

一种缺页恢复方法、设备及存储介质

5 本申请要求于 2023 年 08 月 16 日提交中国专利局、申请号为 202311046986.1、申请名称为“一种缺页恢复方法、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

10 本申请涉及计算机技术领域，尤其涉及一种缺页恢复方法、设备及存储介质。

背景技术

IOPF（IO Page fault，IO 缺页），即 IO（Input/Output，输入/输出）设备发送的缺页请求。

15 目前，在具有多核处理器的物理机中，物理机所连接的各个 IO 设备所发起的缺页请求都被集中到指定的某个物理内核上进行处理，这不仅给该物理内核带来了巨大的工作压力，而且，还会导致缺页恢复出现排队问题，影响缺页恢复的效率。

发明内容

20 本申请的多个方面提供一种缺页恢复方法、设备及存储介质，用以改善对 IOPF 的处理效率。

本申请实施例提供一种缺页恢复方法，适用于物理机中的虚拟机管理器，所述物理机上装配有 IO 设备，所述物理机包含多个物理内核，所述方法包括：

25 在监测到所述物理机上任一 IO 设备发起缺页请求的情况下，确定所述缺页请求所属的目标虚拟机；

从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；
将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。

30 本申请实施例还提供一种计算设备，包括存储器、处理器和通信组件；

所述存储器用于存储一条或多条计算机指令；

所述处理器与所述存储器和所述通信组件耦合，用于执行所述一条或多条计算机指令，以用于执行前述的缺页恢复方法。

本申请实施例还提供一种存储计算机指令的计算机可读存储介质，当所述计算机指令被一个或多个处理器执行时，致使所述一个或多个处理器执行前述的缺页恢复方

法。

在本申请实施例中，创新性地提出在虚拟机管理器中新增对 IOPF 的调度功能。可响应于监测到物理机上任一 IO 设备发起的缺页请求，确定所述缺页请求所属的目标虚拟机；从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。这样，通过明确物理机中发生的各个 IOPF 所属的虚拟机，可以虚拟机为约束条件进行 IOPF 调度，这不仅可避免 IOPF 集中到物理机上的某个指定物理内核上处理，而且可有效规避 IOPF 调度对无关虚拟机造成扰动，从而保证对 IOPF 的合理调度，进而改善对 IOPF 的处理效率。

10

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

15

- 图 1 为本申请一示例性实施例提供的一种缺页恢复方法的流程示意图；
- 图 2a 为本申请一示例性实施例提供的一种缺页恢复方法的逻辑示意图；
- 图 2b 为本申请一示例性实施例提供的一种 IOPF 调度器的示例性结构示意图；
- 图 3 为本申请另一示例性实施例提供的一种缺页恢复方法的流程示意图；
- 图 4 为本申请一示例性实施例提供的一种可选实现方式的逻辑示意图；
- 图 5 为本申请另一示例性实施例提供的一种计算设备的结构示意图。

20

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范

25

围。

目前，物理机上发生的所有 IOPF 都集中在指定物理内核上处理，给指定物理内核带来了巨大的工作压力，而且造成 IOPF 的排队问题，影响对 IOPF 的处理效率。为此，本申请的一些实施例中：创新性地提出在虚拟机管理器中新增对 IOPF 的调度功能。可响应于监测到物理机上任一 IO 设备发起的缺页请求，确定所述缺页请求所属的目标虚拟机；从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。这样，通过明确物理机中发生的各个 IOPF 所属的虚拟机，可以虚拟机为约束条件进行 IOPF 调度，这不仅可避免 IOPF 集中到物理机上的某个指定物理内

30

35

核上处理，而且可有效规避 IOPF 调度对无关虚拟机造成扰动，从而保证对 IOPF 的合理调度，进而改善对 IOPF 的处理效率。

以下结合附图，详细说明本申请各实施例提供的技术方案。

图 1 为本申请一示例性实施例提供的一种缺页恢复方法的流程示意图。图 2a 为本申请一示例性实施例提供的一种缺页恢复方法的逻辑示意图。参考图 2a，本实施例提供的缺页恢复方法可由物理机中的虚拟机管理器（virtual Machine monitor, VMM）实施。其中，虚拟机管理器可以是一种运行在基础物理资源和操作系统之间的中间软件层，可允许多个操作系统和应用共享物理资源，也可叫做 Hypervisor。

以下先对本实施例中涉及到的几个技术概念进行简单说明。

10 IOPF（IO Page fault，IO 缺页），即 IO 设备发起的缺页请求。目前，越来越多的 IO 设备开始支持页请求接口（Page Request Interface, PRI）协议，使得 IO 设备可以像 CPU 一样，在内存访问场景中进行缺页请求（Page Fault）。这样，在云计算等场景下，可解除 IO 设备所访问的物理机内存不允许进行内存交换（Swap）的限制。后文中，将 IO 设备发起的缺页请求描述为 IOPF。

