(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101493785 B
(45)授权公告日 2014.06.18

(21)申请号 200910128146.3

(22)申请日 2005.06.28

(30)优先权数据
10/880,057 2004.06.28 US

(62)分案原申请数据
200510082092.3 2005.06.28

(73)专利权人 英特尔公司
地址 美国加利福尼亚州

(72)发明人 V·乌利希 S·贝内特
E·科塔-罗布尔斯 S·舍恩贝格
A·安德森 R·乌利希 G·奈格

(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 张欣

(51)Int.Cl.
G06F 9/455(2006.01)
G06F 9/46(2006.01)

(56)对比文件
WO 2002052404 A2,2002.07.04, 说明书第7页第11行-第8页第6行、第9也第9-10行、第10页第17行至第11页第7行、第14页第14行至第15页第14行、第17页第18行至第18页第2行以及附图1、7.

WO 2002052404 A2,2002.07.04, 说明书第7页第11行-第8页第6行、第9也第9-10行、第10页第17行至第11页第7行、第14页第14行至第15页第14行、第17页第18行至第18页第2行以及附图1、7.

CN 1506861 A,2004.06.23, 说明书第5页第3段、第6段以及附图1.

US 4787031 A,1988.11.22, 摘要第6-10行、摘要第13-15行、第6栏第1-3行、第8栏第67行-第9栏第15行、第10栏第26行-第10栏第37行、第11栏第35-40行、第15栏第1-5行、第18栏第35-37行以及附图2A、2B.

US 2001045061 A1,2001.11.29, 全文.

审查员 张涛

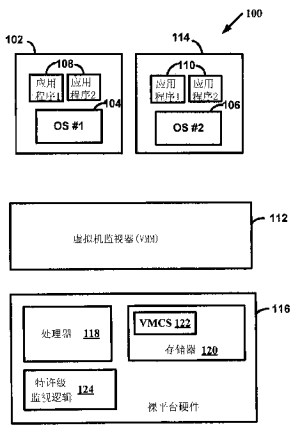
权利要求书2页 说明书8页 附图8页

(54)发明名称

根据客户软件的特许级支持向虚拟机监视器转移

(57)摘要

一种在客户软件当前特许级符合特许级退出判据时让虚拟机监视器承担系统控制的系统与方法。处理器检测当前特许级符合该判据。然后把控制从客户软件转到虚拟机监视器，而后者可对某些特许级变化或值拒绝承担控制。



CN 101493785 B

1. 一种用于根据客户软件的特许级支持向虚拟机监视器转移的方法，包括：

确定客户软件的初始特许级；

评估在虚拟机中执行的所述客户软件的当前特许级；以及

如果当前特许级符合特许级退出判据，则把控制转到虚拟机监视器，其中判断当前特许级是否符合特许级退出判据包括：如果启动了特许级改变时退出的控制策略，并且当前特许级不同于初始特许级，则把控制转到虚拟机监视器。

2. 如权利要求 1 所述的方法，还包括：

在虚拟机控制结构中存储所述特许级退出判据。

3. 如权利要求 1 所述的方法，还包括：

确定客户软件的初始特许级；以及

如果启动了特许级增高时退出的控制策略，并且当前特许级高于初始特许级，则把控制转到虚拟机监视器。

4. 如权利要求 1 所述的方法，还包括：

确定客户软件的初始特许级；以及

如果启动了特许级减低时退出的控制策略，并且当前特许级低于初始特许级，则把控制转到虚拟机监视器。

5. 如权利要求 1 所述的方法，还包括：

确定客户软件的初始特许级；以及

如果启动了与所述初始特许级到所述当前特许级的转换相关联的特许级源目的退出控制策略，则把控制转到虚拟机监视器。

6. 如权利要求 1 所述的方法，还包括：

存储目标特许级。

7. 如权利要求 6 所述的方法，还包括：

如果启动了特许级匹配时退出的控制策略，且当前特许级与目标特许级相同，则把控制转到虚拟机监视器。

8. 如权利要求 6 所述的方法，还包括：

如果启动了特许级达上限时退出的控制策略，且当前特许级比目标特许级高，则把控制转到虚拟机监视器。

9. 如权利要求 6 所述的方法，还包括：

如果启动了特许级达下限时退出的控制策略，且当前特许级比目标特许级低，则把控制转到虚拟机监视器。

10. 如权利要求 6 所述的方法，其特征在于，所述目标特许级可变。

11. 如权利要求 6 所述的方法，其特征在于，所述目标特许级固定。

12. 如权利要求 1 所述的方法，还包括：

在执行客户软件中的指令之前将控制转到虚拟机监视器。

13. 如权利要求 1 所述的方法，还包括：

在执行客户软件中的指令之后将控制转到虚拟机监视器。

14. 一种用于根据客户软件的特许级支持向虚拟机监视器转移的设备，其特征在于包括：

用于确定客户软件的初始特许级的装置；

用于评估在虚拟机中执行的所述客户软件的当前特许级的装置；

用于若当前特许级符合特许级退出判据则把控制转到虚拟机监视器的装置，其中用于判断当前特许级是否符合特许级退出判据的装置包括：用于如果启动了特许级改变时退出的控制策略且当前特许级不同于所述初始特许级则把控制转到虚拟机监视器的装置。

15. 如权利要求 14 所述的设备，还包括：

用于在虚拟机控制结构中存储所述特许级退出判据的装置。

16. 如权利要求 14 所述的设备，其中还包括：

用于确定客户软件的初始特许级的装置；以及

用于如果启动了特许级增高时退出的控制策略并且当前特许级高于初始特许级则把控制转到虚拟机监视器的装置。

17. 如权利要求 14 所述的设备，其中还包括：

用于确定客户软件的初始特许级的装置；以及

用于如果启动了特许级减低时退出的控制策略并且当前特许级低于初始特许级则把控制转到虚拟机监视器的装置。

18. 如权利要求 14 所述的设备，其中还包括：

用于确定客户软件的初始特许级的装置；以及

用于如果启动了与所述初始特许级到所述当前特许级的转换相关联的特许级源目的退出控制策略则把控制转到虚拟机监视器的装置。

19. 如权利要求 14 所述的设备，还包括：

用于存储目标特许级的装置。

20. 如权利要求 19 所述的设备，其中还包括：

用于如果启动了特许级匹配时退出的控制策略且当前特许级与目标特许级相同则把控制转到虚拟机监视器的装置。

21. 如权利要求 19 所述的设备，其中还包括：

用于如果启动了特许级达上限时退出的控制策略且当前特许级比目标特许级高则把控制转到虚拟机监视器的装置。

22. 如权利要求 19 所述的设备，其中还包括：

用于如果启动了特许级达下限时退出的控制策略且当前特许级比目标特许级低则把控制转到虚拟机监视器的装置。

23. 如权利要求 19 所述的设备，其特征在于，所述目标特许级可变。

24. 如权利要求 19 所述的设备，其特征在于，所述目标特许级固定。

25. 如权利要求 14 所述的设备，还包括：

用于在执行客户软件中的指令之前将控制转到虚拟机监视器的装置。

26. 如权利要求 14 所述的设备，还包括：

用于在执行客户软件中的指令之后将控制转到虚拟机监视器的装置。

根据客户软件的特许级支持向虚拟机监视器转移

[0001] 本申请是申请人于 2005 年 6 月 28 日提交的，申请号为 200510082092.3 的，发明名称为“根据客户软件的特许级支持向虚拟机监视器转移”的发明专利申请的分案申请。

[0002] 发明背景

[0003] 本发明诸实施例涉及虚拟机，更具体地说，本发明诸实施例涉及虚拟机 (VM) 中运行的客户软件的特许级符合特许级退出判据时，让虚拟机监视器 (VMM) 夺回对处理器的控制。

[0004] 普通 VMM 可在计算机上运行，向其它软件呈现一个或多台虚拟机的概况。每台 VM 都可当作一个独立的平台，运行其自己的操作系统 (OS) 诸如“客户操作系统”和应用程序，这统称为“客户软件”。比如说，客户软件准备在 VM 里或 VM 上运行。客户软件期望像在专用计算机上而不是在 VM 上运行那样的操作，即客户软件希望控制各种计算机操作，并在操作时访问硬件资源。硬件资源包括控制寄存器等处理器固有的资源和诸如描述符表格等留驻在存储器里的资源。但在虚拟机环境内 VMM 应能最终控制这些资源，使 VM 正常操作并加以保护。为此，VMM 通常拦截和仲裁客户软件对硬件资源的所有访问。

