

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G06F 9/06

G06F 9/45



[12] 发明专利申请公开说明书

[21] 申请号 03122642.6

[43] 公开日 2003 年 12 月 17 日

[11] 公开号 CN 1461991A

[22] 申请日 2003.4.18 [21] 申请号 03122642.6

[30] 优先权

[32] 2002.5.30 [33] US [31] 10/159528

[71] 申请人 惠普公司

地址 美国加利福尼亚州

[72] 发明人 N·帕塔米安 L·莫里歇蒂

A·涅内 A·特里克

[74] 专利代理机构 中国专利代理(香港)有限公司

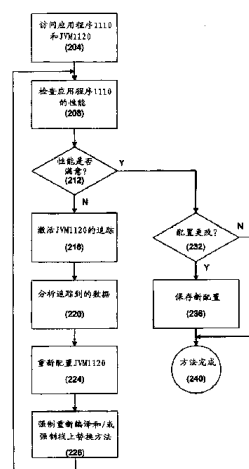
代理人 程天正 王忠忠

权利要求书 2 页 说明书 9 页 附图 3 页

[54] 发明名称 在计算机程序正在运行时改进它们的性能

[57] 摘要

本发明在各种实施方案中提供了用于改进程序性能的技术。在一种实施方案中，该程序以 Java 语言编写，并运行在包括具有一配置的 Java 虚拟机 (JVM) 的 Java 运行时环境 (JRE) 中。一个可访问该 JVM 的控制面板被调用。通过该控制面板，用户观察该 Java 程序的执行、分析执行的结果、改变 JVM 的配置并强制重新优化正在运行的程序的关键部分。通过重复上述观察该程序执行、分析结果、改变 Java 机的配置以及强制重新优化该程序的关键部分的步骤，该用户改进该程序的整体性能。



- 1、一种用于改进计算机程序(1110)的性能的方法,包括步骤:
运行该程序(1110),并因此运行一个包括影响该程序(1110)性能
5 参数的虚拟机(1120);该参数具有一个值;
调用可访问该程序执行、虚拟机(1120)和参数的控制面板;以及
通过该控制面板并且在该程序(1110)和虚拟机(1120)正在运行
时,
调整该参数的值为一个新值;
执行带有含新值的参数的程序(1110);以及
10 分析该执行步骤的结果,以及
基于该分析步骤的结果而采取一个行动。
- 2、权利要求1的方法,其中该程序(1110)以Java语言编写,并且
该虚拟机(1120)是Java虚拟机(1120)。
- 3、权利要求1的方法,其中采取行动的步骤包括重复该调整、执
15 行、分析和采取行动的步骤。
- 4、权利要求1的方法,其中采取行动的步骤是选自强制编译该程
序(1110)的方法、强制栈上替换该程序(1110)的方法、重新配置执
行该程序(1110)的计算机(300)的存储器堆、执行对计算机(300)
的存储器(308、312)的无用单元收集、以及重新配置该虚拟机(1120)
20 中的一个或者一个组合。
- 5、权利要求1的方法,其中该参数的值存储在该控制面板和虚拟
机(1120)可访问的共享存储器(308、312)中。
- 6、权利要求1的方法,其中调整步骤使用进程间通信协议。
- 7、权利要求1的方法,其中该参数是虚拟机(1120)配置的一部
25 分。
- 8、权利要求1的方法,进一步包括在通过网络(130、340、352)
与运行该程序的计算机(300)相连的显示设备(332)上显示该控制面
板的步骤。
- 9、权利要求1的方法,进一步包括为后面的使用而保存新值的步
30 骤。
- 10、权利要求1的方法,进一步包括激活该虚拟机(1120)的追踪
特性的步骤。

11、一种用于调谐虚拟机(1120)的方法,包括步骤:
执行程序(1110),并因此调用该具有影响该程序性能的配置的虚拟机(1120);
调用可访问该程序执行和该虚拟机(1120)的的配置的控制面板;以
5 及
通过该控制面板,并在该程序(1110)和虚拟机(1120)正在执行
时,
分析由该程序(1110)的执行而产生的数据,
基于分析步骤的结果,调整虚拟机(1120)的配置,以及
10 运行带有该虚拟机(1120)的已调整配置的程序(1110);
其中该虚拟机(1120)提供用于执行该程序(1110)的程序代码的
环境。