15 内存交换（Swap）技术，是将使用不频繁的内存页复制到磁盘等外部存储区域，并释放物理机内存以增加系统可用内存的技术。使用内存交换技术，可明显提高物理机内存的使用率，该技术在内存高密度型数据中心以及云计算中发挥着重要的作用。而内存交换可能导致前述的缺页问题（Page fault）。

参考图 2a，本实施例中的物理机上装配有 IO 设备，本实施例中的 IO 设备可以理解 20 为 PCI 设备/PCIe 设备，也即是，基于总线协议接入所述物理机的外接设备。本实施例中对物理机上装配的 IO 设备的类型不做限定，例如，IO 设备可以是网卡设备、GPU（graphics processing unit，图形处理器）、DPU（Data Processing Unit，数据处理器）等。而且，本实施例中的物理机可包含多个物理内核。另外，本实施例中的物理机上，还基于虚拟化技术，运行有虚拟机。前述的虚拟机管理器即可用于对物理机上运行的 25 虚拟机进行管理。

传统的缺页恢复方案中，通常是由外部中断通过异步方式触发对 IOPF 的处理工作。而且，这些中断都集中由物理机中的指定物理内核进行处理，这样，正如背景技术中介绍的那样，物理机中发生的所有 IOPF 的缺页恢复工作都将集中在该指定物理机内核上，会造成前述的 IOPF 排队问题，影响对 IOPF 的处理效率。

30 为此，本实施例提出，在虚拟机管理器中新增 IOPF 调度器，并通过新增的 IOPF 调度器对物理机中发生的 IOPF 进行调度。这与传统的默认由指定物理内核集中处理全部 IOPF 的方案相比，可有效改善对 IOPF 的处理效率。本实施例中的 IOPF 调度器可与虚拟机控制器中的其它组件交互，以实现 IOPF 调度功能。

在此基础上，参考图 1，本实施例中提出的 IOPF 调度器可执行以下的缺页恢复方法： 35

步骤 100、在监测到物理机上任一 IO 设备发起 IOPF 的情况下，确定 IOPF 所属的目标虚拟机；

步骤 101、从多个物理内核中为目标虚拟机选取用于处理 IOPF 的目标物理内核；

步骤 102、将 IOPF 调度至目标物理内核，以利用目标物理内核处理 IOPF。

5 图 2b 为本申请一示例性实施例提供的一种 IOPF 调度器的示例性结构示意图，参考图 2b，本实施例中，在提出 IOPF 调度器的基础上，还进一步提出可将 IOPF 调度器至少分为两层：硬件抽象层和内核层。其中，硬件抽象层用于对接 IO 设备所使用到的页请求接口，以将从该硬件接口发出的 IOPF 提供至内核层。

本实施例中，基于硬件抽象层，可支持在多种 IOPF 实现架构下的 IOPF 的监测，
10 为此，本实施例中，硬件抽象层中可包含用于对接不同 IOPF 实现架构下的页请求接口的驱动程序。其中，参考图 2b，两种示例性的 IOPF 实现架构可以是：IO 设备基于输入输出内存管理单元 IOMMU 中的 PRI 接口发起 IOPF；IO 设备自身内置有自定义页请求接口，并基于自身内置的自定义页请求接口发起 IOPF。呼应于这两种示例性的 IOPF 实现架构，本实施例中的硬件抽象层中至少可包含用于对接 IOMMU 中 PRI 接口
15 的驱动程序和用于对接 IO 设备内置的自定义页请求接口的驱动程序。当然，本实施例中的硬件抽象层中还可包含更多页请求接口驱动程序，以用于对接更多 IOPF 实现架构中的页请求接口，在此不做穷举。值得说明的是，图 2b 中的两种示例性的 IOPF 实现架构中，IOMMU 中的 PRI 接口是指遵循常规页请求接口协议（Page Request Interface, PRI）的页请求接口；而 IO 设备自身内置的自定义页请求接口则可遵循自定义的页请求接口协议，关于自定义的页请求接口协议在此不做展开描述，本实施例可支持各种
20 自定义的页请求接口协议。

这样，本实施例中，基于 IOPF 调度器中的硬件抽象层，可支持对接更多 IOPF 实现架构，而不再仅局限于 IOMMU 这一种，并可将从不同 IOPF 实现架构中接收到的硬件的 IOPF，转化为一致的抽象的 IOPF，以提供至内核层。且无需对 VMM 中的其它
25 组件进行改造。

基于此，参考图 1，在步骤 100 中，IOPF 调度器可监测物理机上装配的 IO 设备发起的 IOPF。这里，正如前文提及的，IOPF 可以是 IO 设备可基于自身内置的自定义页请求接口发起 IOPF，也可以是基于 IOMMU 中的 PRI 接口而发起 IOPF。参考图 1，
30 在步骤 100 中，在监测到 IOPF 的情况下，并不会直接提交至传统方案中用于缺页恢复的默认物理内核，而是按照步骤 100 中所述，会为 IOPF 确定出其所属的目标虚拟机。