[0005] 大多数指令组体系结构 (ISA) 规定了把较少特许的应用程序与更富特许的操作系统功能相隔离的多种特许级，例如一种原有技术的 32 位结构有四种特许级，称为环 0 ～环 3，环 0 最具特许，环 3 最少特许。处理器在不同特许级之间提供受控的切换法。切换通过调用专用指令可以是直接明示的，或通过提出异常或故障，或运用中断等外部事件可以是暗示的，例如在执行诸如调用 (CALL)、软件中断 (INT) 或中断返回 (IRET) 的指令时，可能出现特许级变化。由于其它的同步或异步事件，例如像异常、外部中断、故障、任务切换、陷阱和其它类似事件，也会发生特许级变化。

[0006] 多处理器或多线程系统的操作系统使用软件构成的保证相互排斥的锁来保护可能被一个以上线程或处理器同时访问的数据。通常在短时间持有锁的情况下，可用所谓的旋转锁。当在一台处理器或一线程上操作的软件试图获取已被在另一台处理器或一线程上操作的软件取得的锁时，该软件就设法在紧密代码循环 (tight code loop) 中重新获取该锁。虽在该紧密循环中运行，但软件并不执行任何有用的工作，硬件处理器线程并无益处。在多线程处理器或多个处理器系统上，一个线程或处理器的运行会取走其它线程或处理器的资源，诸如耗用带宽、执行单元或功率。因此，自旋周期要尽可能短。

[0007] 虚拟多处理器或多线程系统的 VMM 可在独立的 VM 或虚拟处理器 (VP) 中执行运行于每种客户软件实例的软件。在非 VM 系统上，这些客户软件实例应在截然不同的处理器或多线程上执行。VP 与所有客户软件实例都统称为虚拟系统。在不考虑这种客户锁定特性时，这种 VMM 会明显劣化，因此当 VP 保持锁定时 VMM 不得抢占该 VP，除非要为该虚拟系统抢占所有的 VP。因为当用软件来实现锁时锁定原函数不能被硬件直接检测，可以应用直观推断法或间接观察法。

[0008] 一种这样的直观推断法基于公共 OS 特性。在 OS 不执行在特许模式操作中或处于低功率状态时，该 OS 可能保持不锁定。VMM 可以利用这一认知，只抢占正执行在非特许模式或处于低功率态的虚拟处理器。推迟对以特许模式操作的客户软件的抢占，直到客户软件

切换到非特许模式之后。

附图简介

[0009] 图 1 示出可在其上实施本发明一个实施例的处理系统一实施例的框图。

[0010] 图 2 在流程图中示出本发明一实施例的特许级退出控制的过程，各独立控制用于增减特许级。

[0011] 图 3 在流程图中示出本发明一实施例的特许级退出控制过程，组合的控制用于增减特许级。

[0012] 图 4 在流程图中示出本发明一实施例检测特许级变化的过程。

[0013] 图 5 在流程图中示出本发明一实施例的异步事件处理过程。

[0014] 图 6 在流程图中示出根据客户软件在虚拟机环境中的特许级从客户软件转到虚拟机监视器的过程的一实施例。

[0015] 图 7 在流程图中示出根据客户软件在虚拟机环境中的特许级从客户软件转到虚拟机监视器过程的一实施例。

[0016] 图 8 在流程图中示出在虚拟机环境中实行抢占调度过程的一实施例。

[0017] 详细描述

[0018] 提出一种让虚拟机监视器 (VMM) 在虚拟机 (VM) 上运行的客户软件的特许级符合一定特许级退出判据时保证对系统的控制的系统与方法。处理器检出客户软件的特许级符合该判据，于是将控制转向 VMM。特许级退出判据包括特许级增大、特许级减小、特许级任何变化、匹配预定等级的特许级、大于预定等级的特许级、小于预定等级的特许级或特许级的特定转换。VMM 可以拒绝承担对某些特许级值或值变化的控制。在 VM 中执行任何指令之前或之后，都可以转向 VMM。

[0019] 本发明诸实施例还涉及执行这些操作的设备，该设备可以为所需目的而得到，或包括由存贮在计算机里的计算机程序选择性激活或重组的通用计算机。这种计算机程序可以存入计算机可读存储媒体，诸如但不限于任一类型的盘片（包括软盘、光盘、CD-ROM 与磁光盘）、只读存储器 (ROM)、随机存取存储器 (RAM)、可擦可编程只读存储器 (EPRPM)、电可擦可编程只读存储器 (EEPROM)、磁卡或光卡，或任一种适合存贮电子指令的媒体，每一种都耦接计算机系统总线。指令可用一个或多个处理设备（如中央处理单元等）来执行。在其它实施例中，本发明诸步骤可用包含可重配或硬线连接的用于步骤执行的逻辑电路的专用硬件元件或者任意组合的编程计算机元件与定制硬件元件来执行。

[0020] 另外，设计要经历从创造、模拟到制造的各个阶段。代表设计的数据可以若干方式表示该设计。首先在模拟当中，适宜用硬件描述语言或另一种功能描述语言表示硬件。另在设计过程的某些阶段，可以形成带逻辑门和 / 或晶体管门电路的电路级模型。再者，在某一阶段，大多数设计达到数据代表各种设备在硬件模型中物理位置的程度。在应用常规半导体制造技术的情况下，代表硬件模型的数据可以是对于用来生产集成电路的掩模规定各种特征是否存在于不同的掩蔽层上的数据。在任一设计表示法中，数据可以任一机器可读媒体形式来存储。为发送此类信息的调制或以其他方式生成的光波或电波、存储器或者盘片等磁或光存储器，可以是机器可读媒体，这类媒体都可“携带”或“指示“设计或软件信息。当发送了指示或携带代码或设计的电气载波时，对该电信号执行复制、缓冲或再传输，新的

拷贝就制成了。这样，通信提供商或网络提供商就可复制实施本发明技术的物件（载波）。

[0021] 图 1 示出可在其中运用本发明的虚拟机环璋 100 的一实施例。图 1 的 VM 环境和包括硬件、软件或二者组合的处理逻辑电路，都可实现本发明诸不同的实施例。在本例中，裸平台硬件 116 包括能例如操作标准操作系统 (OS) 或 VMM112 的计算平台。

[0022] VMM 112 虽然一般以软件构成，但可向更高级软件送出裸机接口。此种更高级软件包括标准或实时的 OS，虽然本发明在这方面不限于该范围。另外，例如 VMM 可在另一 VMM 内或在另一 VMM 上运行。VMM 及其一般特征与功能已为本领域技术人员熟悉，如可用软件、固件或各种技术的组合来实现。

[0023] 平台硬件 116 可以是个人计算机 (PC)、主机、手持设备、便携计算机、机顶盒或任一其它计算系统。平台硬件 116 包括处理器 118、存储器 120 与特许级监视逻辑电路 124。

[0024] 处理器 118 可以是任一种能执行软件的处理器，诸如微处理器、数字信号处理器、微控制器等。处理器 118 包括执行本发明方法实施例的微代码、可重编逻辑、可编程逻辑或硬代码逻辑。虽然图 1 只示出一台这种处理器 118，但系统内可以有一台或多台处理器。

[0025] 存储器 120 可以是硬盘、软盘、随机存取存储器 (RAM)、只读存储器 (ROM)、闪速存储器、以上器件的任意组合或者其它任一种处理器 118 可读的机器媒体。存储器 120 可存贮执行本发明诸方法实施例的指令和 / 或数据。