在计算机程序正在运行时改进它们的性能

技术领域

- 5 本发明总地涉及计算机程序，尤其涉及在这样的程序运行时改进它们的性能。

背景技术

- 起初可接受地运行的应用程序在执行了一段时间后，由于各种原因而变成差的执行者，上述原因包括程序的工作负荷特性的改变。例如在
10 Java 运行时环境（JRE）中，Java 虚拟机（JVM）依靠动态优化的编译器来加速 Java 应用程序的执行。在应用程序运行时，JVM 识别出调用最频繁的方法，即“热”方法，优化它们，并且把被优化的方法放入代码高速缓存中，从而会由于使用了这些方法的高速缓存版本而使得随后的对这些方法的调用能更有效。代码高速缓存是存储器中存储被优化的热
15 方法代码的指定区域。一组热方法通常与特定类型的负荷或该程序执行的动作集相关联。运行两种不同负荷的程序提供两种不同的热方法集。例如，在为个人预备税时，税务程序与一组热方法相关联，这些方法不同于与为公司预备税的同一税务程序相关联的一组热方法，因为为个人预备税的方法与为公司预备税的方法不同。因此，在程序负荷发生改变
20 时，通常会识别出热方法。由于额外的热方法被识别出并被优化，所以应用程序运行得更有效。但是，在应用程序运行了一段时间之后，热方法会达到一个稳态，即没有识别出新的热方法。应用程序性能就依赖于稳态方法被优化到多好。

- 在 Java 环境运送到客户处时，通常会预置影响应用程序优化的等
25 级和范围的阈值。积极（aggressive）的阈值和保守（conservative）的阈值都会导致性能差。例如，对于要在编译期间被直接插入的大量方法代码的积极的阈值会引起过多的指令高速缓存（I-高速缓存）未中，而保守的阈值则会留下较大部分未优化的代码。这两种情况都导致应用程序性能差。

- 30 典型地，当用户认识到正在执行的应用程序性能差时，该用户会检查该应用程序，识别出 JVM 中影响应用程序性能的参数，并且给这些参数重新分配更合适的值。但是，为了使新改变有效，必须安装基于该修

5 正的参数的新环境，并且这要求关掉现有的应用程序。在关掉该应用程序时，它不能被使用，这会导致事务中断。在各种情况中，为了纠正该问题，要求该用户与提供 Java 环境的企业机构合作。在这些情况中，该用户必须亲自到机构所在地，必须在那里重现该应用程序的软件和硬件环境以及导致性能差的状态，所有这些的代价是非常高的。在许多情况下，产生引起性能差的状态并不容易。另外，在应用程序被调试之后，新 Java 环境通常被修改，于是该用户必须在他的机器中安装包括这些修改的新 Java 环境。这也要求关掉该应用程序，而这将导致了中断，就像上面讨论的。

10 基于上述的内容，希望提供机制来解决上面的不足之处以及有关问题。

发明内容

15 在各种实施方案中，本发明提供用于改进程序性能的技术。在一种实施方案中，以 Java 语言写该程序，并且在包括含一种配置的 Java 虚拟机 (JVM) 的 Java 运行时环境 (JRE) 中运行该程序。可访问该 JVM 的控制面板被调用。通过这个控制面板，用户观察 Java 程序的执行，分析执行结果，改变 JVM 的配置以及强制正在运行的程序的关键部分重新优化。通过重复上面观察该程序执行、分析结果、改变 Java 机的配置和强制该程序的关键部分重新优化的这些步骤，该用户能改进该程序
20 的整体性能。

附图说明

本发明通过附图中的实例来说明，在附图中，同样的参考数字指出同样的元素，并且在附图中：

25 图 1 显示了可以在其上实施本发明的实施方案的系统；
图 2 是说明用于根据一种实施方案来改进程序性能的方法的流程图；以及
图 3 是可以实施本发明的实施方案的计算机系统。

具体实施方式

30 在下列描述中，为了解释，众多特殊细节被陈述以提供对本发明的彻底理解。但是，显而易见的是本领域中的技术人员没有这些特殊细节也可以实践本发明。在其它实例中，众所周知的结构和设备都以框图的形式示出以避免遮蔽本发明。

图 1 显示了本发明的实施方案可以在其上实施的系统 100。系统 100 包括第一计算机 110 和第二计算机 120，为说明而把它们分别称为应用程序服务器 110 和性能调试机 120。服务器 110 和机器 120 通过通信链路 130 连接。