值得说明的是，在传统方案中，VMM 并不关注其接收到的 IOPF 对应的虚拟机，这是由于 VMM 只需将所有的 IOPF 默认提交至指定物理内核即可，而指定物理内核也只需无差别的处理各个 IOPF，保证相应的缺页得以恢复即可，双方都无需关注 IOPF
35 的发起方。

而本实施例中，则创新性地提出为 IOPF 确定其所属的目标虚拟机，从而引入虚拟机作为 IOPF 调度的重要参考维度和约束条件，这可有效避免因本实施例引入的 IOPF 调度方案而对无关虚拟机造成扰动。其中，无关虚拟机可理解为与当次 IOPF 无关的虚拟机。

- 5 本实施例中，可采用多种实现方式为 IOPF 确定其所属的目标虚拟机。在一种可选的实现方式中：可确定 IOPF 对应的调度域 domain；基于物理机上运行的各个虚拟机各自对应的调度域列表 domain tables，查询 IOPF 所属的目标虚拟机，其中，目标虚拟机对应的调度域列表中包含 IOPF 对应的调度域。

- 10 实际应用中，为了实现设备直通，通常会为不同的虚拟机分别配置调度域列表 domain tables，以限制隶属于单个虚拟机的调度域列表下的各个 VF (Virtual Function，虚拟功能) 设备共用同一页表进行直接内存访问 (Direct Memory Access, DMA)，从而保证在不同虚拟机下用于实现设备直通访问的页表各不相同，也即是保证不同虚拟机下的设备直通访问相互隔离。其中，VF 设备是通过 IO 设备进行虚拟化而产生的虚拟设备，不同的虚拟机上可按需装配这些虚拟设备。关于调度域的背景知识在此不做更多介绍。在该实现方式中，可从 VMM 中的其它组件中获取到物理机上运行的各个虚拟机各自对应的调度域列表 domain tables，例如，用于对接 VFIO 等设备直通协议的相关组件等，在此不做限定。

- 20 在该实现方式中，一种确定 IOPF 对应的调度域的示例性方案可以是：从 IOPF 中解析出触发 IOPF 的 IO 设备对应的 IO 设备信息；基于 IO 设备标识与调度域 domain 的映射关系表中，查询 IOPF 对应的 IO 设备标识所映射到的调度域，作为 IOPF 所对应的调度域。其中，该映射关系表可从 VMM 中的其它组件中获取到，应当理解的是，单个虚拟机的调度域列表中通常包含多个调度域 domain，前文提及到的 VF 设备被划分至调度域中，VF 设备所对应的 IO 设备也将被划分至相应的调度域中，从而产生该映射关系表。而在该示例性方案中，IO 设备在基于页请求接口发起 IOPF 时，则可将自身的设备标识 (Source Identifier) 携带在 IOPF 中，例如，IO 设备标识可以是 BDF 信息，即总线 BUS、设备 Device 和函数 Func 信息，以支持通过查询该映射关系表而为 IOPF 确定出对应的调度域。

- 30 基于此，在该实现方式中，IOPF 调度器可拥有与 VMM 等同的调度域 domain 内资源管理视角，从而，可通过比对 IOPF 对应的调度域和物理机上运行的各个虚拟机各自对应的调度域列表，确定出当前的 IOPF 所属的目标虚拟机。

应当理解的是，上述实现方式中的技术细节仅是示例性的，而并不限于此。另外，本实施例中还可采用其它实现方式来为 IOPF 确定出其所属的虚拟机，而并不限于此，在此不做更多示例。

- 35 继续参考图 1，在确定出 IOPF 所述的目标虚拟机后，在步骤 101 中，可从多个物理内核中为目标虚拟机选取用于处理 IOPF 的目标物理内核。基于在步骤 100 中引入

的目标虚拟机，在步骤 101 中提出，将以 IOPF 为单位的调度问题转换为以目标虚拟机为单位的调度问题，这样，在步骤 101 中，可从物理机上目标虚拟机可占用的物理内核中，选出目标物理内核，以用于处理当前的 IOPF。其中，目标虚拟机可占用的物理内核可理解为未被其它虚拟机占用或锁定的物理内核。

5 基于此，在步骤 101 中实质是以目标虚拟机作为调度单位并可将尽量不干扰无关虚拟机的运行作为调度约束条件，以进行 IOPF 调度。这样，可更加合理地选取出目标物理内核，进而可有效规避对 IOPF 调度工作对无关虚拟机造成干扰。