[0026] VMM112 向其它软件（如“客户”软件）提供一台或多台虚拟机 (VM) 的概要，而虚拟机向各种客户提供同样或不同的概要。系统内有一台或多台 VM，如图 1 示出两台 VM，即 102 和 114。在每台 VM 上运行的客户软件包括诸如 104 或 106 等客户 OS 和各种客户软件应用程序 108 与 110。在 VM 102 和 114 内运行的客户软件希望在客户软件在其上正在运行的 VM 102 和 114 内访问物理资源（如处理器寄存器存储器和 I/O 设备），并处理各种事件，包括系统设备产生的中断、异常、等等。在虚拟机环境中，为保持 VM 102 和 114 的正常操作和起到保护作用，VMM 112 应能最终控制物理资源，为此在必要时可拦截客户软件对计算机物理资源的访问。

[0027] 被客户软件访问的资源分为“特许”或“非特许”。对特许的资源，VMM112 可接通客户软件所需的功能，同时保持对这些特许资源的最终控制权。非特许资源不必受 VMM 112 控制，可被客户软件访问。

[0028] 另外，各客户 OS 104 和 106 都想处理各种故障事件，诸如异常（如页面故障、一般保护故障、陷阱、失灵等）、中断（如硬件中断、软件中断）和平台事件（如初始化 (INIT) 与系统管理中断 (SMI)）。其中有些故障事件是“特许的”，因为这些事件必须由 VMM112 处理以保证来自 VM102 与 114 以及 VM102 与 114 之间的正常操作与保护。

[0029] 在出现特许的故障事件或客户软件试图访问特许资源时，可将控制转到 VMM112。这里把控制从客户软件转到 VMM112 称为 VM 退出。在 VM 退出后接收控制之后，VMM112 可作各种处理，之后再把控制还给客户软件。这里把控制从 VMM 转给客户软件称为 VM 进入。

[0030] 为调用特许的 OS 功能，应用程序使用了专用指令，诸如 INT 或 SYSTENTER。调用操作前，通常把系统调用标识符和参数装入处理器的寄存器。拦截系统调用考虑到多种应用场景，例如通过跟踪系统调用，侵入检测系统可找出安全破坏的原因。另一个有用场景是通过根据系统调用启用的临时模式来检测执行周期，导出预定的定期实时任务模式。

[0031] 在一实施例中，处理器 118 按存储在 VM 控制结构 (VMCS)122 里的数据来控制

VM102 和 114 的操作。在一实施例中，VMCS122 被存入存储器 120。在另一实施例中，VMCS122 被存入处理器 118。在有些实施例中，用多个 VMCS 结构支持多个 VM。

[0032] VMCS122 是一种含有客户软件状态，VMM112 状态、执行控制信息和其它信息的结构，其中执行控制信息指示 VMM112 是如何希望限制或控制客户软件操作的。VM 中对客户软件的移入和移出以及客户软件的操作是使用一组存贮在 VMCS 里的 VM 控制点来控制的。执行控制点规定了必须将控制从客户软件转到 VMM 的情况。退出控制点控制客户状态的保留和 VMM 状态在 VM 退出时的加载。在一组 VM 退出信息数据段内设置了描述最近 VM 退出的信息。VM 退出时，将处理器状态中由客户软件使用的部分保留至 VMCS122，并从 VMCS122 加载 VMM112 所要求的处理器状态部分。进入控制点控制 VMM 状态的保留，并在 VM 进入时加载客户状态。在 VM 进入时，利用存贮在 VMCS122 里的数据来恢复客户状态，并将控制还给客户软件。

[0033] 在一实施例中，处理器 118 包括特许级监视逻辑 (PLML)124，负责评估客户的当前特许级，以根据 VMM112 规定的特许级退出判据判断是否应产生 VM 退出。在一实施例中，该特许级退出判据存储在 VMCS122 中。下面描述所述特许级退出判据的特定实施例。若 PLML124 断定当前客户特许级符合特许级退出判据，处理逻辑就促使 VM 从客户软件退到 VMM112。

[0034] 在各种 ISA 中，对特许级可指定一数值，较高数值表明增高的特许，较低数值表明减低的特许。在其它实施例中，较高数值可以指示减低的特许。在本文讨论中，特许级“增高”表示客户软件正变得更富特许，“减低”表示正变得更少特许，与涉及的数值无关。同样地，“小于”另一特许级的特许级是减低了特许的特许级，与涉及的数值无关。

[0035] 诸控制点被加到 VM 控制点以表示特许级退出判据。在一实施例中，特许级增高退出控制若置成使能值，表示在客户软件操作期间发生特增高许级时，将产生 VM 退出。这一控制点判断增高特许级的事件传送或指令执行是否造成 VM 退出。在一实施例中，在完成了使特许级变化的指令执行之后（即撤退后），因发生特许级增高而造成的 VM 退出。在一实施例中，作为 VM 退出信息一部分而向 VMM 报告的客户指令指针值，可以指向准备以新的特许级执行的第一指令。在一实施例中，没有做出明确的规定要报告造成特许级变化的指令地址。在另一实施例中，特许级减低退出控制若置成使能值，就表明在客户软件操作期间发生特许级减低时，应产生 VM 退出。该控制点判断减低特许级的事件或指令是否造成 VM 退出。在一实施例中，在造成特许级变化的指令执行后（即撤退后），因发生特许级减低而发生 VM 退出。作为 VM 退出信息一部分而报告给 VMM 的客户指令指针值，可以指向准备以新的特许级执行的第一指令。在一实施例中，没有做出明确的规定要造成特许级变化的指令的地址。

[0036] 图 2 的流程图示出了支持特许级增高与特许级减低控制点的方法的一实施例。图 2 示出在客户特许级变换时对特许级增高和特许级减低退出控制点所作的试验。图 2 中，特许级变化造成的 VM 退出比其它 VM 退出源更优先，这是 为了简化图示，但实际上，有些 VM 退出源的优先权可以更高，而有些则更低。在图 2 的实施例中，虽然示出了特许级增高和特许级减低两个控制点，但在不同的实施例中，这些控制点可被独立支持或结合其它特许级退出控制点被支持。

[0037] 图 2 中，在收到来自 VMM112 的 VM 进入请求时（框 210），过程即开始（框 205）。执

行 VM 进入检查，诸如装载系统状态等（框 215）。执行客户指令（框 220）。若指令的执行使 VM 退出（框 225），则 VM 退出使控制转到 VMM112，并向其报告 VM 退出的原因（框 230），结束该过程（框 235）。若在指令的执行时不发生 VM 退出（框 225），则撤退客户指令（框 240）。在评估了客户软件的当前特许级后，若客户软件的特许级被判定为增高了（框 245），而且特许级增高 (PLI) 控制点被使能（框 250），则 VM 退出使控制转到 VMM112，并 VMM112 报告 VM 退出是由 PLI 引起的（框 230）。若客户软件的特许级被判定为增高了（框 245），且特许级增高 (PLI) 控制点不被使能（框 250），则处理逻辑执行下一客户指令（框 220）。

[0038] 若客户软件的特许级未被定为增高（框 245），则处理逻辑检查特许级减低 (PLD)。在评估了客户软件的当前特许级后，若客户软件的特许级被判定为减低了（框 245），且 PLD 控制点被使能（框 250），则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出由 PLD 造成（框 230）。若客户软件的特许级未被定为减低（框 255）或 PLD 控制点不被使能（框 260），就执行下一客户指令（框 220）。

[0039] 特许级退出判据可被使能与禁止，以便作选择性退出，这尤其有利于多处理器调用，因为通常只是偶尔要求特许级变化造成的 VM 退出。另外，例如 VMM 可能只对从 OS 核心转到 VM 中用户级代码起作用。

[0040] 在另一实施例中，上述特许级增高和特许级减低执行控制点提供的功能可以组合成单一的特许级变化退出控制，规定任何特许级变化都会造成 VM 退出，如图 3 所示。图 3 中，过程在从 VMM112 收到 VM 进入请求时（框 310）开始（框 305）。执行 VM 进入检查，诸如装载系统状态等（框 315）。执行客户指令（框 320）。若指令的执行使 VM 退出（框 325），则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出的原因（框 330），结束处理（框 335）。若在执行指令时不发生 VM 退出（框 325），则撤退客户指令（框 340）。评估了客户软件的当前特许级后，若断定客户软件的特许级变了（框 345），且特许级变化 (PLC) 退出控制点被使能（框 350），则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出由 PLC 造成（框 330）。若 PLC 条件不被满足（框 345）或 PLC 退出控制 VMM112 未使能（框 350），则执行下一客户指令（框 320）。

