(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2019 年 7 月 4 日 (04.07.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/128158 A1

(51) 国际专利分类号:
H04L 12/26 (2006.01) *G06F 9/455* (2006.01)
H04L 29/08 (2006.01) *G06F 9/48* (2006.01)

(21) 国际申请号: PCT/CN2018/092618

(22) 国际申请日: 2018 年 6 月 25 日 (25.06.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711419235.4 2017年12月25日 (25.12.2017) CN

(71) 申请人: 上海交通大学 (SHANGHAI JIAO TONG UNIVERSITY) [CN/CN]; 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。

(72) 发明人: 李健 (LI, Jian); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 管海兵 (GUAN, Haibing); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 胡小康 (HU, Xiaokang); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 张望 (ZHANG, Wang); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。

(74) 代理人: 上海旭诚知识产权代理有限公司 (SUNRAY INTELLECTUAL PROPERTY ATTORNEYS); 中国上海市浦东新区东方路710号汤臣金融大厦1806室, 郑立, Shanghai 200122 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** INTEGRATED SCHEDULING METHOD AND DEVICE FOR INTERRUPTABILITY HOLDER PREEMPTION PROBLEM

(54) **发明名称:** 针对中断持有者抢占问题的综合调度方法及装置

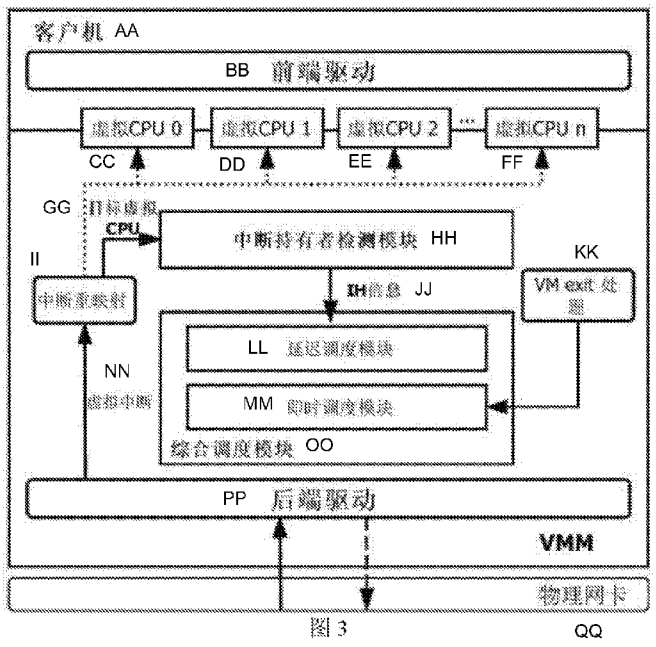


图 3

AA Client
BB Front-end drive
CC Virtual CPU 0
DD Virtual CPU 1
EE Virtual CPU 2
FF Virtual CPU n
GG Target virtual CPU
HH Interruptability holder (IH) detection module
II Interruptability remapping
JJ IH information
KK VM exit processing
LL Deferred scheduling module
MM Instant scheduling module
NN Virtual interruption
OO Integrated scheduling module
PP Rear-end drive
QQ Physical network card

(57) **Abstract:** The present invention discovers and defines a possible problem existing in an interruptability remapping mechanism under a virtual symmetric multiprocessing environment: an interruptability holder preemption (IHP) problem. The problem leads to a failure of the interruptability remapping mechanism, and thus reduces the I/O performance of a virtual machine. To solve the IHP problem, the present invention provides a proactive VCPU integrated scheduling method based on interruptability holder information. According to the method, a VCPU integrated scheduling method is constructed by globally mastering and analyzing an operation state of an interruptability holder and giving consideration to the global scheduling fairness of the system on the basis of a Kernel-Based

WO 2019/128158 A1

[见续页]

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

Virtual Machine (KVM) widely used at present and a semi-virtual network model; the IHP problem is effectively eliminated, and the I/O performance of the virtual machine is obviously improved.

(57) 摘要：本发明发现并定义了虚拟对称多处理环境下中断重映射机制可能存在的一个问题：中断持有者抢占(IHP，Interruptability Holder Preemption)问题。该问题导致中断重映射机制失效，降低了虚拟机的I/O性能。为了解决IHP问题，本发明提供一种前摄性的、基于中断持有者信息的VCPU综合调度方法。该方法基于目前广泛使用的KVM(Kernel Based Virtual Machine)及其半虚拟化网络模型，通过对中断持有者运行状态的全局把握与分析，并兼顾系统的全局调度公平性，构建一种VCPU综合调度方法，有效消除IHP问题，显著提升虚拟机的I/O性能。

针对中断持有者抢占问题的综合调度方法及装置

[1] 技术领域

[2] 本发明涉及计算机技术领域，尤其涉及一种针对中断持有者抢占问题的综合调度 方法及装置。

[3] 背景技术

[4] 系统虚拟化为允许多个虚拟机同时运行在一个物理服务器上创造了环境。虚拟机 监控器(VMM，Virtual Machine Monitor)将底层硬件抽象为虚拟资源，并管理虚拟机的运 行。I/O虚拟化是系统虚拟化的重要分支，负责为虚拟机提供I/O设备。随着网络互联速度的 持续增加，I/O虚拟化已经成为云数据中心的关键议题之一。

