

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 9/50 (2006.01)

G06F 9/455 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200710119526.1

[43] 公开日 2009年1月28日

[11] 公开号 CN 101354663A

[22] 申请日 2007.7.25

[21] 申请号 200710119526.1

[71] 申请人 联想(北京)有限公司

地址 100085 北京市海淀区上地信息产业基地创业路6号

[72] 发明人 宋伟

[74] 专利代理机构 北京银龙知识产权代理有限公司

代理人 许静

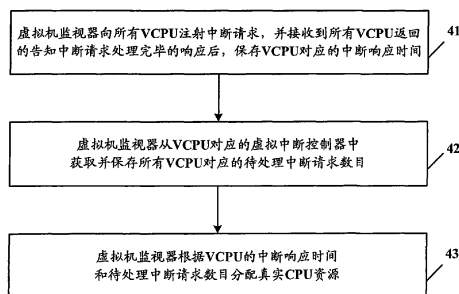
权利要求书2页 说明书9页 附图2页

[54] 发明名称

应用于虚拟机系统的真实 CPU 资源的调度方法及调度装置

[57] 摘要

本发明提供了一种应用于虚拟机系统的真实 CPU 资源的调度方法和调度装置, 其中, 该方法包括: 步骤 A, 虚拟机监视器获取虚拟 CPU 的负载情况; 步骤 B, 虚拟机监视器根据虚拟 CPU 的负载情况分配真实 CPU 资源, 负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。利用本发明的方法和装置, 可在获取虚拟 CPU 的负载情况后分配真实 CPU 资源, 使负载较重的 VCPU 能获得更多的真实 CPU 资源, 同时, 本发明直接由 VMM 实现, 与上层 GOS 的类型无关, 具有通用性。



1. 一种应用于虚拟机系统的真实 CPU 资源的调度方法，其特征在于，包括：

步骤 A，虚拟机监视器获取虚拟 CPU 的负载情况；

步骤 B，虚拟机监视器根据虚拟 CPU 的负载情况分配真实 CPU 资源，负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。

2. 根据权利要求 1 所述的方法，其特征在于，所述虚拟 CPU 的负载情况根据中断响应时间判断，所述中断响应时间越长，虚拟 CPU 负载越重。

3. 根据权利要求 1 所述的方法，其特征在于，所述虚拟 CPU 的负载情况根据待处理中断请求数目判断，所述待处理中断请求数目越多，虚拟 CPU 负载越重。

4. 根据权利要求 1 所述的方法，其特征在于，所述负载情况根据中断响应时间和中断请求数目综合判断。

5. 根据权利要求 1、2、3 或 4 所述的方法，其特征在于，所述步骤 B 中，虚拟机监视器分配给虚拟 CPU 的真实 CPU 资源的总和等于现有的真实 CPU 资源。

6. 根据权利要求 1、2、3 或 4 所述的方法，其特征在于，所述步骤 B 中，虚拟机监视器分配给虚拟 CPU 的真实 CPU 资源的比值与虚拟 CPU 的负载指数的比值相同。

7. 一种应用于虚拟机系统的真实 CPU 资源的调度装置，其特征在于，包括设置于虚拟机监视器中的负载情况获取模块和分配模块，其中：

负载情况获取模块，用于获取虚拟 CPU 的负载情况；

分配模块，用于根据虚拟 CPU 的负载情况分配真实 CPU 资源，负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。

8. 根据权利要求 7 所述的装置，其特征在于，所述负载情况根据中断响应时间和/或中断请求数目判断。

9. 根据权利要求 7 所述的装置，其特征在于，所述分配模块具体用于根据虚拟 CPU 的负载情况，将真实 CPU 资源全部分配给分配虚拟 CPU，负载越

重的虚拟 CPU 分配到的真实 CPU 资源越多。

10. 根据权利要求 7 所述的装置，其特征在于，所述分配模块具体用于根据虚拟 CPU 的负载指数的比值分配真实 CPU 资源。

应用于虚拟机系统的真实 CPU 资源的调度方法及调度装置

技术领域

本发明涉及虚拟机系统，特别是一种应用于虚拟机系统的真实 CPU 资源的调度方法及调度装置。

背景技术

如图 1 所示，在虚拟机系统中，虚拟机监视器（Virtual Machine Monitor, VMM）之上同时运行有多个操作系统，而每个操作系统中都包括一个或多个虚拟 CPU（Virtual CPU, VCPU），因此，需要虚拟机监视器根据一定的资源调度算法，将真实的 CPU 资源分配给 VCPU，而如何分配该真实 CPU 资源也会影响到虚拟机系统的性能。

