

(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
2014 年 12 月 4 日 (04.12.2014)



(10) 国际公布号
WO 2014/190486 A1

- (51) 国际专利分类号:
G06F 15/167 (2006.01)
- (21) 国际申请号: PCT/CN2013/076329
- (22) 国际申请日: 2013 年 5 月 28 日 (28.05.2013)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 雷晓松 (LEI, Xiaosong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: 北京同立钧成知识产权代理有限公司 (LEADER PATENT & TRADEMARK FIRM); 中国北京市海淀区西直门北大街 32 号枫蓝国际 A 座 8F-6, Beijing 100082 (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, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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))。

(54) Title: METHOD AND SYSTEM FOR SUPPORTING RESOURCE ISOLATION UNDER MULTI-CORE ARCHITECTURE

(54) 发明名称: 支持多核架构下资源隔离的方法及系统

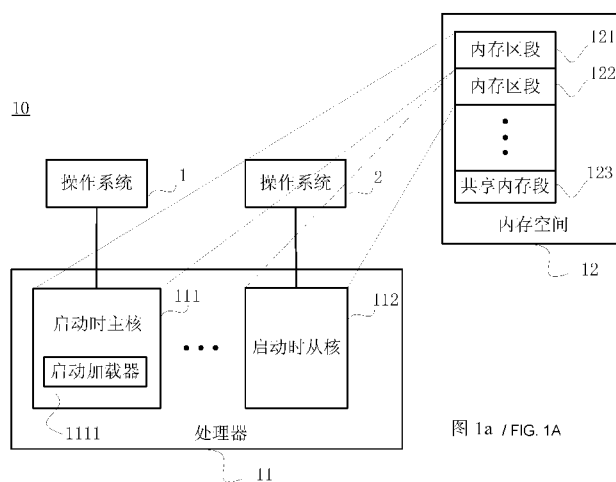


图 1a / FIG. 1A

(57) Abstract: Embodiments of the invention provide a method and a system for supporting resource isolation under a multi-core architecture. In the method and the system for supporting resource isolation under a multi-core architecture provided in the embodiments of the present invention, operating systems operating on different processing cores of a multi-core processor can operate independently without interference to each other, by using isolation of inter-core operating systems, isolation of memory sections and isolation of I/O resources, so that advantages of high integration level and low comprehensive cost of the multi-core processor can be fully utilized, and failure regions of the multi-core processor can be maintained at a single hard disk, thereby providing relatively high reliability.

(57) 摘要: 本发明实施例提供一种支持多核架构下资源隔离的方法及系统。本发明实施例提供的支持多核架构下资源隔离的方法及系统, 通过核间操作系统隔离、内存区段隔离、I/O 资源隔离的方式, 使得在多核处理器的不同处理核上运行的操作系统能够独立运行而互不影响, 从而, 充分利用了多核处理器高集成度、综合成本低的优势, 实现了多核处理器的故障域维持在单块硬盘, 具有较高的可靠性。

- 1 OPERATING SYSTEM
2 OPERATING SYSTEM
11 PROCESSOR
12 MEMORY SPACE
111 MAIN CORE DURATION THE START
112 AUXILIARY CORE DURATION THE START
121 MEMORY SECTION
122 MEMORY SECTION
123 SHARED MEMORY SECTION
1111 START LOADER

支持多核架构下资源隔离的方法及系统

技术领域

- 5 本发明实施例涉及计算机技术，尤其涉及一种支持多核架构下资源隔离的方法及系统。

背景技术

- 10 基于大规模统一资源池的云存储系统，以数据存储和管理为核心，支持多业务的统一部署，高效利用磁盘，可按需提供减少投资损失，通过合理调度实现有效利用资源、降低综合能耗。

现有技术采用多核 ARM 处理器带少量硬盘的服务器，例如基于精简指令集(Reduced Instruction Set Computing，简称 RISC)的 ARM 服务器方式，可以发挥 ARM 处理器低功耗、高集成度、综合成本低的优势。

- 15 然而，现有技术的多核 ARM 处理器带多块硬盘的情况，多核处理器上运行一个操作系统，在该操作系统失效时，服务器的失效粒度为多块硬盘，并且由于单块硬盘的容量大而输入/输出(Input/Output，简称 I/O)吞吐带宽却较小，使得数据恢复量以及网络传输数据量都以倍数增长，故障恢复粒度较粗，给系统造成较大的压力。