[0041] 图 4 在流程图中示出本发明一实施例的检测特许级变化的一种方法。图中，通过将代表老特许级 (OPL) 的一变量置于当前特许级 (CPL) 而使之初始化（框 410），过程便开始（框 405）。处理或执行单一指令或单一异步事件诸如中断（框 415）。在指令执行或事件处理后（框 415），若特许级与指令或事件处理前不同（框 420），则特许级变了，采取某一动作（框 425），如图 2 和 3 所示。否则，执行下一指令或过程事件（框 415）。

[0042] 注意，在图 2 和 3 中，在执行指令后但撤退指令前，作出评估，判断是否出现 VM 退出。实际上，这一评估作为试图执行指令的一部分。例如，访问一控制寄存器（如 CR0）会导致具有错误语义的 VM 退出。换言之，在任一体系结构状态被指令修改前都会出现 VM 退出。其它 VM 退出可在修改了某些体系结构状态之后但在撤退指令（诸如使任务切换接着使 VM 退出的指令）之前发生。有些 VM 退出可在撤退了指令之后发生。特许级评估可以是这样一种情形：在撤退后被评估。其它 VM 退出可因其它原因在其它情况下发生。

[0043] 图 5 的流程图示出的方法识别同步事件（诸如执行特许级变化指令）以及异步事件（在执行客户软件时中断到来）。图 5 中，过程在从 VMM112 收到 VM 进入请求（框 510）时开始（框 505）。进行 VM 进入检查，如装载系统状态等（框 515）。若异步事件已定（框

520)，就执行客户指令（框 525)。若指令（框 525) 使 VM 退出（框 530)，则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出的原因（框 535)，结束处理（框 540)。若指令的执行不使 VM 退出（框 530)，则撤退客户指令（框 545)。若特许级增高 (PLI) 条件被满足（框 550) 且 PLI 控制点被使能（框 555)，则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出由 PLI 造成（框 535)。若 PLI 条件未满足（框 550) 或 PLI 控制点未被使能（框 555)，则处理逻辑检查特许级减低 (PLD)。若 PLD 条件满足（框 560) 且 PLD 控制点被使能（框 565)，则 VM 退出使控制转到 VMM112，并且 VMM112 报告 VM 退出由 PLD 造成（框 535)。若 PLD 条件不满足（框 560) 或 PLD 控制点未被使能（框 565)，则处理逻辑检查异步事件在该点是否悬而未决（框 520)。若异常事件悬而未决（框 520) 且该事件造成 VM 退出（框 570)，则 VM 退出使控制转到 VMM112，并向 VMM112 报告 VM 退出由该事件造成（框 535)。若异步事件悬而未决（框 520) 且未造成 VM 退出（框 570)，则把该事件引入客户软件（框 575)，再作上述的特许级检查。

[0044] 在另一实施例中，一执行控制指示特许级目标值，特许级目标值在 VMM 控制下被存入 VMCS，VMM 可将特许级目标值定为任一有效特许级。在另一实施例中，该特许级目标值有一固定值。有些实施例可提供一个以上这样的特许级目标值，例如，一个实施例有两个不同的特许级目标值，各自固定为专用值。

[0045] 在一实施例中，特许级目标值直到使能才起作用。特许级目标值控制点可通过把 VMCS 中一匹配的特许级目标使能控制定为使能值（诸如 1) 而被使能。该控制点表明在执行客户软件时，若客户软件特许级与该特许级目标值匹配，就产生 VM 退出。执行控制可以有多个特许级目标值和多个匹配目标特许使能控制点，各自对应于一个特许级目标值控制点。

[0046] 在一实施例中，在准备以匹配的特许级执行的第一指令的执行前，可能发生因当前特许级与特许级目标值匹配而造成 VM 退出。在一实施例中，可能被作为 VM 退出信息一部分而报告给 VMM 的客户指令指针值，指向准备以匹配特许级执行的第一指令。在一实施例中，没有做出明确的规定要报告使特许级改为匹配值的指令的地址。

[0047] 图 6 是一过程的实施例的流程图，该过程用于根据虚拟机环境中客户软件的特许级从客户软件转到 VMM。在图示例中，根据客户软件的特许级和特许级目标值控制 (PLTVC) 的值，发生 VM 退出。图 6 中，在从 VMM112 接收 VM 进入请求时（框 620)，过程开始（框 610)。执行 VM 进入检查，诸如装载系统状态等（框 630)。若 PLTVC 不被使能（框 640)，就执行客户指令（框 650)。评估了客户软件的当前特许级 (CPL) 后，若 PLTVC 被使能（框 640)，而且 CPL 等于特许级目标值 (PLTV)（框 660)，则 VM 退出使控制转到 VMM112（框 670)，结束处理（框 680)。若 CPL 不等于 PLTV（框 660)，就执行客户指令（框 650)。

[0048] 在另一实施例中，特许级目标值控制点用特许级上限值 (ceiling value) 控制来扩展。若将该控制点定为使能值（如定为 1)，就表示在执行客户软件时，若客户软件的特许级大于该特许级目标值，就发生 VM 退出。在一实施例中，执行控制点有多个特许级目标值和相应的特许级上限值控制点。在另一实施例中，若把特许级目标值定为最高特许级，就不会转到更富特许的状态，因此不用使能控制就能有效地禁止该控制点。

[0049] 在又一实施例中，特许级目标执行控制点用特许级下限值 (floor value) 控制来扩展，表明在执行客户软件时，若客户软件的特许级小于特许级目标值，将产生 VM 退出。在

一实施例中，执行控制点有多个特许级目标值和相应的特许级下限值控制点。在另一实施例中，若把特许级目标值定为最低特许级，就不会转到较少特许的状态，因而不用使能控制就能有效地禁止该控制点。

[0050] 图 7 是一过程的实施例的流程图，该过程用于根据客户软件在虚拟机环境中的特许级从客户软件转到 VMM。在图示例中，在当前特许级大于 PLTV 时，根据客户软件的特许级和特许级上限值控制 (PLCVC) 的值发生 VM 退出。图 7 中，当从 VMM112 收到 VM 进入请求时 (框 720)，过程开始 (框 710)。

[0051] 执行 VM 进入检查，诸如装载系统状态等 (框 730)。若 PLCVC 不被使能 (框 740)，就执行客户指令 (框 750)。评估了客户软件的 CPL 后，若 PLCVC 被使能 (框 740)，而且 CPL 超出 PLTV(框 760)，则 VM 退出使控制转到 VMM112(框 770)，结束处理 (框 780)。若 CPL 不超出 PLTV(框 760)，就执行客户指令 (框 750)。

[0052] 在实践中，要用一个监视计时器，因为在运行多驱动器线程的延长时段内，操作系统有时保持在核心模式中。图 8 是一过程的实施例的流程图，在虚拟机环境中执行抢占调度。图 8 中，当处理逻辑判断当前 VM(CVM) 是否要求抢占时 (框 810)，过程就开始 (框 805)。处理逻辑判断 CVM 是否为虚拟系统的组成部分 (框 815)。若 CVM 不处于特许模式 (PM)(框 820)，则 CVM 被抢占 (框 825)，结束处理 (框 830)。若 CVM 处于 PM(框 820)，则建立监视器计时器 (WDT)(框 835)。于是在 CVM 不处于特许模式时 (框 840)，适当设置特许级退出控制 (PLEC) 以让 VM 退出。接着，处理逻辑把控制转到 CVM(框 845)。VMM112 在 VM 从 CVM 退出后接受控制 (框 850)。若 VM 退出由 WDT 造成 (框 855)，则 CVM 被抢占 (框 825)，结束处理框 (830)。若 VM 退出不是 WDT 造成 (框 855)，则处理 VM 退出 (框 860)，处理逻辑判断 CVM 是否不处于 PM(框 865)。若 CVM 不处于 PM(框 865)，则 CVM 被抢占 (框 825)，结束处理 (框 830)。若 CVM 处于 PM(框 865)，则 PLEC 置位 (框 840)，控制转到 CVM(框 845)，WDT 和 PLEC 保持有效。