目前虚拟机的调度算法大体分两种：BVT（Borrowed Virtual Time）方法和 SEDF 方法。

BVT 方法中，VMM 根据目前的真实 CPU 资源的情况，将所有的真实 CPU 资源平均分配给上层操作系统中的 VCPU，让每一个 VCPU 占用同样多的真实 CPU 资源。

但众所周知，在虚拟机环境中，各个操作系统中的 VCPU 的负载肯定不一样，而 BVT 调度算法就让负载大的 VCPU 和负载小的 VCPU 获得几乎同样多的真实 CPU 时间，这样就可能出现分配给负载大的 VCPU 的真实 CPU 时间不够，而分配给负载小的 VCPU 的真实 CPU 时间有剩余，造成了 CPU 资源的浪费，没有最大程度的利用真实 CPU 的资源。

而 SEDF 方法中，VMM 根据上层各操作系统中最近对 CPU 资源的请求和目前的进程数，给予各操作系统“打分”，为请求资源多的操作系统更多的分配真实 CPU 资源。

SEDF 方法解决了 BVT 方法中存在的问题，可以更好的利用真实 CPU 资源，但是需要在客户操作系统（Guest Operation System, GOS）中有特定的程

序来统计这些资源，对应于不同的操作系统（如 XP、Linux、Vista），需要设计不同的程序，不具有通用性。

发明内容

本发明的目的是提供一种应用于虚拟机系统的真实 CPU 资源的调度方法及调度装置，由 VMM 实现，最大程度地利用真实 CPU 的资源。

为了实现上述目的，本发明提供了一种应用于虚拟机系统的真实 CPU 资源的调度方法，包括：

步骤 A，虚拟机监视器获取虚拟 CPU 的负载情况；

步骤 B，虚拟机监视器根据虚拟 CPU 的负载情况分配真实 CPU 资源，负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。

上述的方法，其中，所述虚拟 CPU 的负载情况根据中断响应时间判断，所述中断响应时间越长，虚拟 CPU 负载越重。

上述的方法，其中，所述虚拟 CPU 的负载情况根据待处理中断请求数目判断，所述待处理中断请求数目越多，虚拟 CPU 负载越重。

上述的方法，其中，所述负载情况根据中断响应时间和中断请求数目综合判断。

上述的方法，其中，所述步骤 B 中，虚拟机监视器分配给虚拟 CPU 的真实 CPU 资源的总和等于现有的真实 CPU 资源。

上述的方法，其中，所述步骤 B 中，虚拟机监视器分配给虚拟 CPU 的真实 CPU 资源的比值与虚拟 CPU 的负载指数的比值相同。

为了更好的实现上述目的，本发明还提供了一种应用于虚拟机系统的真实 CPU 资源的调度装置，包括设置于虚拟机监视器中的负载情况获取模块和分配模块，其中：

负载情况获取模块，用于获取虚拟 CPU 的负载情况；

分配模块，用于根据虚拟 CPU 的负载情况分配真实 CPU 资源，负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。

上述的装置，其中，所述负载情况根据中断响应时间和/或中断请求数目判断。

上述的装置，其中，所述分配模块具体用于根据虚拟 CPU 的负载情况，将真实 CPU 资源全部分配给分配虚拟 CPU，负载越重的虚拟 CPU 分配到的真实 CPU 资源越多。