[5] 虚拟对称多处理(VSMP，Virtual Symmetric Multiprocessing)技术使得一个虚 拟机可以拥有多个虚拟处理器(VCPU，Virtual CPU)，这不仅带来了更高的并发度和资源利 用率，也有利于进一步提升I/O吞吐量和响应性。基于硬件的中断递交技术，如Message Signaling Interrupt、Posted Interrupt等，可以将I/O 设备中断高效地递交至目标虚拟 机进行处理。然而，在VSMP环境下，由于多个VCPU需要时分共享一个物理CPU核，这使得虚拟中断的处理过程变得复杂：负责处理虚拟中断的VCPU可能会被VMM调度器调度出去，导致中 断事件的处理延迟，进而影响虚拟机的I/O性能。

[6] 很多已有研究致力于提升虚拟对称多处理环境下的I/O吞吐量和响应性。现有的 解决方案大致可以分为三种：旁核(side-core)策略、调度相关策略以及中断重映射机制。旁核策略使用一个单独的CPU核专用于虚拟机的I/O处理，虽然可以保证I/O性能，但很容易 造成CPU资源的浪费。调度相关策略通过减少VMM调度器的时间片来减少每个VCPU在运行队 列中的等待时间，从而降低I/O延迟，但该方法会造成频繁的上下文切换，不利于CPU密集型 (PU-intensive)任务的运行。相比之下，中断重映射机制通过将I/O设备中断智能地重映射至虚拟机多个VCPU中最合适的一个，既能提升I/O性能，也不会引入其它副作用。

[7] 目前，中断重映射机制已经成为虚拟对称多处理环境下最具有吸引力的I/O性能解决方案之一。但是，在实际过程中，中断重映射机制可能会发生失效的风险，如何避免中断重映射机制失效是目前亟待解决的问题。

[8] 发明内容

[9] 如背景技术中所述，中断重映射机制会发生失效的风险。发明人对此进行研究，发现，如果虚拟机运行的是Linux操作系统，该系统中针对网络I/O设备的NAPI处理模式可能 会和中断重映射机制发生冲突，使得中断重映射机制失效。NAPI是Linux上采用的一种提高 网络处理效率的技术，它的核心概念就是不采用中断的方式读取数据，而代之以首先采用中断唤醒数据接收的服务程序，然后关闭设备中断进入轮询(POLL)模式读取数据，等轮询 模式结束时再开启设备中断。由于NAPI需要操作I/O设备的可中断性，而中断重映射需要修改设备中断的目的地址，如果不加以协调，会产生冲突以致中断重映射失效。

[10] 至此，发明人首先发现并定义了虚拟对称多处理环境下中断重映射机制可能存在的问题：中断持有者抢占(IHP，Interruptability Holder Preemption)问题。

[11] IHP问题的定义：假设虚拟机的某个VCPU当前正在处理某个网络I/O设备的中断，并进入了NAPI处理过程的轮询(POLL)模式，则该设备的可中断性被关闭，此时称该VCPU为中断持有者(IH,Interruptability Holder)。如果此时该IH被VMM调度器调度出去，那么就发生了IHP问题。在IH被再次调度并重新开启设备中断的这段时间里，该网络I/O设备无法 产生中断，即使它收到了外界的数据包。由于没有中断产生，中断重映射失效，该设备收到的数据包无法被即使处理，降低了I/O性能。

[12] 因此，发明人在上述研究的基础上提出一种针对虚拟多对称处理环境下中端持有 者抢占问题的综合调度方法及装置，以解决IHP问题，该方法基于目前广泛使用的KVM (Kernel-Based Virtual Machine)及其半虚拟化网络模型，通过对中断持有者运行状态的全局把握与分析，并兼顾系统的全局调度公平性，构建一种VCPU综合调度方法，有效消除 IHP问题，显著提升虚拟机的I/O性能。

[13] 本发明的技术方案提供一种针对中断持有者抢占问题的综合调度方法，包括如下步骤：步骤1：当虚拟网络设备接收到数据包时，从多个虚拟CPU中选择一个

虚拟CPU作为目标虚拟CPU，然后注入虚拟中断通知数据包的到达；步骤2：分析所述虚拟中断的信息，并将所述目标虚拟CPU标记为中断持有者；步骤3：目标虚拟CPU接收所述虚拟中断，触发对应的 中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模式读取数据包；步骤4：当要将 一虚拟CPU调离出物理CPU时，判断所述虚拟CPU是否被标记为中断持有者，若是，则继续调度该虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟CPU；步骤5：以轮询方式判 断优选虚拟CPU是否已经开启虚拟网络设备的中断，若已打开，则立即将所述优选虚拟CPU 调离物理CPU。

[14] 可选的，步骤1中根据中断重映射机制选择一虚拟CPU作为目标虚拟 CPU。

[15] 可选的，步骤4中，当要被调离出物理CPU的虚拟CPU未被标记为中断持有者，则立 即将该虚拟CPU调离出物理CPU。

[16] 可选的，步骤5中，在VM Exit处理函数中判断优选虚拟CPU是否已经打开虚拟网络设备的中断。

[17] 本发明的技术方案还提供一种针对中断持有者抢占问题的综合调度装置，包括：多个虚拟CPU；选择模块，用于当虚拟网络设备接收到数据包时，从多个虚拟CPU中选择一个虚拟CPU作为目标虚拟CPU，然后注入虚拟中断通知数据包的到达，所述目标虚拟CPU用于接收所述虚拟中断，触发对应的中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模 式读取数据；中断持有者检测模块，用于分析所述虚拟中断的信息，并将所述目标虚拟CPU 标记为中断持有者；调度模块，与所述中断持有者检测模块连接，用于当要将一虚拟CPU调离出物理CPU时，根据所述中断持有者检测模块标记的中断持有者信息，判断所述虚拟CPU 是否被标记为中断持有者，若是，则继续调度该虚拟 CPU运行在物理CPU上，并将该虚拟CPU 记录为优选虚拟CPU，并且当优选虚拟CPU开启虚拟网络设备的中断后，立即将所述优选虚拟CPU调离物理 CPU；判断模块，与所述调度模块连接，用于以轮询方式判断优选虚拟CPU 是否已经开启虚拟网络设备的中断，并将判断信息发送至调度模块。