5 服务器 110 运行各种应用程序，其中之一显示为程序 1110。在一种实施方案中，程序 1110 是在包括 Java 虚拟机 (JVM) 1120 的 Java 运行时环境 (JRE) 中工作的 Java 程序。典型地，程序 1110 在 JVM 1120 的控制下优化和执行。程序 1110 进而又通过网络 (比如通信链路 130、互联网等) 正常地提供服务给用户。例如，这些服务包括万维网、数据库、电子邮件、安全、通信等。

10 Java 虚拟机 1120 提供一种用于执行与运行 JVM 1120 的计算机平台无关的 Java 字节码的环境。通常 JVM 1120 以运行在服务器 110 的硬件和操作系统上的软件来实施。因而，JVM 1120 提供一种环境允许以将在服务器 110 上执行的字节码的形式来表示一般程序。JVM 1120 负责优化该 Java 程序以及把 Java 字节码翻译成由服务器 110 执行的动作。JVM 1120 包括各种影响像程序 1110 这样的应用程序性能的参数。

15 机器 120 包括控制面板 1220 能在其上显示的监视器或屏幕。在图 1 这个实例中，控制面板 1220 与服务器 110 离得很远，即控制面板 1220 通过通信链路 130 来与服务器 110 通信。但是，控制面板 1220 也可以是本地的，例如在服务器 110 本地的监视器上显示。在一种实施方案中，控制面板 1220 是从 JVM 1120 处调用的软件包，它允许用户访问并修改 JVM 1120。另外，通过控制面板 1220 并以图像用户接口 (GUI) 的形式，与提供该 Java 环境的机构相关联的客户支持工程师远程分析并改进程序 1110 的性能。

25 通信链路 130 是用于服务器 110 与机器 120 通信的机制。通信链路 130 可以是单个网络或者是利用一种通信协议或通信协议组合的网络组合，这些通信协议是诸如传输控制协议/互联网协议 (TCP/IP)、公共交换电话网 (PSTN)、数字用户线路 (DSL)、电缆网、卫星兼容的、无线兼容的等。在一种实施方案中，通信链路 130 是互联网。

30 Java 虚拟机的配置

Java 虚拟机 1120 执行各种功能，并包括各种参数，这些功能和参数都影响程序 1110 的性能。例如，JVM 1120 执行管理服务器 110 的存

存储器中数据的无用单元收集，以使得正在运行的 Java 程序不会用完存储器。管理存储器要花费时间和资源，因而会影响程序 1110 的性能。改变无用单元收集的频率会改变程序 1110 的性能。在一种实施方案中，JVM 1120 包括指示无用单元收集应执行的频率的阈值以及指示执行无用单元收集的存储器块的优先级的机制。堆 (heap) 指的是无用单元收集的对象被放置的存储位置。在堆已满时，启动无用单元收集。较小的堆会引起更频繁地进行无用单元收集，并且因此可降低系统性能。JVM 1120 包括用于调整堆大小的参数。

“热” Java 程序方法被十分频繁地调用，并且以编译形式执行，而不是以解释形式执行，因为执行被编译的方法的速度比执行被解释的方法更快。在一种实施方案中，JVM 1120 包括确定一种方法是否是热方法的阈值。例如，阈值 5000 指示只有至少被调用 5000 次后一个方法才被认为是热方法。同样，阈值 10000 指示只有至少被调用 10000 次后该方法才被认为是热方法等等。由于热方法在程序中被频繁调用，所以一组热方法可以被称为是该程序的关键部分。

Java 方法可以不同的积极性等级编译。被更积极地编译的方法能有更大范围或更大区域的代码被考虑来优化，并且因此与被较不积极编译的方法相比，有更多的代码被编译。这可导致需要花更长时间来编译和/或导致更高的指令高速缓存未命中率，因为有大量代码的被更积极编译的方法不能像有较少代码的方法那样常常被存储在高速缓存中。在一种实施方案中 JVM 1120 包括各种标记用于用户设定编译方法的积极性。JVM 1120 也允许用于在要求时强制一个方法被编译的机制。在一种实施方案中，由于编译器是“优化编译器”，所以强制重新编译可以被认为是强制优化，并且优化什么和/或优化多少是基于优化编译器的参数和/或阈值。另外，由于热方法可以被认为是程序的关键部分，因此热方法的强制重新编译可以被认为是程序的强制优化。

JVM 1120 也包括强制栈上替换 (on-stack replacement) 的机制，它指的是在该程序的一部分正在运行时替换它，通常这是用该部分或者该方法的更好的优化版本来进行的。