上述的装置，其中，所述分配模块具体用于根据虚拟 CPU 的负载指数的比值分配真实 CPU 资源。

利用本发明的方法和装置，可在获取虚拟 CPU 的负载情况后分配真实 CPU 资源，使负载较重的 VCPU 能获得更多的真实 CPU 资源，同时，本发明直接由 VMM 实现，与上层的 GOS 的类型无关，具有通用性。

附图说明

- 图 1 为虚拟机系统的简要结构示意图；
- 图 2 为本发明的第一实施例的方法的流程示意图；
- 图 3 为本发明的第二实施例的方法的流程示意图；
- 图 4 为本发明的第三实施例的方法的流程示意图。

具体实施方式

本发明的应用于虚拟机系统的真实 CPU 资源的调度方法和调度装置，直接通过虚拟机监视器来实现，其首先获取每个 VCPU 的负载情况，然后根据该 VCPU 的负载情况将真实 CPU 资源尽可能多的分配给负载较重的 VCPU。该 VCPU 的负载情况可通过各种参数来判断，下面以不同的情况对本发明的方法和装置进行详细说明。

<第一实施例>

在本发明的第一实施例中，该负载情况具体根据中断响应时间判断。

所谓中断响应时间即：VMM 向 VCPU 发出中断到 VMM 接收到告知中断处理完毕的响应之间的时间。

很明显，该中断响应时间越长，表明对应的 VCPU 负载越重。

如图 2 所示，本发明的第一实施例的方法包括：

中断响应时间获取步骤 21，虚拟机监视器向所有 VCPU 注射中断请求，并接收到所有 VCPU 返回的告知中断请求处理完毕的响应后，保存 VCPU 对

应的中断响应时间;

真实 CPU 资源分配步骤 22, 根据 VCPU 的中断响应时间分配真实 CPU 资源; 中断响应时间越长的 VCPU 分到的真实 CPU 资源越多。

在步骤 21 中, 需要向 VCPU 注射中断, 该中断可以是:
时钟中断, 如 RTC、PIT 及 HPET 等;
网卡中断; 或
声卡中断等。

在步骤 22 中, 需要根据 VCPU 的中断响应时间分配真实 CPU 资源, 中断响应时间越长的 VCPU 分到的真实 CPU 资源越多。下面以实际的例子来说明如何分配。

第一种真实 CPU 资源分配方式

假设第一种真实 CPU 资源分配方式中, 现有真实 CPU 资源为 200 时间片, 而虚拟机系统中包括 3 个 VCPU: VCPU1、VCPU2、VCPU3, 其对应的中断响应时间分别为 5、3 和 2。

步骤 22 中, 则可以分别将 80 时间片、60 时间片和 40 时间片分配给 VCPU1、VCPU2、VCPU3。

第一种真实 CPU 资源分配方式只需要保证满足以下两个条件即可:

条件 1、分配给 VCPU 的真实 CPU 资源的总和小于现有真实 CPU 资源;

条件 2、中断响应时间越长的 VCPU 分到的真实 CPU 资源越多。

然而上述的分配方案在分配真实 CPU 资源时有可能出现下述情况, 如分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源分别为 50 时间片、40 时间片和 30 时间片的情况, 这样, 分配的真实 CPU 资源为 120 时间片, 有大量的真实 CPU 资源 (80 时间片) 没有被分配, 因此, 真实 CPU 资源没有得到有效的利用。

鉴于上述问题的存在, 本发明的第一实施例中进一步将条件 1 修改为: 分配给 VCPU 的真实 CPU 资源的总和等于现有的真实 CPU 资源。

这样, 在保证负载越重的 VCPU 可以分到越多真实 CPU 资源的同时, 使真实的 CPU 资源可以得到最大程度的使用。

第二种真实 CPU 资源分配方式

假设第二种真实 CPU 资源分配方式中,现有真实 CPU 资源为 200 时间片,而虚拟机系统中包括 3 个 VCPU: VCPU1、VCPU2、VCPU3,其对应的中断响应时间分别为 5、3 和 2。

在第二种真实 CPU 资源分配方式中,VCPU 所分配到的真实 CPU 资源的比值等于 VCPU 的中断响应时间的比值,当然也需要分配给 VCPU 的真实 CPU 资源的总和小于现有的真实 CPU 资源。

按照上述的第二种分配方式,分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源可以是如下情况:

...