举例来说，若当前的 IOPF 隶属于虚拟机 A，则在步骤 101 中可尽量避免将该 IOPF 调度至其它虚拟机所占用或锁定的物理内核上，以避免对其它虚拟机造成资源抢占而产生跨虚拟机攻击。实际应用中，若发生资源抢占很可能导致其它虚拟机出现故障，而本实施例则可有效规避这种问题。

关于步骤 101 中选取目标物理内核的实现方案有多种，将在后文实施例中进行详细说明，在此不做过多限定。

15 在选取出目标物理内核之后，参考图 1，在步骤 102 中，可将 IOPF 调度至目标物理内核，以利用目标物理内核处理 IOPF。其中，利用目标物理内核处理 IOPF，是指利用目标物理内核按照 IOPF 进行缺页恢复。本实施例中，缺页恢复过程中可包括但不限于为缺页分配空闲物理页、从外部存储设备中读取缺页数据以及将缺页数据存储至前述所分配的空闲物理页中等工作，在此不做展开详述。

20 本实施例中，可采用多种实现方式将 IOPF 调度至目标物理内核。一种可选的实现方式中：可在物理机中的指定物理内核上监测到 IOPF 对应的缺页中断时，禁止指定物理内核启动该 IOPF 对应的缺页恢复任务；将缺页恢复任务调度至目标物理内核；其中，指定物理内核为用于缺页恢复的默认内核。其中，VMM 具有感知缺页中断的功能，这是由于，中断处理程序是 VMM 注册到缺页中断上的，基于此，在该实现方式中，可从 VMM 中的相关组件中监测到 IOPF 对应的缺页中断。值得说明的是，缺页中断和前述的缺页请求（也即是 IOPF）并不是同一概念，IO 设备除了发出 IOPF 之外，还会向指定物理内核发出与 IOPF 对应的缺页中断。在该实现方式中，缺页请求可由前述的输入输出内存管理单元 IOMMU 触发，或者由 IO 设备内置的自定义页请求接口触发，当然，在其它 IOPF 实现架构中，也可由其它主体触发，在此不做限定。

30 按照传统方案，指定物理内核将接收到 IOPF，而且在接收到对应的缺页中断时，将停止其它的工作而为缺页中断启动对应的缺页恢复任务。在该实现方式中，则会禁止指定物理内核启动相应的缺页恢复任务，而是将缺页恢复任务调度至目标物理内核上。

35 在一种示例性方案中：可控制指定物理内核远程调用目标物理内核，以利用目标物理内核处理缺页恢复任务。关于物理内核之间的远程调用技术在此不做展开详述，基于物理内核之间的远程调用技术，可无需进行任务迁移，而通过发起简洁的远程调

用请求即可触发目标物理内核执行缺页恢复任务。当然，在该实现方式中，还可采用其它示例性方案实现将缺页恢复任务调度至目标物理内核。例如，将指定物理内核上的缺页恢复任务整体迁移至目标物理内核。又例如，一方面可控制指定物理内核向目标物理内核发送针对当前 IOPF 的中断；另一方面可将目标物理内核的标识信息通知给 VMM 中的内存管理单元（Memory Management, MM），以供内存管理单元将缺页请求转发至目标物理内核，进而使得目标物理内核在接收到指定物理内核发送的该中断时启动对该 IOPF 的处理。在此不做更多示例。

综上，本实施例中，创新性地提出在虚拟机管理器中新增对 IOPF 的调度功能。可响应于监测到物理机上任一 IO 设备发起的缺页请求，确定所述缺页请求所属的目标虚拟机；从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。这样，通过明确物理机中发生的各个 IOPF 所属的虚拟机，可以虚拟机为约束条件进行 IOPF 调度，这不仅可避免 IOPF 集中到物理机上的某个指定物理内核上处理，而且可有效规避 IOPF 调度对无关虚拟机造成扰动，从而保证对 IOPF 的合理调度，进而改善对 IOPF 的处理效率。

图 3 为本申请另一示例性实施例提供的一种缺页恢复方法的流程示意图。参考图 3，该方法可包括：

步骤 300、在监测到物理机上任一 IO 设备发起 IOPF 的情况下，确定 IOPF 所属的目标虚拟机；

步骤 301、基于空闲物理内核与虚拟机之间的关联使用关系，判断目标虚拟机是否关联有空闲物理内核；

步骤 302、若是，则将与目标虚拟机存在关联使用关系的空闲物理内核，作为目标物理内核；

步骤 303、将 IOPF 调度至目标物理内核，以利用目标物理内核处理 IOPF。

其中，步骤 300 和步骤 303 可参考前述实施例中关于步骤 100 和 102 的描述，在此不做重复赘述。

本实施例中，提供了一种确定目标物理内核的可选实现方式。图 4 为该可选实现方式的逻辑示意图。

参考图 4 中的 3a 环节，在该可选实现方式中：可基于空闲物理内核与虚拟机之间的关联使用关系，判断目标虚拟机是否关联有空闲物理内核；若是，则将与目标虚拟机存在关联使用关系的空闲物理内核，作为目标物理内核。其中，该关联使用关系描述了物理机中相关虚拟机各自可使用的空闲物理内核。在该可选实现方式中并不限定关联使用关系的存储结构。举例来说，该关联使用关系可采用键值对（key-value）的方式来记录相关虚拟机各自可使用的空闲物理内核，其中虚拟机的标识可作为键 key，而空闲物理内核的标识则可作为取值 value。基于此，若在该关联使用关系中查找到目