[0053] 在另一实施例中，设置了控制点以在特许级之间的特定转换时 (诸如从第一级转到第四级和从第三级转到第二级) 引起 VM 退出。例如在有 4 个特许级的 ISA 中，可设置 12 个控制点让 VMM 在 12 种可能的源与目的地特许级组合中选择任意一种，使 VM 退出。在一实施例中，设置的控制点少于特许级组合的全叉积 (full cross product)。控制点在这里称为特许级源目的地退出控制点 (PLSDEC)。

[0054] 在一实施例中，对引起特许级转换的所有事件和指令都设置了执行控制点，使 VMM112 对所有此类特许级变化都重获控制。这一替代法要求在 VMM112 中附加的支持，以判断特定的指令执行是否真的造成特许级变化。例如，在一个 ISA 中，中断返回指令 (IRET) 可能引起特许级变化，但并不是对一切情况都是这样的。若 IRET 指令的执行无条件地造成 VM 退出而不管执行该指令是否会改变特许级，则 VMM 要模拟该指令或用其他方式来单步执行该指令以判断是否引起特许级变化。对 VMM 软件还有其它要求，诸如模拟该 IRET，以判断故障的指令是否造成期望的特许级变化。

[0055] 引起特许级变化的某些操作或事件，会因其它原因而造成 VM 退出。这类其它 VM 退出原因可能是优先级更高的，或在执行这种指令后在评估特许级退出判据前先被评估了。此时，VMM 可判断造成 VM 退出的事件或指令还将引起特许级变化，例如虚拟结构可让 VMM 设置执行控制以在软件 (SW) 中断指令的执行时造成 VM 退出。以用户级代码执行 SW 中断指

令，会引起特许级变化。配置执行控制点，以在特许级变化被置成使 VM 退出的同时执行 SW 中断指令时，使 VM 退出。出现这种情况时，在执行 SW 中断指令期间首先评估 SW 中断造成的 VM 退出，由此导致在特许级变化前的 VM 退出。此时，SW 中断退出条件实际上有问题并且在完成指令执行前被评估，而特许级变化实际上是陷阱并且在完成指令后被评估。在一实施例中，多个 VM 退出源其实都是陷阱，多个 VM 退出源都有问题。特许级退出判据的评估可以比任何或全部其它 VM 退出源具有更高或更低的优先级。

[0056] 以上描述为了说明起见，提出了众多特定的细节，以便透彻理解本发明。但本领域的技术人员将会明白，没有这些特性的细节也可实施本发明。

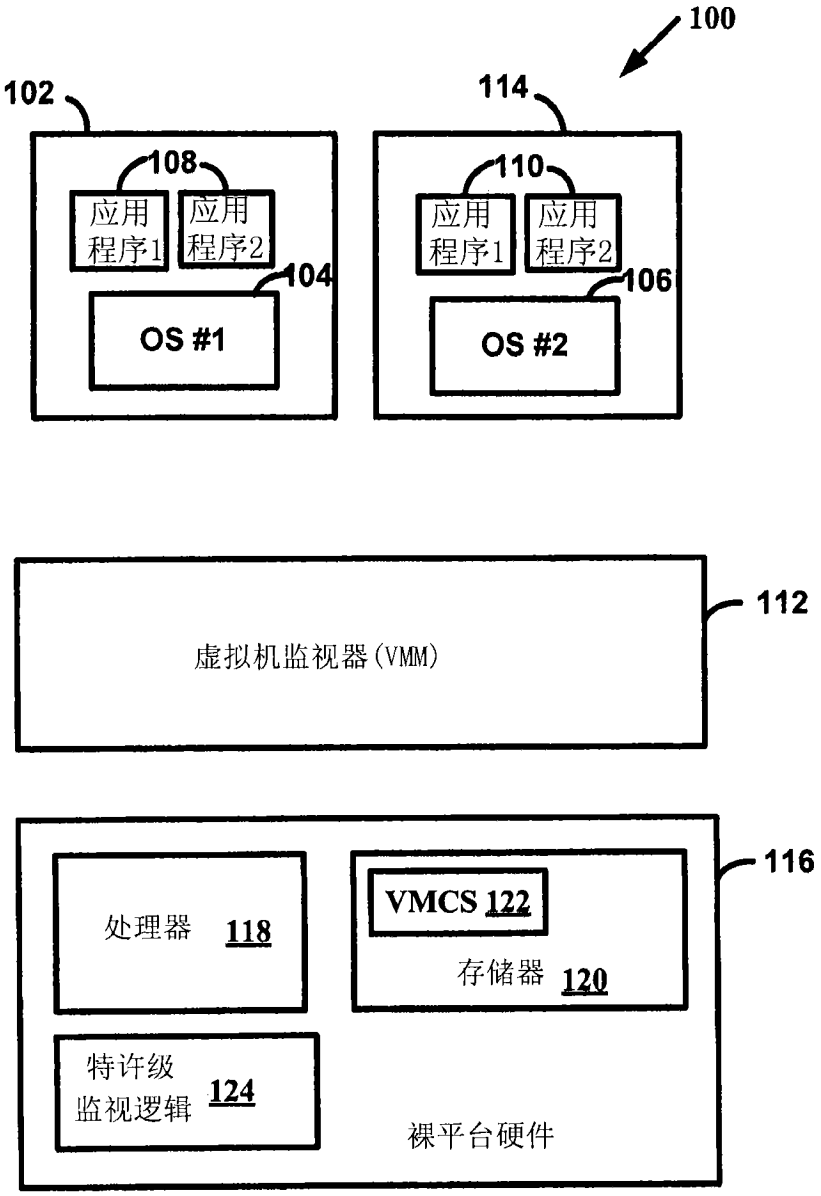

100
102
114
108
110
应用程序1
应用程序2
应用程序1
应用程序2
104
106
OS #1
OS #2
虚拟机监视器(VMM)
112
116
处理器 118
VMCS 122
存储器 120
特许级监视逻辑 124
裸平台硬件