当 JVM 1120 在服务器 110 中是第一次安装时，它被配置到一组参数，并且因此执行一些预先定义的功能。但是，在应用程序 1110 和 JVM 1120 正在运行时，用户能比如通过控制面板 1220 来调整这些参数。

修改 Java 虚拟机的配置

本发明的实施方案允许一个用户实时或“动态地”改变 JVM 1120 的配置，这通常指的是程序 1110 正在运行或正在被执行的事实。改变 JVM 1120 的配置包括调整各种参数和阈值，并且在想要调谐、调试和/或改进应用程序 1110 和/或 JVM 1120 的性能时应用。为了说明，在本文档中使用术语改进性能。但是，该说明性的概念可应用于改进、调谐、调试和它们的等价体。典型地，调谐 JVM 1120 出现在以下情况中：有新功能性被添加到应用程序 1110、应用程序 1110 的负荷改变、引入新的 JVM 1120 版本、新应用程序第一次通过 JVM 1120 运行等。

10 经由控制面板 1220，用户能够激活 JVM 1120 的追踪，例如通过把 JVM 1120 内部的追踪特性设定为它的“开”位置。通过追踪，该用户获得有用的信息来识别出性能差的原因。然后该用户分析获得的信息并在适当时改变 JVM 1120 的配置。改变 JVM 1120 的配置包括调整各种阈值，这些阈值控制 JVM 1120 执行各种动作的频率和持续时间，或者直接影响程序 1110 的优化等级，比如通过 JVM 1120 的优化编译器处理该程序的那些部分等。例如，该用户可以改变无用单元收集的频率并即刻激活该新频率、在代码高速缓存中再生该代码等。

依靠参数和实现，JVM 1120 中的每个参数都与一个控制机制相对应，比如显示在控制面板 1220 上的开/关按钮、滑动条、选择列表等。20 用户调整相应的控制机制以调整该参数。例如为了强制编译或者强制栈上替换，一种方法对应于一个开/关按钮，并且点击这个按钮使得相应的强制编译或强制栈上替换启动或禁止。又比如，通过对应于该阈值范围的滑动条来表示设定阈值的参数，比如设定触发无用单元收集的阈值，或指示方法是热方法的阈值等。调整滑动条就能调整相应的阈值。

25 在一种实施方案中，期望在程序 1110 运行时可调整的参数都存储在服务器 110 的共享存储器中，它指的是能被一个以上进程修改的存储器区域。控制面板 1220 和 JVM1120 被认为是可访问这个共享存储器的两个进程。通过作为一个进程的控制面板 1220，用户可以改变共享存储器中的参数值，并且因此允许应用程序 1110 立即以新改变的参数执行，30 这导致了实时改进应用程序 1110 的性能。在可选实施方案中，该参数存储在 JVM 1120 的数据空间中，并且控制面板 1220 和 JVM 1120 仍然是两个进程。然后该用户使用进程间通信（IPC）协议来直接调整 JVM

1220 的数据空间中的参数。在可选实施方案中，控制面板 1220 是 JVM 1120 的一部分，并且用户通过控制面板 1220 改变来自 JVM 1220 的数据空间的参数。一个进程可以被定义为一个执行的程序或任务，它指的是一个程序和操作系统所使用的簿记信息 (book keep information) 5 的组合。该任务通过一个任务号来识别该程序。一个进程可以在一个能独立于该进程的其它部分而执行的线程中创建。

用户通过控制面板 1220 可以实时观察到程序 1110 的性能，并采取适当的措施来改进这样的性能。例如，该用户激活 JVM 1120 的追踪特性，这种特性允许用户看到程序 1110 和 JVM 1120 正在如何运行、识别 10 程序 1110 的哪些方法被编译而哪些被解释、高速缓存未中率是多少、各种参数的阈值设定为什么值等等。基于该观察和实时获得的数据，该用户可以通过分析现有的数据来解决这些问题，通过以 JVM 1120 的各种新数据和新配置集来运行程序 1110 而试验等等。需要时，该用户使用可获得性能工具来分析数据、程序 1110、JVM 1120 等。性能分析 15 工具的实例包括加利福尼亚的 Santa Clara 的 Intel 的 Prospect、Gprof、Vtune，加利福尼亚的 Palo Alto 的惠普公司的 HPJmeter 和 Glance 等。例如，如果观察到一种方法被调用得十分频繁，但是它的相应热方法阈值被设定为一个太高的值，那么该用户减少这个值，从而使该方法被认为是热方法，并且因此以编译形式运行，而不是以解释的形式运行。相反，如果由于该方法被编译得太积极而使得该方法引起太多的高速缓存未中，那么该用户就以较不积极的选项重新编译该方法。为了帮助该调试进程，该用户强制方法 A 的编译，并且/或者强制方法 B 的栈上替换等。