[18] 可选的，所述选择模块用于根据中断重映射机制选择一虚拟CPU作为目标虚拟CPU。

[19] 可选的，所述调度模块包括：延时调度单元，与所述终端持有者检测模块连接，用于继续调度被标记为中断持有者的虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟CPU；即时调度单元，与所述判断模块连接，用于将未被标记为中断持有者的虚拟CPU和已关闭虚拟网络设备中断的优选虚拟CPU立即调离物理CPU；常规调度单元，与所述延时调 度单元连接，用于将需要调离的未被标记为中断持有者的虚拟CPU立即调离出物理CPU。

[20] 可选的，所述判断模块用于在VM Exit处理函数中判断优选虚拟CPU是否已经开启虚拟网络设备的中断。

[21] 本发明提供的一种前摄性的、基于中断持有者信息的VCPU综合调度方法具有以下有益的技术效果：

[22] (1)本发明采用中断重映射机制将虚拟中断递交至目前正在运行的虚拟 CPU上， 显著降低了由于虚拟CPU的调度延时引起的虚拟机网络响应性下降的问题。

[23] (2)本发明的方案只需要记录虚拟中断的目的VCPU信息，并不涉及虚拟机操作系统的改动，对虚拟机完全透明。

[24] (3)中断持有者抢占问题是造成中断重映射机制失效的根本原因，也是造成I/O虚拟化性能瓶颈的重要原因，本发明完全消除了中断重映射机制中断持有者抢占问题，使得虚拟机的网络I/O性能得到极大提高。

[25] 附图说明

[26] 图1为传统的半虚拟化环境下的中断处理流程示意图；

[27] 图2为本发明所述的中断持有者抢占问题的一个实例图；

[28] 图3为本发明的一个实施例的针对中断持有者抢占问题的综合调度方法的流程示意图；

[29] 图4为本发明的一个实施例的针对中断持有者抢占问题的综合调度装置的结构示意图。

[30] 具体实施方式

[31] 下面结合附图对本发明提供的一种针对中断持有者抢占问题的综合调度方法和调度装置的实施例做详细说明。所述实施例在本发明技术方案的前提下进行实施，并给出详细实施方式和具体操作过程，但是适用的平台不仅限于下述实施

例。

[32] 本实施例基于目前广泛使用的KVM(Kernel-Based Virtual Machine)及其半虚拟化网络I/O模型(由虚拟机中的前端驱动和VMM中的后端驱动构成)。

[33] 请参考图1，为传统的半虚拟化环境下的中断处理流程示意图。

[34] 在半虚拟化环境下，虚拟机的网络设备由运行在虚拟机内部的前端驱动和运行在 VMM中的后端驱动两部分组成。其中，前端驱动通过共享内存与后端驱动进行数据和控制信息的共享。当后端设备从物理网卡或虚拟交换机接收到网络数据包并将其拷贝至共享内存 后(路径①)，VMM会向虚拟机注入一个虚拟中断来触发数据包接收。虚拟中断目的虚拟CPU0的模拟LAPIC(Local Advanced Programmable Interrupt Controller)接收到该虚拟中断 (路径②)并触发存储在客户机IDT(Interrupt Descriptor Table)中的中断处理函数(路径③)。中断处理函数首先发起一次EOI(End of Interrupt)写操作通知模拟LAPIC 该虚拟 中断已经处理完成(路径④)，接着，中断处理函数进入NAPI处理过程(路径⑤)，关闭设备 (后端驱动)中断(路径⑥)并进入轮询(POLL)模式机制从共享内存中读取数据(路径⑦)。需要注意的是，NAPI机制利用一个 softirq线程来轮询访问共享内存中的网络数据包。一旦数据包接收完毕，虚拟CPU0会设置前端驱动退出轮询模式并打开其中断能力(路径⑧)，等待接收下一次虚拟中断注入。

[35] 请参考图2，为本发明所述的IHP问题的一个实例图。

[36] 在虚拟对称多处理架构下，虚拟机V共有两个VCPU(虚拟CPU 0和虚拟 CPU 1)，虚 拟CPU 0首先收到一个虚拟中断并在T0时刻禁用了前端设备的中断能力。如果虚拟CPU 0在 T1时刻被调度器调离物理CPU 0，由于此时前端设备禁止中断，后续的经由中断重映射机制重定向的虚拟CPU 1处理的网络 I/O事件将不能被及时递交至虚拟机V，直到VCPU 0在T2时刻重新被调回物理CPU 1并打开前端设备的中断能力。因此，在T1至T2的时间段中，中断重映射机制完全失效，虚拟机的网络I/O性能依然受制于VCPU的调度延迟，造成了本发明前文所述的中断持有者抢占问题。

[37] 请参考图3，为本发明的针对中断持有者抢占问题的综合调度方法的流程示意图， 包括如下步骤:

[38] 步骤S1：当虚拟网络设备收到数据包时，从客户机即虚拟机的多个虚拟 CPU中选择一个虚拟CPU作为目标虚拟CPU，然后注入虚拟中断通知数据包的到达。在该实施例中，虚拟机网络设备由前端驱动和后端驱动两部分组成，包括多个虚拟CPU。具体的，运行在VMM中 的虚拟机后端驱动从网桥或虚拟交换机接收到网络数据包并将数据包拷贝至前端驱动的共享内存中，VMM利用中断重映射机制为对应虚拟机选择一个合适的VCPU作为目标虚拟 CPU，然后注入虚拟中断通知数据包的到达。在其他具体实施例中，也可以采用其他方式选 择合适的VCPU作为目标虚拟CPU。

[39] 步骤S2：分析所述虚拟中断的信息，并将所述目标虚拟CPU标记为中断持有者。具体的，位于VMM内部的中断持有者检测模块分析该虚拟中断的信息并标记目标虚拟CPU为中断持有者(IH)，并形成IH信息。

[40] 步骤S3：目标虚拟CPU接收所述虚拟中断，触发对应的中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模式读取数据包。具体的，该虚拟中断由目标虚拟CPU接收并触发相应的中断处理函数，进入NAPI处理过程，关闭设备(后端驱动)中断并进入轮询 (POLL)模式从共享内存中读取数据。

[41] 步骤S4：当要将一虚拟CPU调离出物理CPU时，判断所述虚拟CPU是否被标记为中断持有者，若是，则继续调度该虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟CPU。具体的，当要将步骤S2中记录的中断持有者(VCPU)调离物理CPU时，位于VMM内部的延

迟调度模块基于中断持有者检测模块提供的IH信息，判断该虚拟CPU是否为中断持有者。如果是，则该延迟调度模块将继续调度该虚拟CPU运行在物理CPU上，并将其记录为 Favored

VCPU(优选虚拟CPU)。如果不是，则立即将该虚拟CPU调离出物理CPU。

[42] 步骤S5：以轮询方式判断优选虚拟CPU是否已经开启虚拟网络设备的中断，若已打 开，则立即将所述优选虚拟CPU调离物理CPU。具体的，位于VMM 内部的即时调度模块在频繁发生的VM Exit处理函数中判断步骤S4中所记录的Favored VCPU是否已经打开后端驱动的中断能力。对于已经打开设备中断的Favored VCPU，即时调度模块立即将调离物理CPU。本发明利用Intel VT-x 技术中的VM

Exit机制，在VM Exit处理函数中，以一种类轮询方式判断Favored VCPU是否已经打开设备中断，一旦中断能力被打开，即时调度模块立即将该 Favored VCPU调离物理CPU，为系统内其他进程提供运行机会，既充分保证系统的全局调度公平性又能完全避免IHP问题，获得整体性能的提升。

[43] 与传统的半虚拟化环境下中断处理流程相比，本实施例的中断处理流程中加入了中断持有者信息的监控以及更加精细、综合的VCPU调度策略。

[44] 请参考图4，为本发明的实施例提供的针对中断持有者抢占问题的综合调度装置的结构示意图。

[45] 所述综合调度装置包括：多个虚拟CPU501、选择模块502、中断持有者检测模块 503、调度模块504以及判断模块505。

[46] 所述选择模块502用于当虚拟网络设备接收到数据包时，从多个虚拟 CPU501中选 择一个虚拟CPU作为目标虚拟CPU，然后注入虚拟中断通知数据包的到达，所述目标虚拟CPU 用于接收所述虚拟中断，触发对应的中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模式读取数据。所述选择模块502可以根据中断重映射机制选择一虚拟CPU作为目标虚拟CPU。

[47] 中断持有者检测模块503，与所述选择模块连接，用于分析所述虚拟中断的信息， 并将所述选择模块502选择的目标虚拟CPU标记为中断持有者。

[48] 调度模块504，与所述中断持有者检测模块503连接，用于当要将一虚拟 CPU调离出物理CPU时，根据所述中断持有者检测模块503标记的中断持有者信息，判断所述虚拟CPU 是否被标记为中断持有者，若是，则继续调度该虚拟CPU运行在物理CPU上，并将该虚拟CPU 记录为优选虚拟CPU，并且当优选虚拟CPU开启虚拟网络设备的中断后，立即将所述优选虚 拟CPU调离物理CPU。

[49] 判断模块505，与所述调度模块504连接，用于以轮询方式判断调度模块 504记录 的优选虚拟CPU是否已经开启虚拟网络设备的中断，并将判断信息发送至调度模块504。所 述判断模块505利用Intel VT-x技术中的VM Exit机制，在频繁发生的VM Exit处理函数中，以一种类轮询方式判断Favored VCPU是否已经打开设备中断。

[50] 该具体实施方式中，所述调度模块504，包括：延时调度单元5041，与所述终

端持有者检测模块503和判断模块505连接，用于继续调度被标记为中断持有者的虚拟CPU运行在 物理CPU上，并将该虚拟CPU记录为优选虚拟CPU，并将优选虚拟CPU的记录信息发送给判断 模块505；即时调度单元5042，与所述判断模块505连接，用于将已关闭虚拟网络设备中断的优选虚拟CPU立即调离物理CPU；常规调度单元5043，与所述延时调度单元5041连接，用于将需要调离的未被标记为中断持有者的虚拟CPU立即调离出物理CPU。

[51] 以上详细描述了本发明的较佳具体实施例。应当理解，本领域的普通技术无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此，凡本技术领域中技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案，皆应在由权利要求书所确定的保护范围内。

