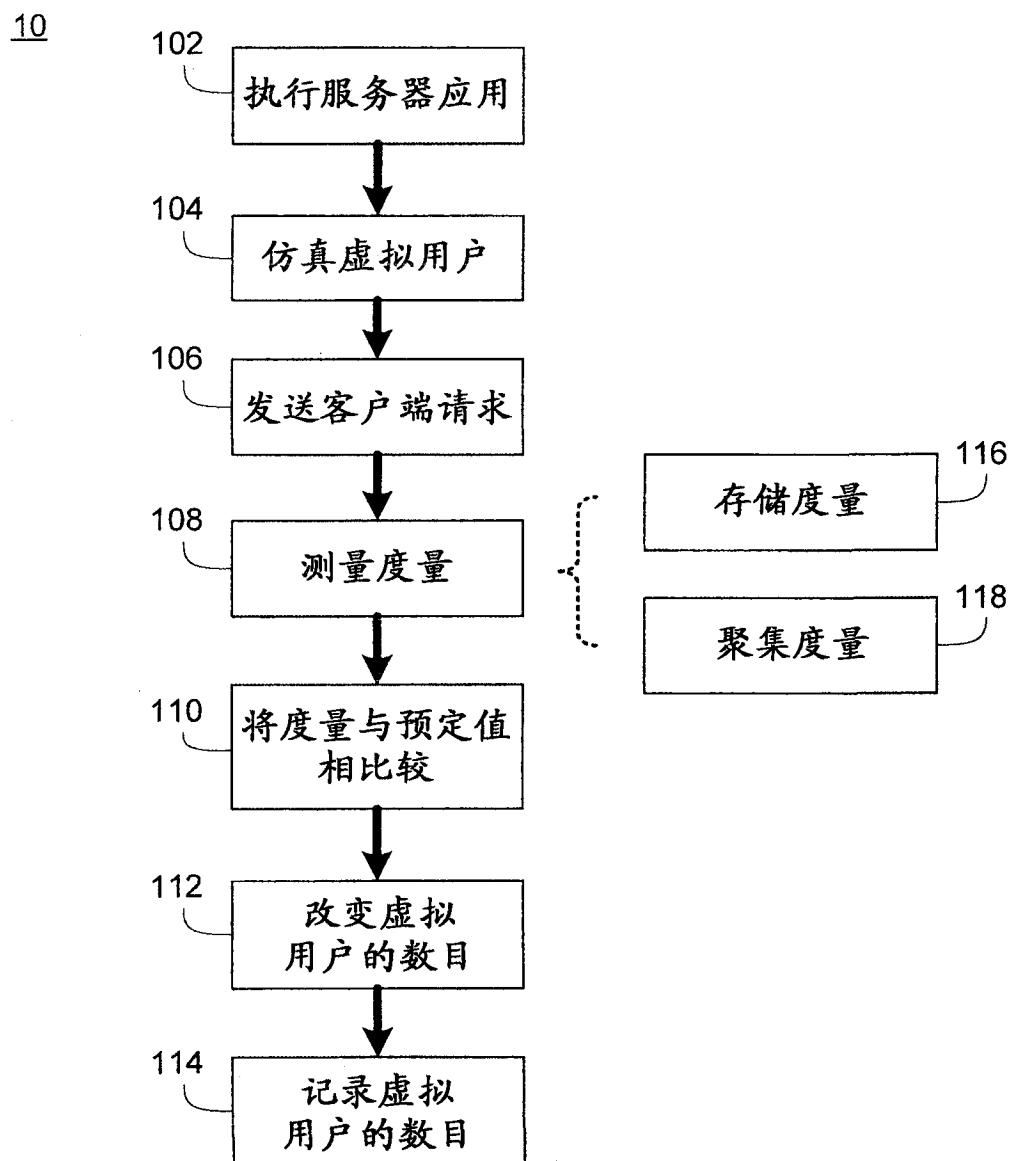


图 2