在一种实施方案中，每次调试会话之后，JVM 1120 的配置模板都会被创建并例如在文件中存储，该模板保存有 JVM 1120 中各种参数的新 25 值。因此每个模板通常都与一个特殊的并且意义明确的程序负荷以及调试和运行应用程序 1110 的环境相关联。配置模板提供给客户一种重新配置要在类似情况下使用的 JVM 1120 的快速方法。例如，在应用程序 1110 正在忙于运行数据库请求时创建模板 T1，在应用程序 1110 正在运行万维网业务时创建模板 T2 等等。当运行数据库或万维网业务的应用程序出现时，该用户能够分别快速地载入模板 T1 或 T2。 30

说明改进程序性能的方法的步骤

图 2 是说明用于根据一种实施方案来改进程序性能的方法的流程图。

在步骤 204, 用户使用控制面板 1220 来访问应用程序 1110 和 Java 虚拟机 1120。

5 在步骤 208, 该用户检查应用程序 1110 的性能。

在步骤 212, 该用户确定是否满意应用程序 1110 的性能。如果不满意这个性能, 那么该用户在步骤 216 激活 JVM 1120 的追踪。

在步骤 220, 该用户分析追踪到的数据。

10 在步骤 224, 该用户重新配置 JVM 1120, 例如通过改变各种阈值和 /或参数来进行。

在步骤 228, 必要时, 该用户强制重新编译热方法和 /或强制栈上替换一些当前正在运行的方法等等。

15 然后该用户重复上面的步骤 208、212、216、220、224 和 228, 直到该用户在步骤 212 确定满意应用程序 1110 的性能。当满意该性能时, 该用户在步骤 232 确定 JVM 1120 的配置是否已经被更改, 并且如果已经更改, 那么该用户在步骤 236 保存该新的配置。该方法就在步骤 240 完成。但是如果 JVM 1120 的配置尚未更改, 那么该用户在步骤 236 跳过保存该配置的步骤, 并且在步骤 240 完成该方法。

20 本发明的实施方案提供了几种优点。为了说明, 使用了术语调谐, 但是这些优点可应用于其它类似情况, 比如改进性能、调整参数等。调谐应用程序 1110 能够相对本地应用程序 1110 的运行位置远程地在实际负荷下实时执行。没有必要建立单独的调谐环境。在调谐变得简单、容易并且廉价时, 可以更频繁地执行它。

计算机系统概述

25 图 3 是其上可以实施本发明的一个实施方案的计算机系统 300。例如计算机系统 300 可以实施为作为服务器 110 或作为调试机 120 等来操作。在一种实施方案中, 计算机系统 300 包括中央处理单元 (CPU) 304、随机存取存储器 (RAM) 308、只读存储器 (ROM) 312、存储设备 316 和通信接口 320, 所有这些都与总线 324 相连。

30 CPU 304 控制逻辑、处理信息并协调计算机系统 300 中的活动。在一种实施方案中, CPU 304 执行存储在 RAM 308 和 ROM 312 中的指令, 例如通过协调从输入设备 328 到显示设备 332 的数据移动来进行。CPU

304 可以包括一个或多个处理器。

通常被称为主存储器的 RAM 308 临时存储 CPU 304 要执行的信息和指令。RAM 308 中的信息可以从输入设备 328 处获得或由 CPU 304 产生，作为 CPU 304 执行的指令所要求的算术进程的一部分。

- 5 ROM 312 存储信息和指令，它们一旦被写入 ROM 芯片中就是只读的并且不能被修改或删除。在一种实施方案中，ROM 312 存储用于计算机系统 300 的配置和初始操作的命令。

像软盘、磁盘驱动器或磁带机这样的存储设备 316 能持久地存储计算机系统 300 使用的信息。

- 10 通信接口 320 使得计算机系统 300 能够与其它计算机或设备对接。例如，通信接口 320 可以是例如调制解调器、综合业务数字网 (ISDN) 卡、局域网 (LAN) 端口等。本领域中的技术人员将认识到调制解调器或 ISDN 卡通过电话线来提供数据通信，而 LAN 端口通过 LAN 来提供数据通信。通信接口 320 也可以允许无线通信。