权利要求书

[权利要求 1] 一种针对中断持有者抢占问题的综合调度方法，其特征在于，包括以下步骤：

步骤1：当虚拟网络设备接收到数据包时，从多个虚拟CPU中选择一个虚拟CPU作为目标 虚拟CPU，然后注入虚拟中断通知数据包的到达；

步骤2：分析所述虚拟中断的信息，并将所述目标虚拟CPU标记为中断持有者；

步骤3：目标虚拟CPU接收所述虚拟中断，触发对应的中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模式读取数据包；

步骤4：当要将一虚拟CPU调离出物理CPU时，判断所述虚拟CPU是否被标记为中断持有者，若是，则继续调度该虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟CPU；

步骤5：以轮询方式判断优选虚拟CPU是否已经开启虚拟网络设备的中断，若已打开，则立即将所述优选虚拟CPU调离物理CPU。

[权利要求 2] 根据权利要求1所述的针对中断持有者抢占问题的综合调度方法，其特征在于，步骤 1中根据中断重映射机制选择一虚拟CPU作为目标虚拟CPU。

[权利要求 3] 根据权利要求1所述的针对中断持有者抢占问题的综合调度方法，其特征在于，步骤 4中，当要被调离出物理CPU的虚拟CPU未被标记为中断持有者，则立即将该虚拟CPU调离出物理CPU。

[权利要求 4] 根据权利要求1所述的针对中断持有者抢占问题的综合调度方法，其特征在于，步骤 5中，在VM Exit处理函数中判断优选虚拟CPU是否已经打开虚拟网络设备的中断。

[权利要求 5] 一种针对中断持有者抢占问题的综合调度装置，其特征在于，包括：

多个虚拟CPU；

选择模块，用于当虚拟网络设备接收到数据包时，从多个虚拟CP

U中选择一个虚拟CPU作为目标虚拟CPU，然后注入虚拟中断通知数据包的到达，所述目标虚拟CPU用于接收所述虚拟中断，触发对应的中断处理函数，关闭所述虚拟网络设备的中断并进入轮询模式读取数据；

中断持有者检测模块，用于分析所述虚拟中断的信息，并将所述目标虚拟CPU标记为中断持有者；

调度模块，与所述中断持有者检测模块连接，用于当要将一虚拟CPU调离出物理CPU时，根据所述中断持有者检测模块标记的中断持有者信息，判断所述虚拟CPU是否被标记为中断持有者，若是，则继续调度该虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟 CPU，并且当优选虚拟CPU开启虚拟网络设备的中断后，立即将所述优选虚拟CPU调离物理 CPU；

判断模块，与所述调度模块连接，用于以轮询方式判断优选虚拟CPU是否已经开启虚拟网络设备的中断，并将判断信息发送至调度模块。

[权利要求 6] 根据权利要求5所述的针对中断持有者抢占问题的综合调度装置，其特征在于，所述选择模块用于根据中断重映射机制选择一虚拟CPU作为目标虚拟CPU。

[权利要求 7] 根据权利要求5所述的针对中断持有者抢占问题的综合调度装置，其特征在于，所述调度模块包括：延时调度单元，与所述终端持有者检测模块连接，用于继续调度被标记为中

断持有者的虚拟CPU运行在物理CPU上，并将该虚拟CPU记录为优选虚拟CPU；即时调度单元， 与所述判断模块连接，用于将未被标记为中断持有者的虚拟CPU和已关闭虚拟网络设备中断的优选虚拟CPU立即调离物理CPU；常规调度单元，与所述延时调度单元连接，用于将需要调离的未被标记为中断持有者的虚拟CPU立即调离出物理CPU。