50 时间片、30 时间片和 20 时间片;

...

60 时间片、36 时间片和 24 时间片;

75 时间片、45 时间片和 30 时间片;

...

100 时间片、60 时间片和 40 时间片。

可以看出,如果按照上述的方式来分配,肯定也会存在真实 CPU 资源利用率低的问题存在,上述的分配情况,分别有 100 时间片、...、80 时间片、...、50 时间片没有被分配。

鉴于上述问题的存在,一种更加合适的分配方式是:保证 VCPU 所分配到的真实 CPU 资源的比值等于 VCPU 的负载指数(在此,该负载指数为中断响应时间,或中断响应时间归一化处理后的数值)的比值的同时,保证分配给 VCPU 的真实 CPU 资源的总和等于现有的真实 CPU 资源,即最后分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源分别为 100 时间片、60 时间片和 40 时间片,这样,既考虑到了 VCPU 的负载,也考虑到了真实 CPU 资源的最大程度的利用。

<第二实施例>

在本发明的第二实施例中,该负载情况具体根据 VCPU 的待处理中断请求数目来确定。

很明显,待处理中断请求数目越多,表明对应的 VCPU 的负载越重。

如图3所示,本发明的第二实施例的方法包括:

待处理中断请求数目获取步骤31,虚拟机监视器从VCPU对应的虚拟中断控制器中获取并保存所有VCPU对应的待处理中断请求数目;

真实CPU资源分配步骤32,根据VCPU的待处理中断请求数目分配真实CPU资源;待处理中断请求数目越多的VCPU分到的真实CPU资源越多。

在步骤32中,需要根据VCPU的待处理中断请求数目分配真实CPU资源,待处理中断请求数目越多的VCPU分到的真实CPU资源越多。下面以实际的例子来说明如何分配。

第一种真实CPU资源分配方式

假设第一种真实CPU资源分配方式中,现有真实CPU资源为200时间片,而虚拟机系统中包括3个VCPU: VCPU1、VCPU2、VCPU3,其对应的待处理中断请求数目分别为5、3和2。

第一种真实CPU资源分配方式只需要保证满足以下两个条件即可:

条件1、分配给VCPU的真实CPU资源的总和等于现有真实CPU资源;

条件2、中断响应时间越长的VCPU分到的真实CPU资源越多。

按照上述分配方案,分配给VCPU1、VCPU2、VCPU3的真实CPU资源可以是:

...

120时间片、60时间片和20时间片;

...

140时间片、50时间片和10时间片;

...

这样在保证负载越重的VCPU可以分到越多真实CPU资源的同时,真实的CPU资源也得到了最大程度的使用。

第二种真实CPU资源分配方式

假设真实CPU资源分配方式中,现有真实CPU资源为200时间片,而虚拟机系统中包括3个VCPU: VCPU1、VCPU2、VCPU3,其对应的待处理中断请求数目分别为5、3和2。

在第二种真实CPU资源分配方式中,VCPU所分配到的真实CPU资源的

比值等于 VCPU 的负载指数（在此，该负载指数为待处理中断请求数目或其归一化得到的数值）的比值，同时，分配给 VCPU 的真实 CPU 资源的总和等于现有的真实 CPU 资源。

按照上述的第二种分配方式，分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源为：100 时间片、60 时间片和 40 时间片，既考虑到了 VCPU 的负载程度（VCPU1、VCPU2、VCPU3 分到的真实 CPU 资源的比值为 5: 3: 2），也考虑到了真实 CPU 资源的最大程度的利用（200 时间片的真实 CPU 资源都分配出去了）。