- 15 总线 324 可以是任何用于传送计算机系统 300 所使用的信息的通信机制。在图 3 的实例中，总线 324 是用于在 CPU 304、RAM 308、ROM 312、存储设备 316、通信接口 320 等之间传送数据的媒体。

- 典型地，计算机系统 300 与输入设备 328、显示设备 332 和光标控制 336 耦合。输入设备 328 比如是包括字母数字键和其它键的键盘，可以
20 把信息和命令传送给 CPU 304。显示设备 332，比如阴极射线管 (CRT) 把信息显示给计算机系统 300 的用户。光标控制 336，比如鼠标、跟踪球或光标方向键可以把方向信息和命令发送给 CPU 304 并控制显示设备 332 上的光标移动。

- 计算机系统 300 可以通过一个或多个网络与其它计算机或设备通
25 信。例如利用通信接口 320，计算机系统 300 可以通过网络 340 与另一个连接到打印机 348 上的计算机 344 通信，或者通过环球网 352 与服务器 356 通信。环球网 352 通常被称为互联网。作为选择，计算机系统 300 可以通过网络 340 来访问互联网 352。

- 计算机系统 300 可以被用来实施上面描述的技术。在各种实施方案
30 中，CPU 304 通过执行带到 RAM 308 中的指令来实现技术步骤。在可选实施方案中，可以用硬连线电路来替代软件指令或与软件指令组合以实施描述的技术。因此，本发明的实施方案不受限于软件、固件、硬件或

电路中的任何一个或其组合。

由 CPU 304 执行的指令可以存储在一个或多个计算机可读媒体中，和/或通过其来承载，该计算机可读媒体是指计算机从中读取信息的任何媒体。例如，计算机可读媒体可以是软盘、硬盘、zip 驱动盒、磁带、5 或任何其它磁媒体、CD-ROM、CD-RAM、DVD-ROM、DVD-RAM、或任何其它光学介质、纸带、穿孔卡片、或任何其它有孔图案的物理媒体、RAM、ROM、EPROM、或任何其它存储芯片或盒式盘。计算机可读媒体也可以是同轴电缆、铜线、光纤、声波或电磁波、电容耦合或电感耦合等等。作为一个实例，CPU 304 要执行的指令是以一个或多个软件程序的形式并10 最初存储在通过总线 324 与计算机系统 300 对接的 CD-ROM 中。计算机系统 300 把这些指令载入 RAM 308 中、执行一些指令并通过通信接口 320、调制解调器和电话线发送一些指令给网络，例如网络 340、互联网 352 等。通过网络电缆接收数据的远程计算机执行接收到的指令，并把该数据发送给计算机系统 300，以将其存储在存储设备 316 中。

15 在上述说明中，已经参考特定的实施方案描述了本发明。但是，显而易见的是可以在不违背本发明的主要精神和范围的情况下进行各种修改和改变。因而，本说明书和附图应被认为是说明性的，而不是限制性的。