标虚拟机的标识，则可确定目标虚拟机关联有空闲物理内核，否则，可确定目标虚拟机未关联有空闲物理内核。当然，还可采用其它存储结构以实现该关联使用关系，在此不做更多示例。

在该可选实现方式中，可从 VMM 中的其它组件中获知物理机上的空闲物理内核，其中，空闲物理内核是指未承载任何虚拟机的物理内核。可知，使用空闲物理内核处理缺页请求不会对任何虚拟机的运行造成干扰。因此，在该可选实现方式中，若目标虚拟机关联有空闲物理内核，则可直接将其所关联的空闲物理内核作为目标物理内核。

在该可选实现方式中，可动态维护前述所使用到的空闲物理内核与虚拟机之间的关联使用关系。一种示例性方案中：可在物理机中的空闲物理内核中分别创建缺页恢复线程，以形成缺页恢复线程池；将缺页恢复线程池中的缺页恢复线程分配至物理机中运行的虚拟机，以产生空闲物理内核与虚拟机之间的关联使用关系；其中，虚拟机独占其所分配到的缺页恢复线程，虚拟机与其所分配到的缺页恢复线程所属的空闲物理内核存在关联使用关系。另外，该关联使用关系可存储在物理机的内存中，以供本实施例中的 IOPF 按需调用，当然，也可存储在该物理机外接的存储设备或者关联的远程存储设备，本实施例对此不做限定，本实施例支持将该关联使用关系存储在物理机可访问到并可及时更新的任何存储位置。

在该可选实现方式中，提出了如图 4 所示的缺页恢复线程池。应当理解的是，该缺页恢复线程池中的缺页恢复线程数量是动态变化的，通过将池中的缺页恢复线程分配给物理机中虚拟机，可间接建立起缺页恢复线程所属的空闲物理内核与虚拟机之间的关联使用关系。

其中，一种示例性的缺页恢复线程分配方案可以是：若缺页恢复线程池中的缺页恢复线程的数量大于或等于虚拟内核的数量，则为物理机上运行的各个虚拟机分别分配缺页恢复线程，以产生空闲物理内核与虚拟机之间的关联使用关系；若缺页恢复线程池中的缺页恢复线程的数量小于虚拟内核的数量，则为物理机上运行的部分虚拟机分别分配缺页恢复线程，以产生空闲物理内核与虚拟机之间的关联使用关系；其中，单个虚拟机可包含一个或多个虚拟内核。实际应用中，单个虚拟机可分配到一个或多个缺页恢复线程，例如，当虚拟机包含多个虚拟内核时，可为其分配多个缺页恢复线程，以应对可能发生的多个 IOPF 并行发生的问题。

在该示例性缺页恢复线程分配方案中，考虑了两种情况。一种情况下，缺页恢复线程池中的缺页恢复线程足够多，各个虚拟机均可分配到足量的缺页恢复线程，这样，每个虚拟机都将关联有空闲物理内核。另一种情况下，缺页恢复线程池中的缺页恢复线程数量不足，这势必导致部分虚拟机无法分配到缺页恢复线程，这种情况下，可基于优先级等指标维度对物理机中的虚拟机进行排序，优先为排序在前的虚拟机分配缺页恢复线程，这样，物理机上运行的部分虚拟机可关联有空闲物理内核，而剩余的虚拟机则不会关联有空闲物理内核。

另外，在该示例性方案中，可通过与 VMM 中的其它组件交互感知物理机中空闲物理内核的变化情况（例如，出现了虚拟机创建事件等），并在感知到物理机中的空闲物理内核发生数量变化的情况下，更新空闲物理内核与虚拟机之间的关联使用关系。这样，可通过低频率更新而保证该关联使用关系的准确性。

5 当然，在该实现方式中还可采用其它方案维护空闲物理内核与虚拟机之间的关联使用关系，而并不限于此。在不同方案中维护这种关联使用关系所需的相关信息可通过与 VMM 中其它组件的交互而获得。

10 进一步，参考图 3 以及图 4 中的 3b 环节，若在步骤 301 中判断出目标虚拟机未关联有空闲物理内核，则本实施例提供的缺页恢复方法还可包括步骤 304：从目标虚拟机包含的虚拟内核中选取目标虚拟内核；将目标虚拟内核所处的物理内核，作为目标物理内核。