[权利要求 8] 根据权利要求5所述的针对中断持有者抢占问题的综合调度装置，

其特征在于，所述 判断模块用于在VM Exit处理函数中判断优选虚拟CPU是否已经开启虚拟网络设备的中断。

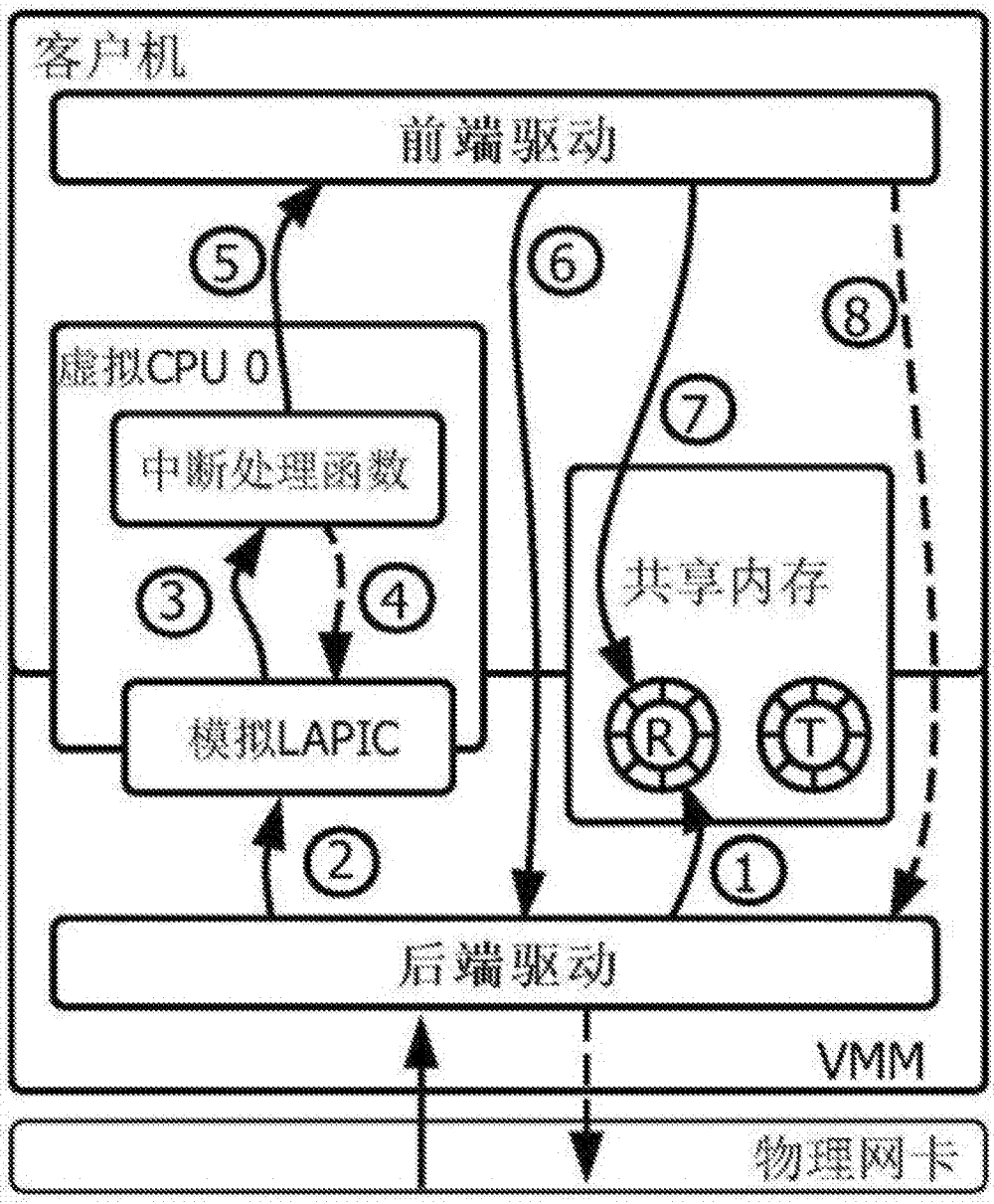

客户机
前端驱动
5
6
8
虚拟CPU 0
7
中断处理函数
共享内存
3
4
模拟LAPIC
R
T
2
1
后端驱动
VMM
物理网卡

图 1

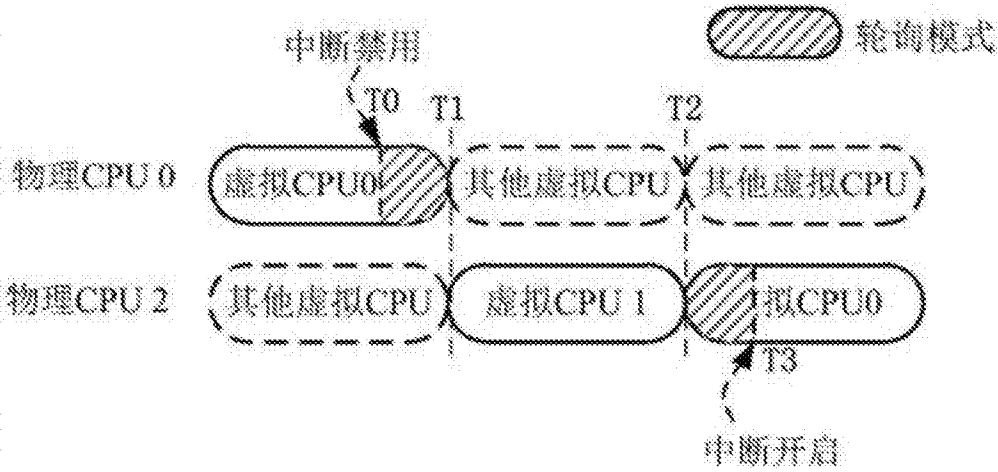

轮询模式
中断禁用
T0
T1
T2
物理CPU 0
虚拟CPU0
其他虚拟CPU
其他虚拟CPU
物理CPU 2
其他虚拟CPU
虚拟CPU 1
虚拟CPU0
T3
中断开启