<第三实施例>

在本发明的第二实施例中，该负载情况具体根据 VCPU 的中断响应时间和待处理中断请求数目确定。

很明显，待处理中断请求数目越多，对应的 VCPU 负载越重；中断响应时间越长，对应的 VCPU 负载越重。

如图 4 所示，本发明的第三实施例的方法包括：

步骤 41，虚拟机监视器向所有 VCPU 注射中断请求，并接收到所有 VCPU 返回的告知中断请求处理完毕的响应后，保存 VCPU 对应的中断响应时间；

步骤 42，虚拟机监视器从 VCPU 对应的虚拟中断控制器中获取并保存所有 VCPU 对应的待处理中断请求数目；

步骤 43，虚拟机监视器根据 VCPU 的中断响应时间和待处理中断请求数目分配真实 CPU 资源，中断响应时间越长的 VCPU 分到的真实 CPU 资源越多，待处理中断请求数目越多的 VCPU 分到的真实 CPU 资源越多。

在步骤 43 中，需要根据 2 个因素：中断响应时间和待处理中断请求数目为 VCPU 分配真实 CPU 资源。

因此，如果将该 2 个因素整合成一个因素考虑的话，可以更加方便的实现。

下面举例说明如何整合中断响应时间和待处理中断请求数目。

假设现有真实 CPU 资源为 200 时间片，而虚拟机系统中包括 3 个 VCPU：VCPU1、VCPU2、VCPU3，其对应的待处理中断请求数目分别为 5、3 和 2，其对应的中断响应时间分别为 2、6 和 4。

对待处理中断请求数目作归一化处理，都除以 5，得到：1、3/5 和 2/5；

对中断响应时间作归一化处理，都除以 6，得到： $1/3$ 、1 和 $2/3$ ；

然后对应相加，得到负载指数： $3/4$ 、 $8/5$ 和 $16/15$ ；

最后根据负载指数（VCPU1、VCPU2、VCPU3 分别为 $3/4$ 、 $8/5$ 和 $16/15$ ）分配真实 CPU 资源，由于 $3/4 < 16/15 < 8/5$ ，因此，VCPU2 分配到的真实 CPU 资源多于 VCPU3 分配到的真实 CPU 资源，而 VCPU3 分配到的真实 CPU 资源多于 VCPU1 分配到的真实 CPU 资源。

进一步地，在步骤 43 中，保证分配给 VCPU 的真实 CPU 资源的总和等于现有真实 CPU 资源，则能保证真实 CPU 资源的最大化利用；

进一步地，分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源的比值等于 $3/4 : 8/5 : 16/15$ ，则充分考虑了 VCPU 的负载。

下面说明另外一种整合方式，如下所述：

对待处理中断请求数目作归一化处理，都除以 5，得到：1、 $3/5$ 和 $2/5$ ，并乘上权重 2 得到：2、 $6/5$ 和 $4/5$ ；

对中断响应时间作归一化处理，都除以 6，得到： $1/3$ 、1 和 $1/3$ ，并乘上权重 1 得到： $1/3$ 、2 和 $2/3$ ；

对应相加，得到负载指数： $7/3$ 、 $16/5$ 和 $22/15$ ；

最后根据负载指数（VCPU1、VCPU2、VCPU3 分别为 $7/3$ 、 $16/5$ 和 $22/15$ ）分配真实 CPU 资源，因此，VCPU2 分配到的真实 CPU 资源多于 VCPU1 分配到的真实 CPU 资源，而 VCPU1 分配到的真实 CPU 资源多于 VCPU3 分配到的真实 CPU 资源。

进一步地，充分考虑真实 CPU 资源的最大化利用的情况下，分配给 VCPU 的真实 CPU 资源的总和等于现有真实 CPU 资源；

进一步地，充分考虑 VCPU 的负载的情况下，分配给 VCPU1、VCPU2、VCPU3 的真实 CPU 资源的比值等于负载指数的比值 $7/3 : 16/5 : 22/15$ 。

本发明的应用于虚拟机系统的真实 CPU 资源的调度装置，包括：

负载情况获取模块，用于从 VCPU 获取表征 VCPU 负载情况的负载情况；

分配模块，用于根据负载情况分配真实 CPU 资源，负载越重的 VCPU 分配到的真实 CPU 资源越多。

该负载情况根据待处理中断请求数目或中断响应时间等判断；

当负载情况根据待处理中断请求数目判断时,待处理中断请求数目越多的 VCPU 分到的真实 CPU 资源越多;