15 一种示例性的目标虚拟内核选取方案可以是：获取目标虚拟机包含的虚拟内核各自对应的负载率；将负载率满足指定要求的虚拟内核，作为目标虚拟内核。参考图 4，实际应用中，可从虚拟化技术（例如，KVM 等技术）中使用到的虚拟机调度器中询问目标虚拟机包含的虚拟内核各自对应的负载率。进一步，在该示例性的目标虚拟内核选取方案中，若不存在负载率满足指定要求的虚拟内核，则按照轮询机制，将本轮命中的虚拟内核，作为目标虚拟内核。其中，轮询机制可采用循环赛 round robin 或者多队列接收 RSS（receive side scaling）等技术实现，在此不做限定。通过轮询机制，可将缺页请求均衡地分散到目标虚拟机包含的各个虚拟内核所处的物理内核。

20 在该示例性的目标虚拟内核选取方案中，针对负载率所预置的指定要求可包括但不限于负载率最低、负载率低于指定负载率阈值且持续时长超过指定时长阈值或者负载率低于指定负载率阈值且未被锁定更多负载等。

 举例来说，若指定要求采用的是“负载率最低，则可从目标虚拟机包含的虚拟内核中，选择负载率最低的虚拟内核，作为目标虚拟内核。

25 举例来说，若指定要求采用的是“负载率低于指定负载率阈值且持续时长超过指定时长阈值”，则可从目标虚拟机包含的虚拟内核中选取同时符合负载率和持续时长这两方面要求的虚拟机内核，作为目标虚拟内核。其中，持续时长是指负载率持续保持在指定负载率阈值之下的时长。若符合该指定要求的虚拟内核有多个，则可进一步从中选取持续时长最长的一个，作为目标虚拟内核。这样，负载率过高或者低负载率状态的持续时长不足的虚拟内核将不会被选中。

30 举例来说，若指定要求采用的是“负载率低于所述指定负载率阈值且未被锁定更多负载”，则可从目标虚拟机的虚拟内核中，选择负载率足够低且未被其它任务锁定额外负载的虚拟内核，作为目标虚拟内核。实际应用中，物理机中可能存在一些高优先级的任务会预先锁定部分虚拟内核，被锁定的虚拟内核将无法用于其它任务。这种
35 情况下，通过该指定要求，负载率虽然足够低但却已经被其它任务锁定的虚拟内核将

不会被选中。

据此，本实施例中，可充分考虑虚拟机层面的约束以及物理机中各个物理内核的实际负载情况，从而为缺页请求选取出合适的目标物理内核。这样，大部分缺页请求可被调度至空闲物理内核上处理，少数缺页请求需要被调度到非空闲物理内核上时，
5 也充分考虑了其它虚拟机的运行需求，这些缺页请求将被调度至目标虚拟机本来就需要占用的物理内核上，且从中优选出了负载率最低的物理内核，以更好地规避对选中的非空闲物理内核上其它虚拟机的干扰。

需要说明的是，在上述实施例及附图中的描述的一些流程中，包含了按照特定顺序出现的多个操作，但是应该清楚了解，这些操作可以不按照其在本文中出现的顺序来执行或并行执行，操作的序号如 801、802 等，仅仅是用于区分开各个不同的操作，
10 序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。

图 5 为本申请另一示例性实施例提供的一种计算设备的结构示意图。如图 5 所示，该计算设备包括：存储器 50 和处理器 51。该计算设备上装配有 IO 设备 52 且该计算
15 设备包含多个物理内核，其中，处理器 51 为用于运行虚拟机管理器的物理内核。

处理器 51，与存储器 50 和 IO 设备 52 耦合，用于执行存储器 50 中的计算机程序，以用于：

在监测到所述物理机上任一 IO 设备 52 发起缺页请求的情况下，确定所述缺页请求所属的目标虚拟机：

20 从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；
将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。

在一可选实施例中，处理器 51 在确定所述缺页请求所属的目标虚拟机时，具体可用于：

25 确定所述缺页请求对应的调度域 domain；

基于所述物理机上运行的各个虚拟机各自对应的调度域列表 domain tables，查询所述缺页请求所属的目标虚拟机，其中，所述目标虚拟机对应的调度域列表中包含所述缺页请求对应的调度域。