图 1

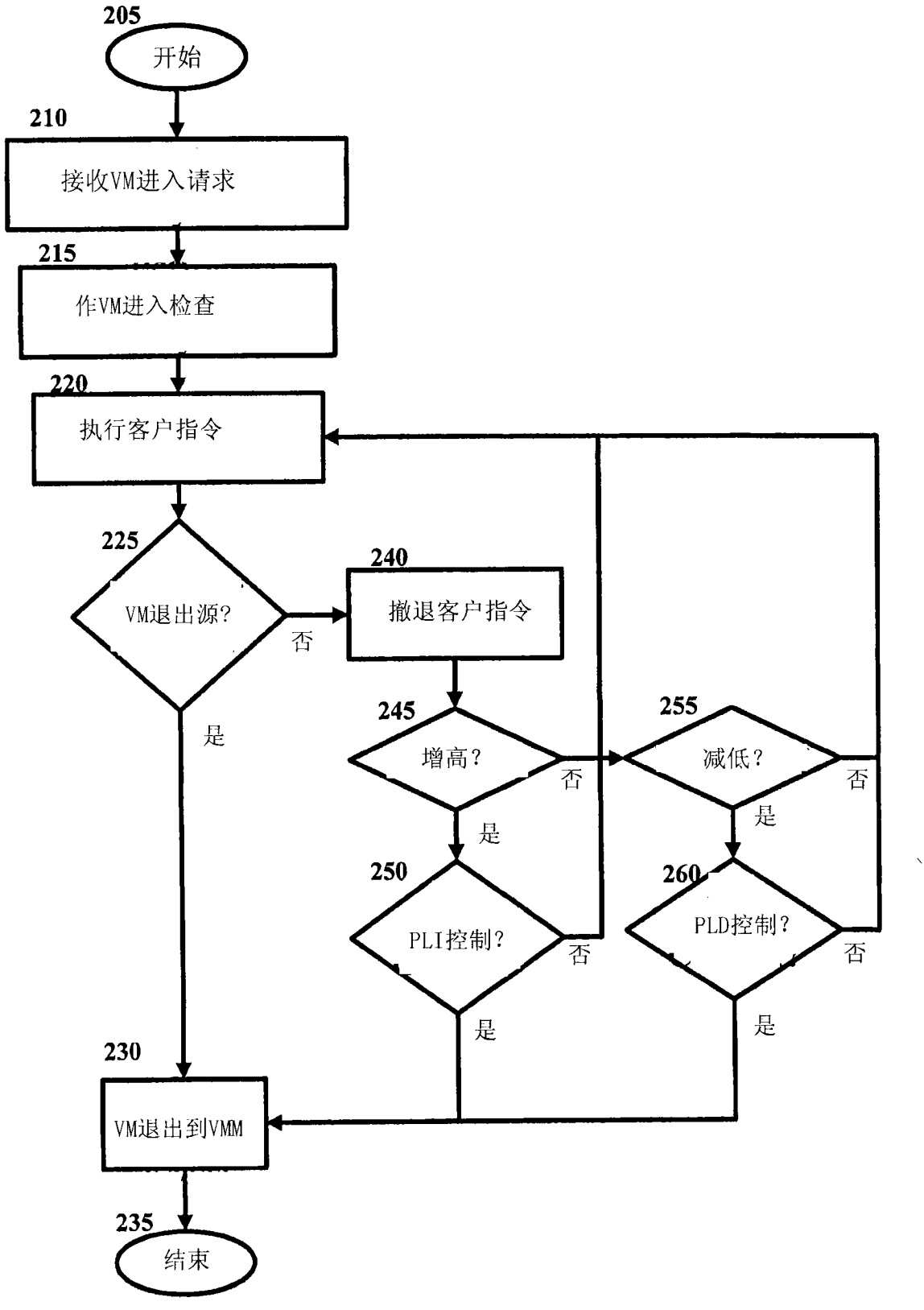

205
开始
210
接收VM进入请求
215
作VM进入检查
220
执行客户指令
225
VM退出源?
否
是
240
撤退客户指令
245
增高?
否
是
250
PLI控制?
否
是
255
减低?
否
是
260
PLD控制?
否
是
230
VM退出到VMM
235
结束

图 2

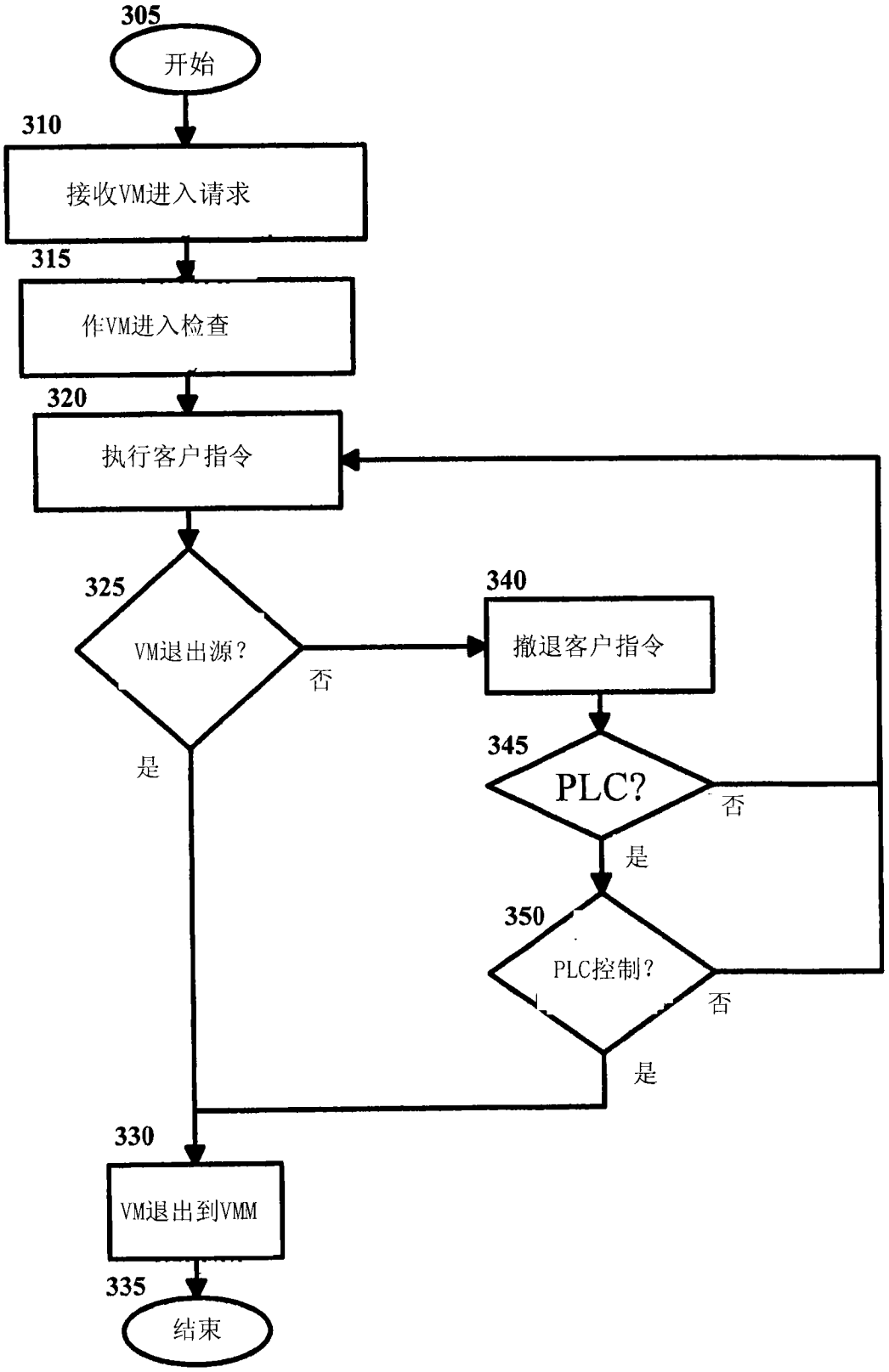

305
开始
310
接收VM进入请求
315
作VM进入检查
320
执行客户指令
325
VM退出源?
否
是
340
撤退客户指令
345
PLC?
否
是
350
PLC控制?
否
是
330
VM退出到VMM
335
结束