图 2

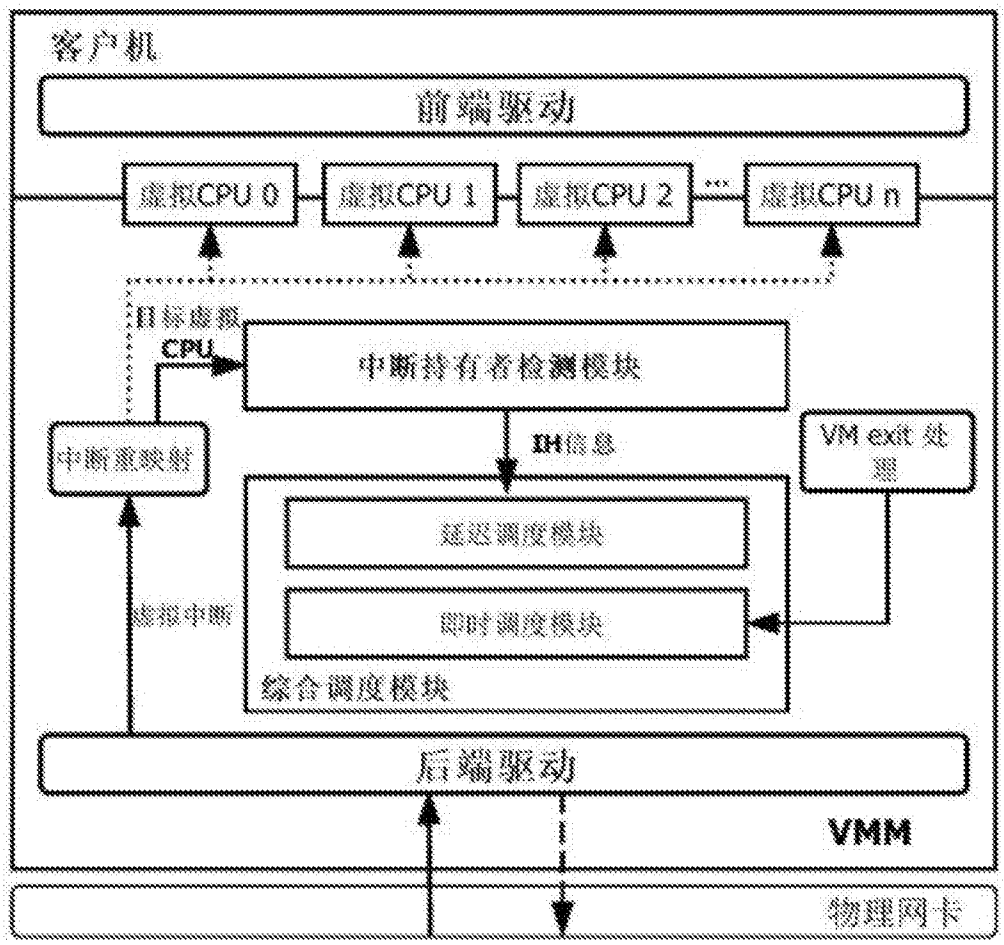

客户机
前端驱动
虚拟CPU 0
虚拟CPU 1
虚拟CPU 2
...
虚拟CPU n
目标虚拟CPU
中断持有者检测模块
中断重映射
IH信息
VM exit 处理
延迟调度模块
即时调度模块
综合调度模块
虚拟中断
后端驱动
VMM
物理网卡

图 3

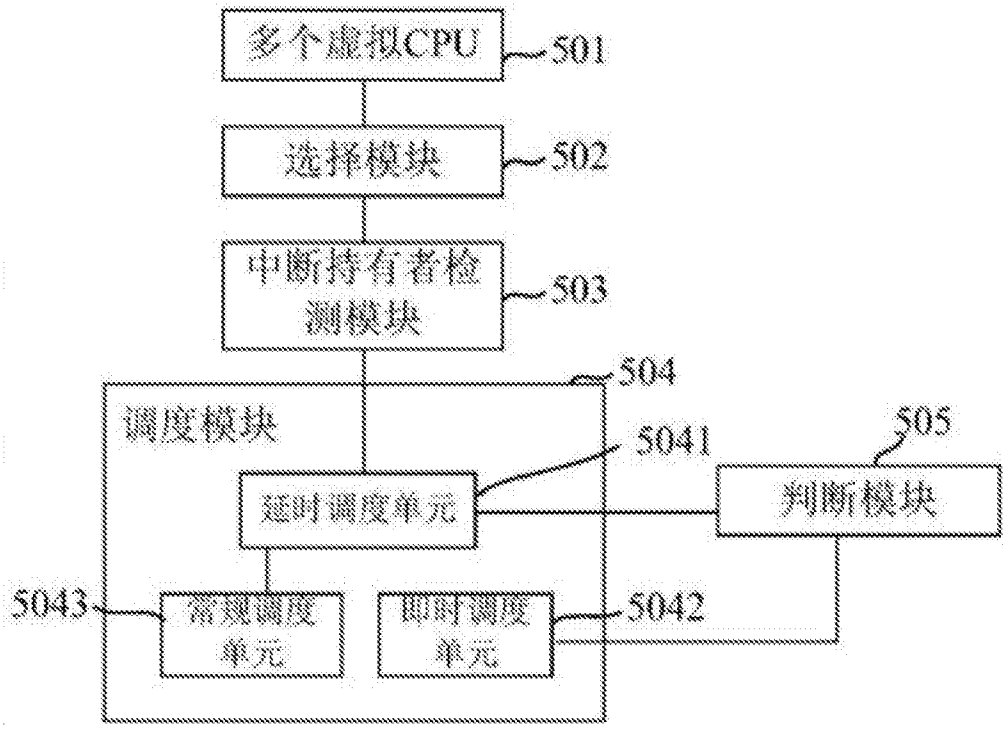

多个虚拟CPU
501
选择模块
502
中断持有者检测模块
503
504
调度模块
5041
505
延时调度单元
判断模块
5043
常规调度单元
即时调度单元
5042

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/092618

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/26(2006.01)i; H04L 29/08(2006.01)i; G06F 9/455(2006.01)i; G06F 9/48(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L,G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; VEN; USTXT; EPTXT; WOTXT; CNKI; IEEE EXPLORE: 中断, 持有者, 抢占, 虚拟CPU, 调离, 脱离, 去激活, 中断重映射, 物理CPU, 注入, 轮询, VCPU, interruptability holder, virtual CPU, POLL, physical CPU, preemption, NAPI, remapping, stellenbosch, inactivation, deactivation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108123850 A (SHANGHAI JIAOTONG UNIVERSITY) 05 June 2018 (2018-06-05) description, paragraphs [0010]-[0016] and [0032]-[0045]	1-8
A	CN 103744716 A (SHANGHAI JIAOTONG UNIVERSITY) 23 April 2014 (2014-04-23) description, paragraphs [0056]-[0061]	1-8
A	黄智强 (HUANG, Zhiqiang). "基于SR-IOV的高性能I/O虚拟化研究与优化 (Research and Optimizations for High Performance I/O Virtualizaion Based on SR-IOV)" *中国优秀硕士学位论文全文数据库(电子期刊)工程科技I辑 (Chinese Master's Theses Full-text Database (electronic journal), Engineering Science & Technology),* No. 7, 15 July 2013 (2013-07-15), ISSN: 1674-0246, pp. I137-38	1-8
A	CN 106095578 A (SHANGHAI JIAOTONG UNIVERSITY) 09 November 2016 (2016-11-09) entire document	1-8
A	CN 105210041 A (ADVANCED RISC MACH LIMITED) 30 December 2015 (2015-12-30) entire document	1-8