当负载情况根据中断响应时间判断时,中断响应时间越长的 VCPU 分到的真实 CPU 资源越多。

以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以作出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

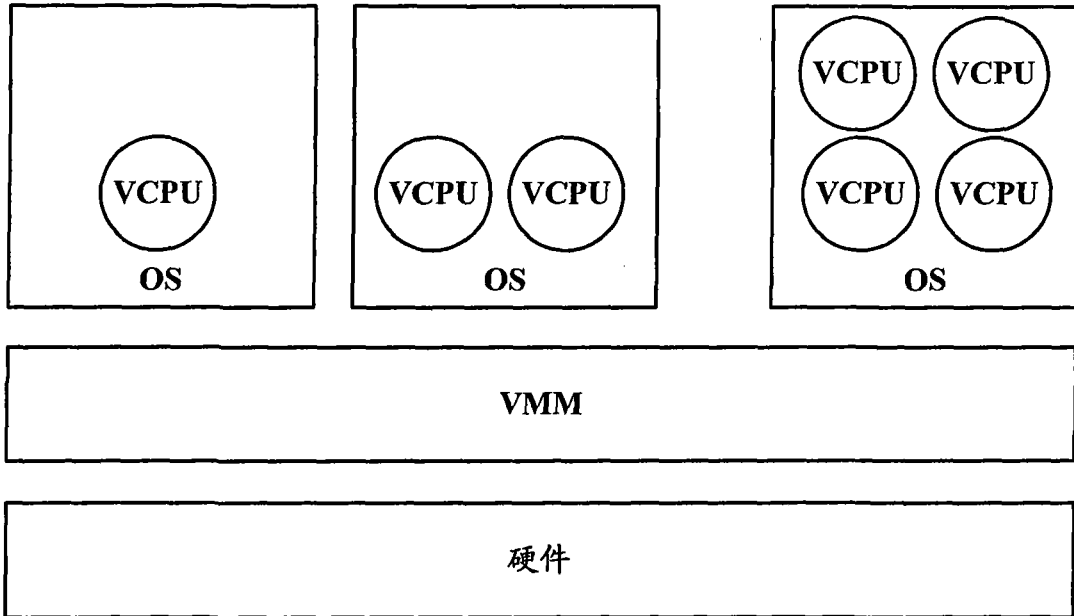


图 1

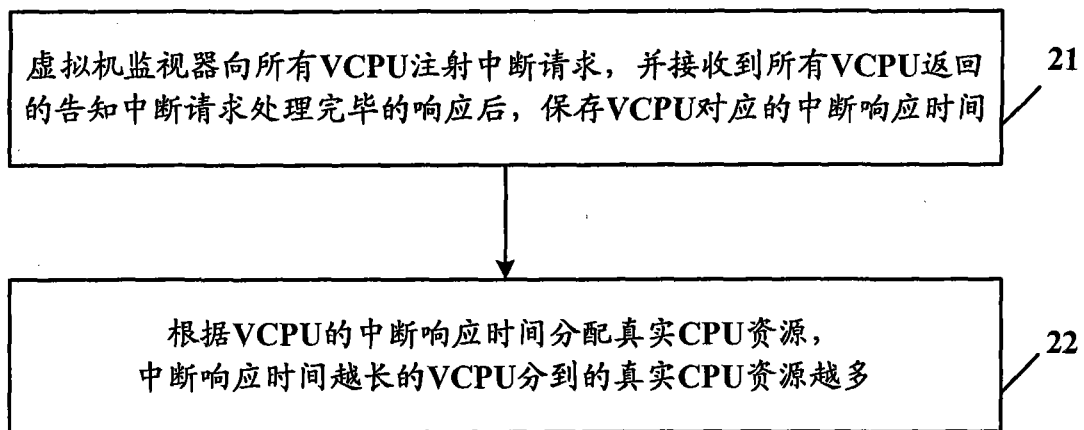


图 2

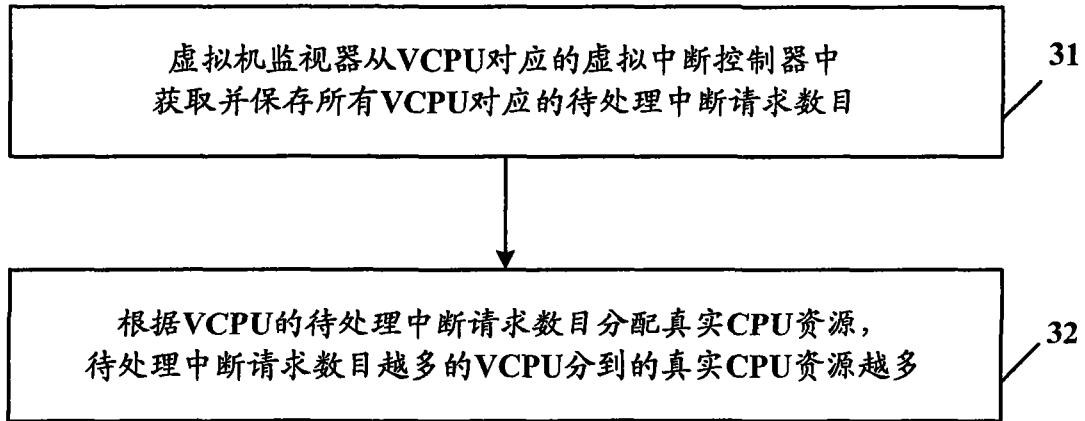


图 3

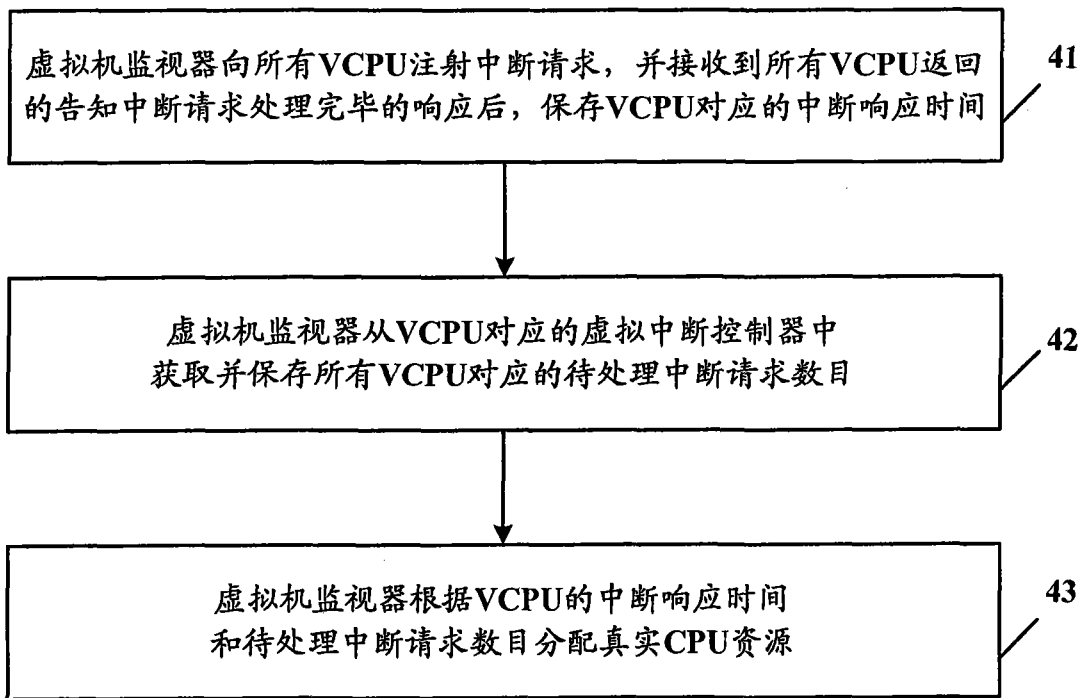


图 4