在一可选实施例中，处理器 51 在从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核时，具体可用于：

30 基于空闲物理内核与虚拟机之间的关联使用关系，判断所述目标虚拟机是否关联有空闲物理内核；

若是，则将与所述目标虚拟机存在关联使用关系的空闲物理内核，作为所述目标物理内核。

35 在一可选实施例中，处理器 51 还可用于：

若否，则从所述目标虚拟机包含的虚拟内核中选取目标虚拟内核；

将所述目标虚拟内核所处的物理内核，作为所述目标物理内核。

在一可选实施例中，处理器 51 还可用于：

5 池；
在所述物理机中的空闲物理内核中分别创建缺页恢复线程，以形成缺页恢复线程

将所述缺页恢复线程池中的缺页恢复线程分配至所述物理机中运行的虚拟机，以产生所述空闲物理内核与虚拟机之间的关联使用关系；

其中，虚拟机独占其所分配到的缺页恢复线程，虚拟机与其所分配到的缺页恢复线程所属的空闲物理内核存在关联使用关系。

10 在一可选实施例中，处理器 51 在将所述缺页恢复线程池中的缺页恢复线程分配至所述物理机中运行的虚拟机，以产生所述空闲物理内核与虚拟机之间的关联使用关系时，可具体用于：

若所述缺页恢复线程池中的缺页恢复线程的数量大于或虚拟内核的数量，则为所述物理机上运行的各个虚拟机分别分配缺页恢复线程，以产生所述空闲物理内核与虚拟机之间的关联使用关系；
15

若所述缺页恢复线程池中的缺页恢复线程的数量小于虚拟内核的数量，则为所述物理机上运行的部分虚拟机分别分配缺页恢复线程，以产生所述空闲物理内核与虚拟机之间的关联使用关系；

其中，单个虚拟机包含一个或多个虚拟内核。

20 在一可选实施例中，处理器 51 在从所述目标虚拟机包含的虚拟内核中选取目标虚拟内核时，可具体用于：

获取所述目标虚拟机包含的虚拟内核各自对应的负载率；

将负载率低于指定要求的虚拟内核，作为所述目标虚拟内核。

在一可选实施例中，处理器 51 还可用于：

25 若不存在负载率低于所述指定要求的虚拟内核，则按照轮询机制，将本轮命中的虚拟内核，作为所述目标虚拟内核。

在一可选实施例中，处理器 51 在将所述缺页请求调度至所述目标物理内核时，具体可用于：

在所述物理机中的指定物理内核上监测到所述缺页请求对应的缺页中断时，禁止
30 所述指定物理内核启动所述缺页请求对应的缺页恢复任务；

将所述缺页恢复任务调度至所述目标物理内核；

其中，所述指定物理内核为用于缺页恢复的默认内核。

在一可选实施例中，处理器 51 在将所述缺页恢复任务调度至所述目标物理内核时，具体可用于：

35 控制所述指定物理内核远程调用所述目标物理内核，以利用所述目标物理内核处

理所述缺页恢复任务。

在一可选实施例中，所述缺页请求由输入输出内存管理单元 IOMMU 触发，或者由所述 IO 设备内置的自定义页请求接口触发；

其中，所述虚拟机管理器中部署有用于调度缺页请求的硬件抽象层和内核层，所述硬件抽象层中包含用于对接所述 IOMMU 中的 PRI 接口的驱动程序和用于对接所述 IO 设备内置的自定义页请求接口的驱动程序，以将接收到的缺页请求提供至所述内核层，所述内核层用于对所述缺页请求进行调度。

进一步，如图 5 所示，该计算设备还包括：通信组件 53、电源组件 54 等其它组件。图 5 中仅示意性给出部分组件，并不意味着计算设备只包括图 5 所示组件。

值得说明的是，上述关于计算设备各实施例中的技术细节，可参考前述的方法实施例中的相关描述，为节省篇幅，在此不再赘述，但这不应造成本申请保护范围的损失。

相应地，本申请实施例还提供一种存储有计算机程序的计算机可读存储介质，计算机程序被执行时能够实现上述方法实施例中可由计算设备执行的各步骤。

上述图 5 中的存储器，用于存储计算机程序，并可被配置为存储其它各种数据以支持在计算平台上的操作。这些数据的示例包括用于在计算平台上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

上述图 5 中的通信组件，被配置为便于通信组件所在设备和其他设备之间有线或无线方式的通信。通信组件所在设备可以接入基于通信标准的无线网络，如 WiFi、2G、3G、4G/LTE、5G 等移动通信网络，或它们的组合。在一个示例性实施例中，通信组件经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，所述通信组件还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

上述图 5 中的电源组件，为电源组件所在设备的各种组件提供电力。电源组件可以包括电源管理系统，一个或多个电源，及其他与为电源组件所在设备生成、管理和分配电力相关联的组件。

本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）

上实施的计算机程序产品的形式。

本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权 利 要 求 书

1、一种缺页恢复方法，适用于物理机中的虚拟机管理器，所述物理机上装配有 IO 设备，所述物理机包含多个物理内核，所述方法包括：

5 在监测到所述物理机上任一 IO 设备发起缺页请求的情况下，确定所述缺页请求所属的目标虚拟机；

 从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核；
 将所述缺页请求调度至所述目标物理内核，以利用所述目标物理内核处理所述缺页请求。