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 August 2018	**03 September 2018**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/092618

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2010138208 A1 (HITACHI, LTD.) 03 June 2010 (2010-06-03) entire document	1-8
A	LI, Jian et al. "vINT: Hardware-Assisted Virtual Interrupt Remapping for SMP VM with Scheduling Awareness" *2015 IEEE 7th International Conference on Cloud Computing Technology and Science,* 03 December 2015 (2015-12-03), pp. 234-241	1-8
A	Guan, Haibing, et al. "SR-IOV Based Network Interrupt-Free Virtualization with Event Based Polling" *IEEE Journal on Selected Areas in Communications,* Vol. 31, No. 12, 02 December 2013 (2013-12-02), ISSN: 0733-8716, pp. 2596-2609	1-8

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/092618

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108123850	A	05 June 2018	None			
CN	103744716	A	23 April 2014	CN	103744716	B	07 September 2016
				US	2016259664	A1	08 September 2016
				US	9697041	B2	04 July 2017
				WO	2015106497	A1	23 July 2015
CN	106095578	A	09 November 2016	None			
CN	105210041	A	30 December 2015	GB	2528600	A	27 January 2016
				WO	2014188155	A1	27 November 2014
				GB	201518020	D0	25 November 2015
				KR	20160010863	A	28 January 2016
				US	9378162	B2	28 June 2016
				US	2014351471	A1	27 November 2014
US	2010138208	A1	03 June 2010	JP	2010128911	A	10 June 2010
				JP	5352848	B2	27 November 2013
				US	8516479	B2	20 August 2013

国际检索报告

国际申请号
PCT/CN2018/092618

A. 主题的分类

H04L 12/26(2006.01)i; H04L 29/08(2006.01)i; G06F 9/455(2006.01)i; G06F 9/48(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L, G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNABS;CNTXT;VEN;USTXT;EPTXT;WOTXT;CNKI;IEEE EXPLORE:中断,持有者,抢占,虚拟CPU,调离,脱离,去激活,中断重映射,物理CPU,注入,轮询,VCPU,interruptability holder, virtual CPU,POLL,physical CPU,preemption,NAPI,remapping,stellenbosch, inactivation,deactivation

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 108123850 A (上海交通大学) 2018年 6月 5日 (2018-06-05) 说明书[0010]-[0016]、[0032]-[0045]段	1-8
A	CN 103744716 A (上海交通大学) 2014年 4月 23日 (2014-04-23) 说明书第[0056]-[0061]段	1-8
A	黄智强. "基于SR-IOV 的高性能I/O 虚拟化研究与优化" 中国优秀硕士学位论文全文数据库(电子期刊)信息科技辑, 第7期, 2013年 7月 15日 (2013-07-15), ISSN：1674-0246, 第I137-38页	1-8
A	CN 106095578 A (上海交通大学) 2016年 11月 9日 (2016-11-09) 全文	1-8
A	CN 105210041 A (ARM 有限公司) 2015年 12月 30日 (2015-12-30) 全文	1-8
A	US 2010138208 A1 (HITACHI LTD) 2010年 6月 3日 (2010-06-03) 全文	1-8

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2018年 8月 3日	2018年 9月 3日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 匡仁炳 电话号码 86-(20)-28950886

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告

国际申请号
PCT/CN2018/092618

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	Jian Li 等. "vINT: Hardware-Assisted Virtual Interrupt Remapping for SMP VM with Scheduling Awareness" 2015 IEEE 7th International Conference on Cloud Computing Technology and Science, 2015年 12月 3日 (2015-12-03), 第234-241页	1-8
A	HaiBing Guan 等. "SR-IOV Based Network Interrupt-Free Virtualization with Event Based Polling" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, 第31卷，第12期，2013年 12月 2日 (2013-12-02), ISSN: 0733-8716, 第2596-2609页	1-8

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/092618

检索报告引用的专利文件			公布日（年/月/日）	同族专利			公布日（年/月/日）
CN	108123850	A	2018年 6月 5日	无			
CN	103744716	A	2014年 4月 23日	CN	103744716	B	2016年 9月 7日
				US	2016259664	A1	2016年 9月 8日
				US	9697041	B2	2017年 7月 4日
				WO	2015106497	A1	2015年 7月 23日
CN	106095578	A	2016年 11月 9日	无			
CN	105210041	A	2015年 12月 30日	GB	2528600	A	2016年 1月 27日
				WO	2014188155	A1	2014年 11月 27日
				GB	201518020	D0	2015年 11月 25日
				KR	20160010863	A	2016年 1月 28日
				US	9378162	B2	2016年 6月 28日
				US	2014351471	A1	2014年 11月 27日
US	2010138208	A1	2010年 6月 3日	JP	2010128911	A	2010年 6月 10日
				JP	5352848	B2	2013年 11月 27日
				US	8516479	B2	2013年 8月 20日