图 3

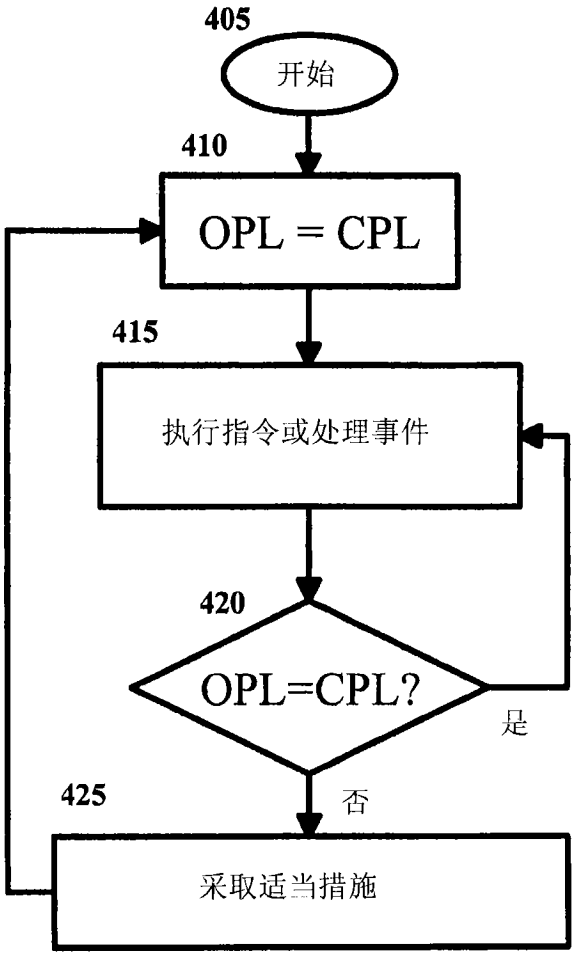


图 4

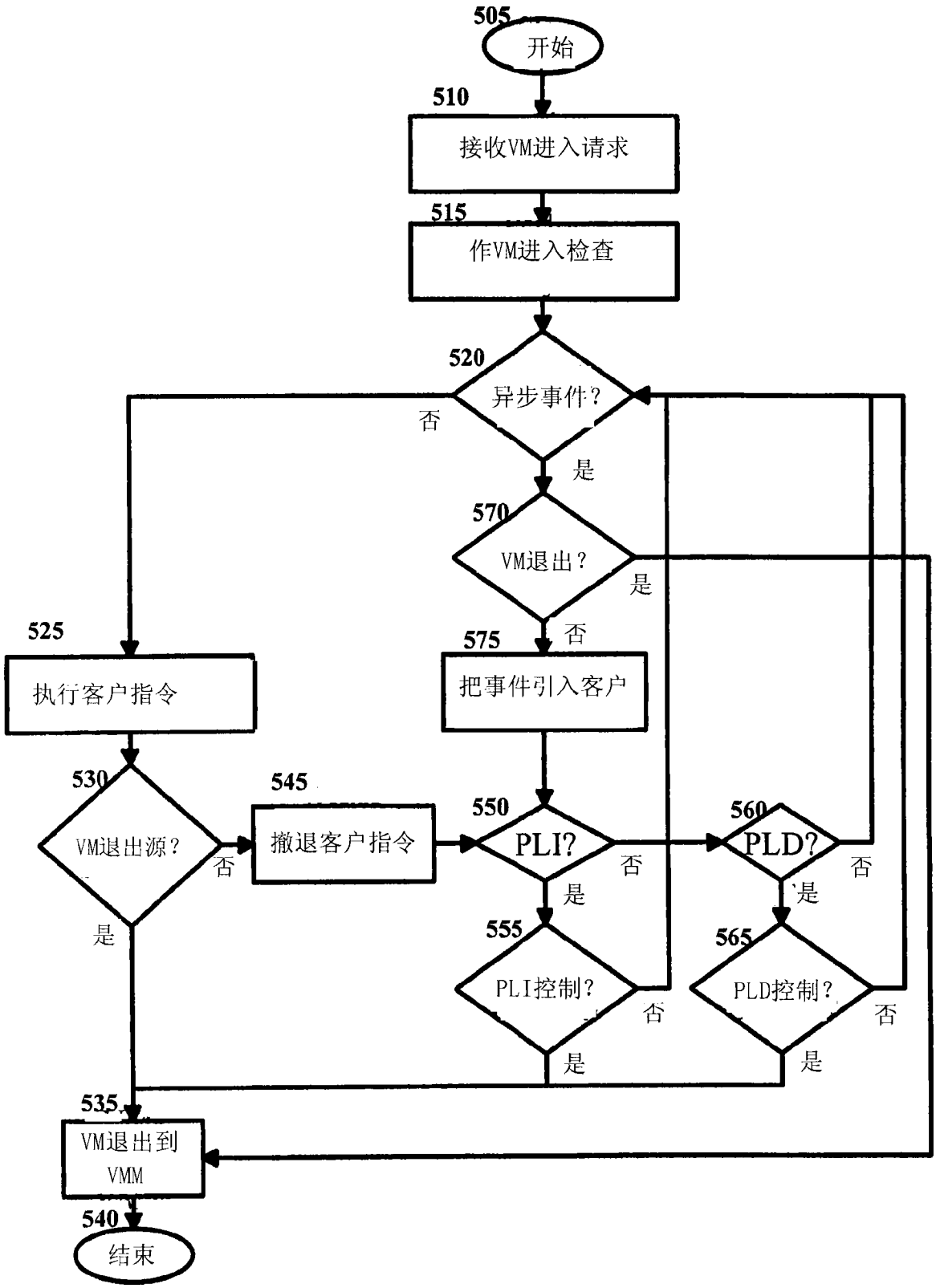

505
开始
510
接收VM进入请求
515
作VM进入检查
520
异步事件?
否
是
570
VM退出?
是
否
525
执行客户指令
575
把事件引入客户
530
VM退出源?
否
是
545
撤退客户指令
550
PLI?
否
是
560
PLD?
否
是
555
PLI控制?
否
是
565
PLD控制?
否
是
535
VM退出到VMM
540
结束

图 5

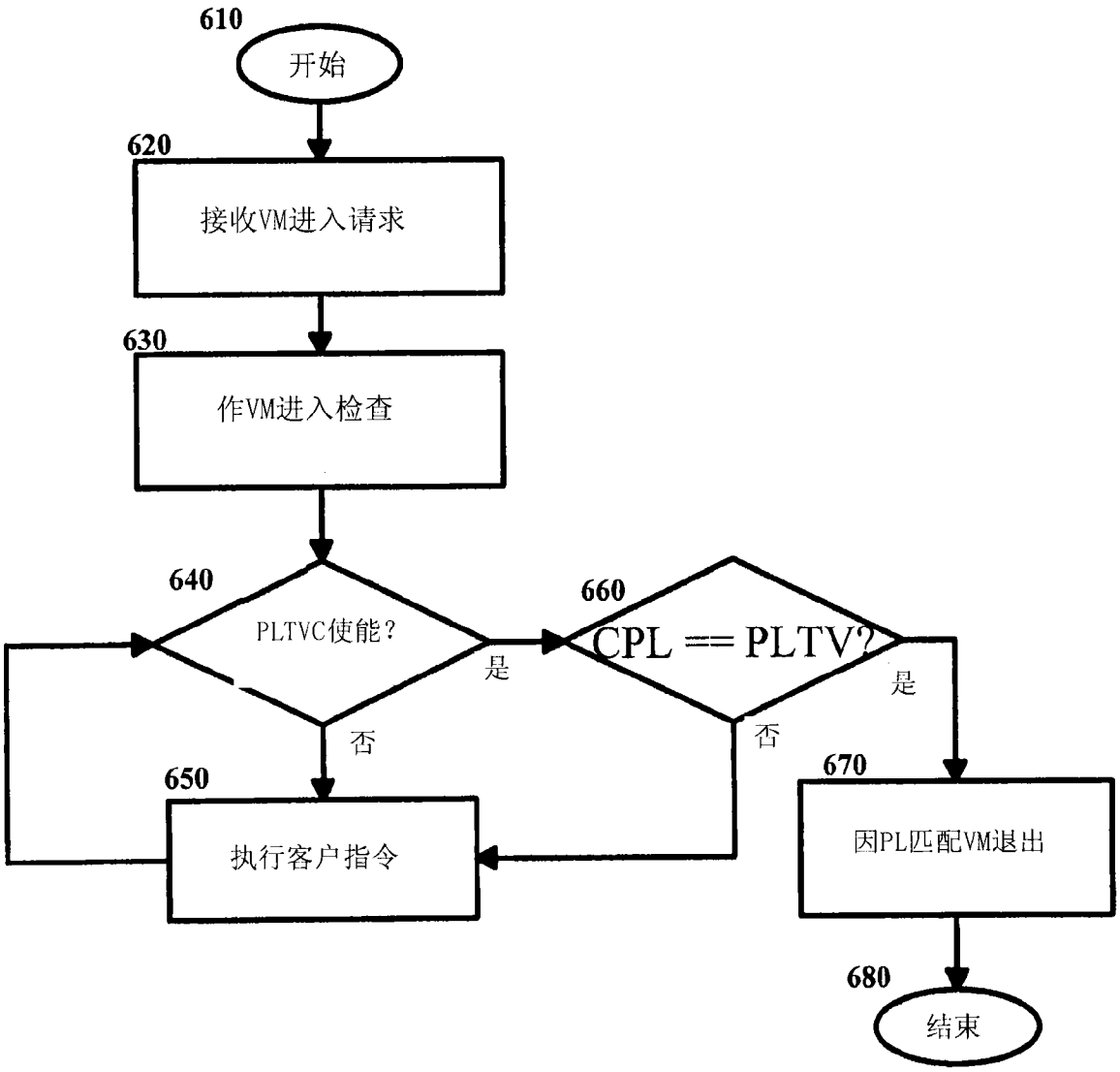

610
开始
620
接收VM进入请求
630
作VM进入检查
640
PLTVC使能?
是
否
660
CPL == PLTV?
是
否
650
执行客户指令
670
因PL匹配VM退出
680
结束