20

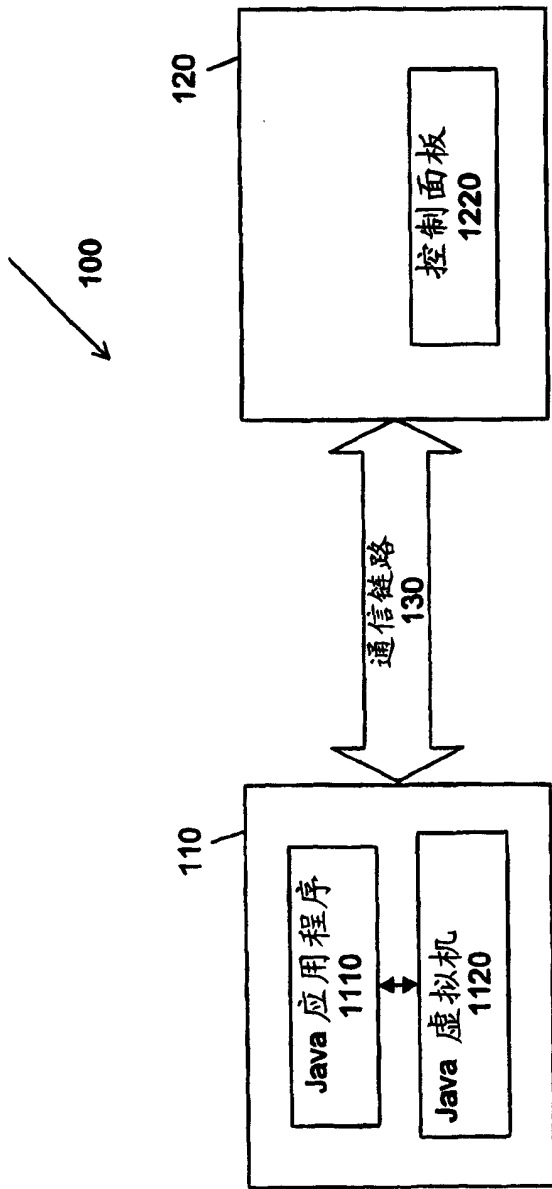


图 1

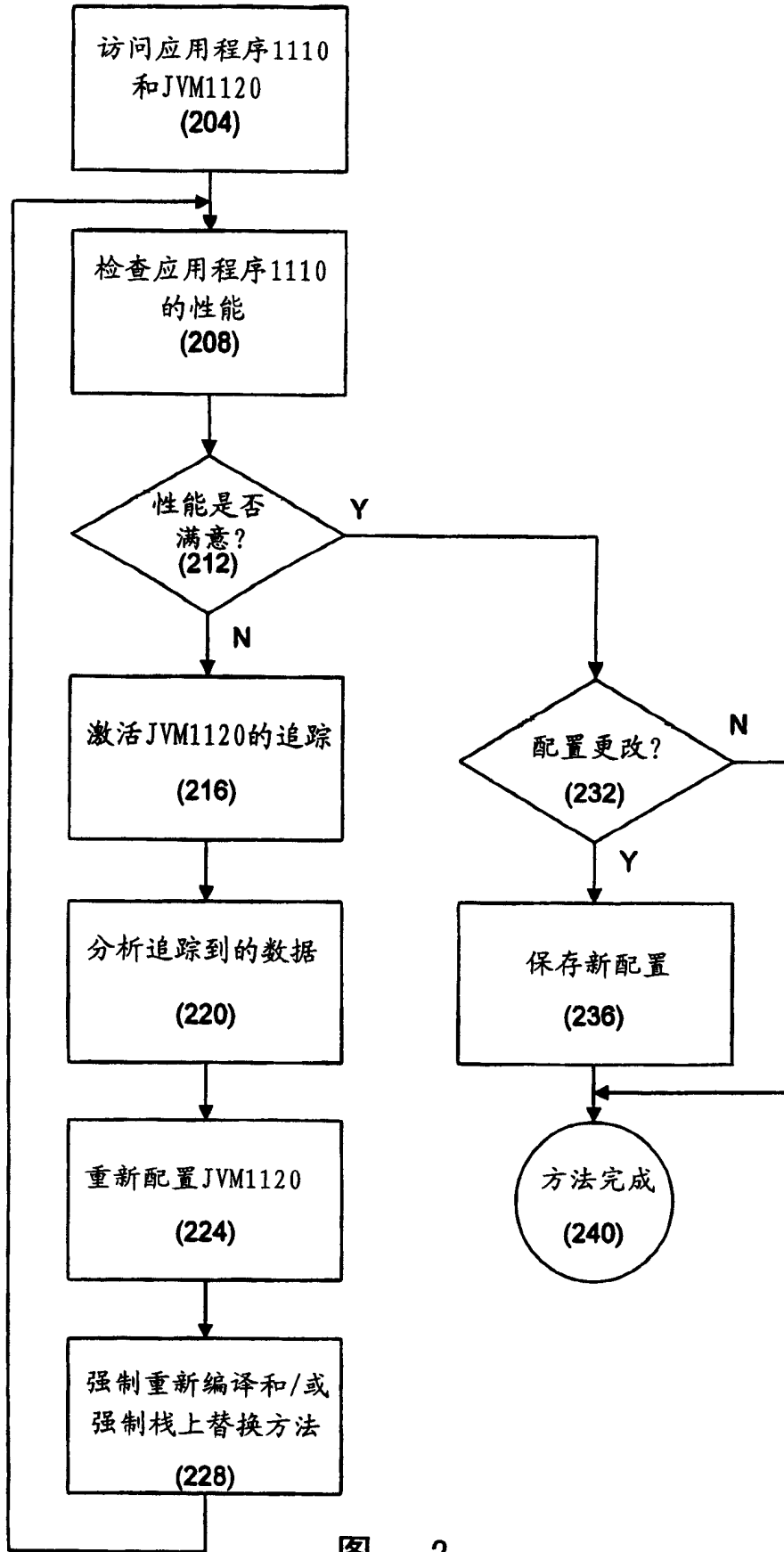