10 2、根据权利要求 1 所述的方法，确定所述缺页请求所属的目标虚拟机，包括：
 确定所述缺页请求对应的调度域 domain；

 基于所述物理机上运行的各个虚拟机各自对应的调度域列表 domain tables，查询所述缺页请求所属的目标虚拟机，其中，所述目标虚拟机对应的调度域列表中包含所述缺页请求对应的调度域。

15 3、根据权利要求 1 所述的方法，从所述多个物理内核中为所述目标虚拟机选取用于处理缺页请求的目标物理内核，包括：

 基于空闲物理内核与虚拟机之间的关联使用关系，判断所述目标虚拟机是否关联有空闲物理内核；

 若是，则将与所述目标虚拟机存在关联使用关系的空闲物理内核，作为所述目标
20 物理内核。

 4、根据权利要求 3 所述的方法，还包括：
 若否，则从所述目标虚拟机包含的虚拟内核中选取目标虚拟内核；
 将所述目标虚拟内核所处的物理内核，作为所述目标物理内核。

 5、根据权利要求 3 或 4 任一项所述的方法，还包括：
25 在所述物理机中的空闲物理内核中分别创建缺页恢复线程，以形成缺页恢复线程池；

 将所述缺页恢复线程池中的缺页恢复线程分配至所述物理机中运行的虚拟机，以产生所述空闲物理内核与虚拟机之间的关联使用关系；

 其中，虚拟机独占其所分配到的缺页恢复线程，虚拟机与其所分配到的缺页恢复
30 线程所属的空闲物理内核存在关联使用关系。

 6、根据权利要求 5 所述的方法，将所述缺页恢复线程池中的缺页恢复线程分配至所述物理机中运行的虚拟机，以产生所述空闲物理内核与虚拟机之间的关联使用关系，包括：

 若所述缺页恢复线程池中的缺页恢复线程的数量大于或等于虚拟内核的数量，则
35 为所述物理机上运行的各个虚拟机分别分配缺页恢复线程，以产生所述空闲物理内核

与虚拟机之间的关联使用关系：

若所述缺页恢复线程池中的缺页恢复线程的数量小于虚拟内核的数量，则为所述物理机上运行的部分虚拟机分别分配缺页恢复线程，以产生所述空闲物理内核与虚拟机之间的关联使用关系；

5 其中，单个虚拟机包含一个或多个虚拟内核。

7、根据权利要求4所述的方法，从所述目标虚拟机包含的虚拟内核中选取目标虚拟内核，包括：

获取所述目标虚拟机包含的虚拟内核各自对应的负载率；

将负载率满足指定要求的虚拟内核，作为所述目标虚拟内核。

10 8、根据权利要求7所述的方法，还包括：

若不存在负载率满足所述指定要求的虚拟内核，则按照轮询机制，将本轮命中的虚拟内核，作为所述目标虚拟内核。

9、根据权利要求7所述的方法，所述指定要求包括负载率最低、负载率低于指定负载率阈值且持续时长超过指定时长阈值或者负载率低于所述指定负载率阈值且未被
15 锁定更多负载。

10、根据权利要求1所述的方法，将所述缺页请求调度至所述目标物理内核，包括：

在所述物理机中的指定物理内核上监测到所述缺页请求对应的缺页中断时，禁止所述指定物理内核启动所述缺页请求对应的缺页恢复任务；

20 将所述缺页恢复任务调度至所述目标物理内核；

其中，所述指定物理内核为用于缺页恢复的默认内核。

11、根据权利要求10所述的方法，将所述缺页恢复任务调度至所述目标物理内核，包括：

25 控制所述指定物理内核远程调用所述目标物理内核，以利用所述目标物理内核处理所述缺页恢复任务。

12、根据权利要求1所述的方法，所述缺页请求由输入输出内存管理单元 IOMMU 触发，或者由所述 IO 设备内置的自定义页请求接口触发；

其中，所述虚拟机管理器中部署有用于调度缺页请求的硬件抽象层和内核层，所述硬件抽象层中包含用于对接所述 IOMMU 中的 PRI 接口的驱动程序和用于对接所述
30 IO 设备内置的自定义页请求接口的驱动程序，以将接收到的缺页请求提供至所述内核层，所述内核层用于对所述缺页请求进行调度。

13、一种计算设备，包括存储器、处理器和通信组件；

所述存储器用于存储一条或多条计算机指令；

35 所述处理器与所述存储器和所述通信组件耦合，用于执行所述一条或多条计算机

指令，以用于执行权利要求 1-12 任一项所述的缺页恢复方法。

- 14、一种存储计算机指令的计算机可读存储介质，当所述计算机指令被一个或多个处理器执行时，致使所述一个或多个处理器执行权利要求 1-12 任一项所述的缺页恢复方法。
- 5