图 6

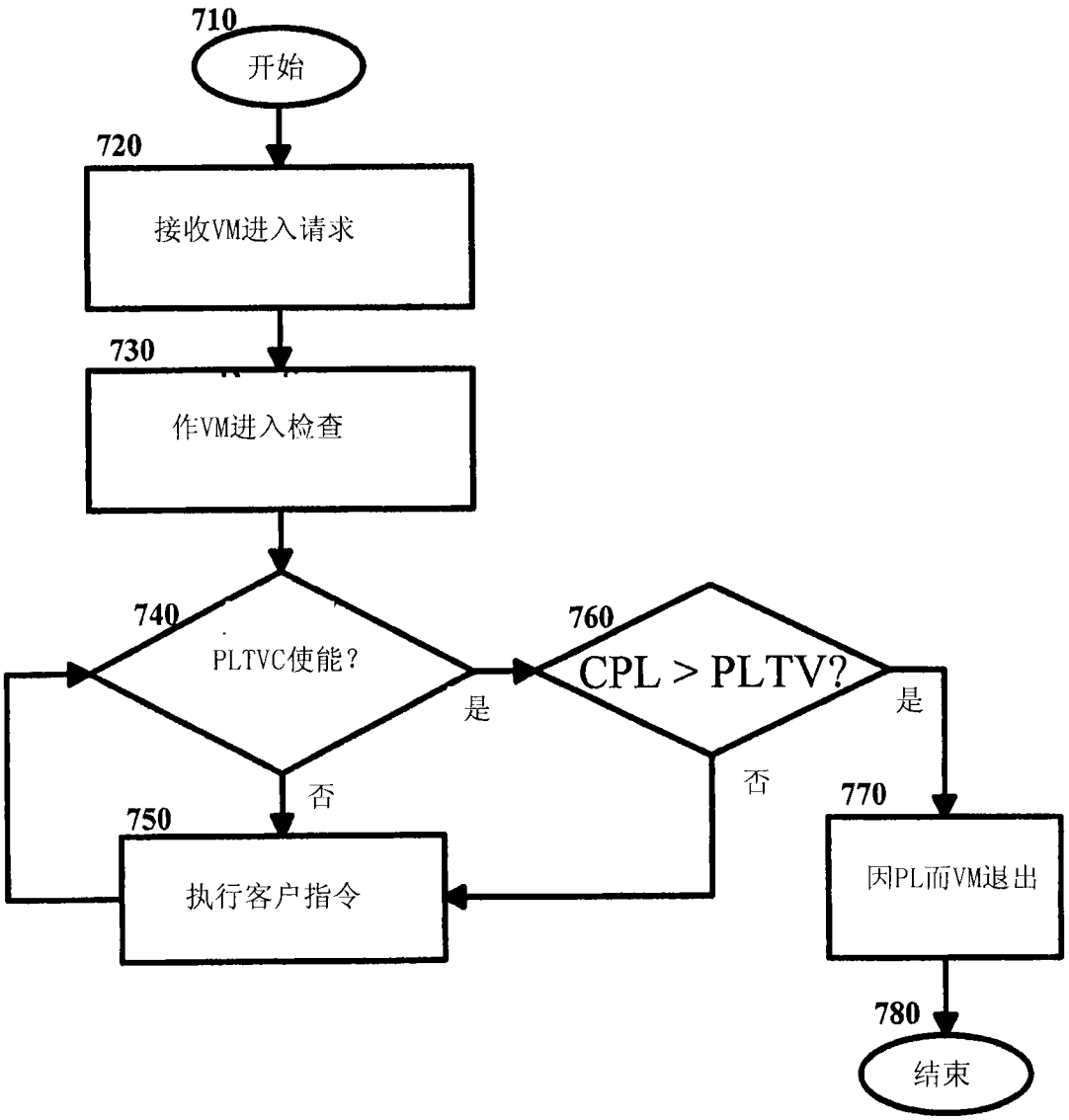

710
开始
720
接收VM进入请求
730
作VM进入检查
740
PLTVC使能?
是
否
760
CPL > PLTV?
是
否
750
执行客户指令
770
因PL而VM退出
780
结束

图 7

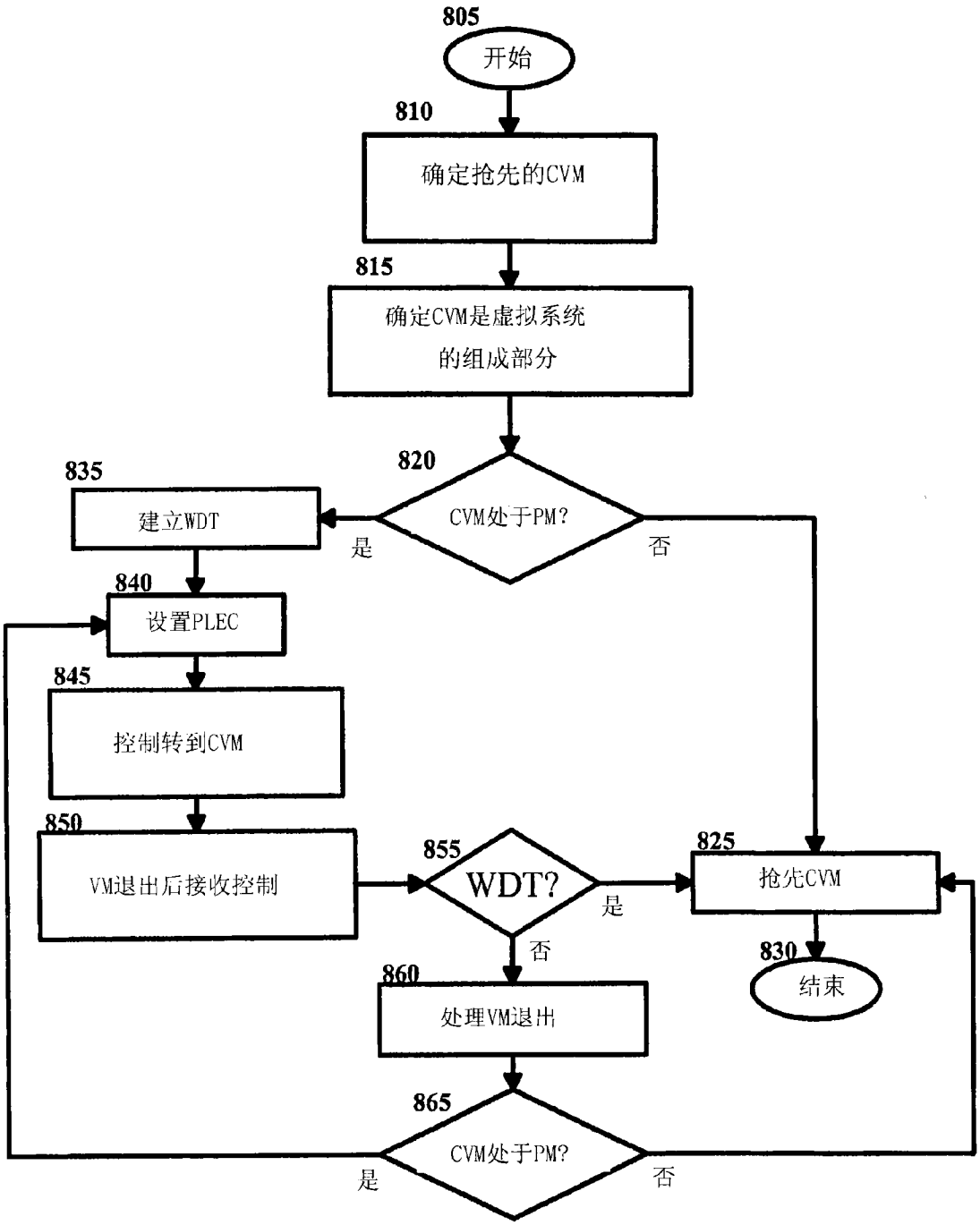

805
开始
810
确定抢先的CVM
815
确定CVM是虚拟系统的组成部分
820
CVM处于PM?
是
否
835
建立WDT
840
设置PLEC
845
控制转到CVM
850
VM退出后接收控制
855
WDT?
是
否
825
抢先CVM
830
结束
860
处理VM退出
865
CVM处于PM?
是
否

图 8