图 2

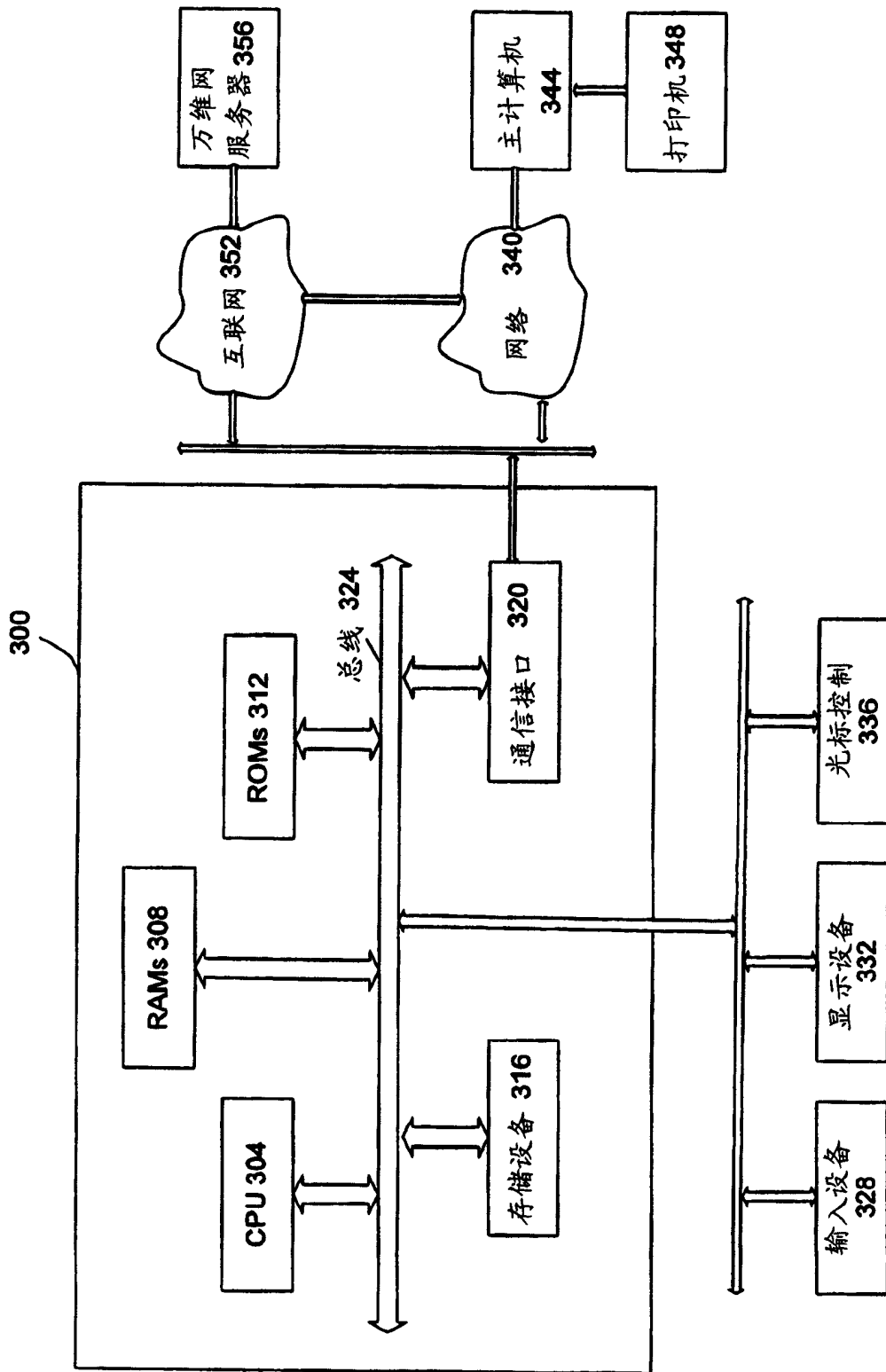


图 3