20

发明内容

本发明实施例提供一种支持多核架构下资源隔离的方法及系统，以使多核处理器的故障域维持在单块硬盘。

- 25 第一方面，本发明实施例提供一种支持多核架构下资源隔离的系统，该系统包括：处理器，所述处理器包括多个处理核，所述多个处理核中包括一个启动时主核以及至少一个启动时从核；内存空间，所述内存空间包括共享内存段，以及各处理核的内存区段；所述共享内存段中存储有运行信息，以及各处理核的运行时入口地址；所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核
30 的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像

文件在各自的内存区段中的存储地址；其中，所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址分别写入到所述共享内存段中；待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

在第一方面的第一种可能的实现方式中，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行输入/输出 I/O 资源的仲裁控制。

根据第一方面的第一种可能的实现方式，在第二种可能的实现方式中，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

根据第一方面的第一种可能的实现方式，在第三种可能的实现方式中，运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

根据第一方面、第一方面的第一种至第三种可能的实现方式的任意一种，在第四种可能的实现方式中，所述启动时主核内设置有启动加载器；所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：所述启动加载器根据所述运行信息，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中。

根据第一方面、第一方面的第一种至第三种可能的实现方式的任意一种，在第五种可能的实现方式中，所述启动时主核内设置有启动加载器；所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：所述启动加载器根据所述运行信息，在所述处理器的至少一个处理核的内存区段中存储所需要运行的操作系统的内核镜像文件，运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据所述运行信息，将所述处理器的需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到各所述其余处理核的内存区段中。

根据第一方面、第一方面的第一种至第五种可能的实现方式的任意一

种，在第六种可能的实现方式中，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件，并且，所述构成所述处理核组的所有处理核共用同一段内存区段；或者，各处理核运行互不相同的操作系统的内核镜像文件。

- 5 第二方面，本发明实施例提供一种支持多核架构下资源隔离的方法，应用于包括处理器和内存空间的支持多核架构下资源隔离的系统，其中：所述处理器包括多个处理核，所述多个处理核中包括一个启动时主核以及至少一个启动时从核；所述内存空间包括共享内存段，以及各处理核的内存区段；所述共享内存段中存储有运行信息，以及各处理核的运行时入口地址；
- 10 所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像文件在各自的内存区段中的存储地址；

所述支持多核架构下资源隔离的方法包括：

- 15 所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址分别写入所述共享内存段中；

待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

- 20 在第二方面的第一种可能的实现方式中，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行输入/输出 I/O 资源的仲裁控制。

- 根据第二方面的第一种可能的实现方式，在第二种可能的实现方式中，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述
- 25 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

根据第二方面的第一种可能的实现方式，在第三种可能的实现方式中，运行时从核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

- 根据第二方面、第二方面的第一种至第三种可能的实现方式的任意一
- 30 种，在第四种可能的实现方式中，所述启动时主核内设置有启动加载器；

所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

所述启动加载器根据所述运行信息，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中。

5 根据第二方面、第二方面的第一种至第三种可能的实现方式的任意一种，在第五种可能的实现方式中，所述启动时主核内设置有启动加载器；

所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

10 所述启动加载器根据所述运行信息，在所述处理器的至少一个处理核的内存区段中存储所需要运行的操作系统的内核镜像文件，运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据所述运行信息，将所述处理器的需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到各所述其余处理核的内存区段中。

15 根据第二方面、第二方面的第一种至第五种可能的实现方式的任意一种，在第六种可能的实现方式中，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件，并且，所述构成所述处理核组的所有处理核共用同一段内存区段；或者，各处理核运行互不相同的操作系统的内核镜像文件。

20 本发明实施例提供的支持多核架构下资源隔离的方法及系统，通过核间操作系统隔离、内存区段隔离、I/O 资源隔离的方式，使得在多核处理器的不同处理核上运行的操作系统能够独立运行而互不影响，从而，充分利用了多核处理器高集成度、综合成本低的优势，实现了多核处理器的故障域维持在单块硬盘，具有较高的可靠性。

25 附图说明

图 1a 为本发明实施例一提供的支持多核架构下资源隔离的系统的结构示意图；

图 1b 为采用本发明实施例一的支持多核架构下资源隔离的系统的云存储系统的拓扑结构示意图；

30 图 2 为图 1a 所示的支持多核架构下资源隔离的系统中的共享内存段

的局部结构示意图；

图 3 为本发明实施例二提供的支持多核架构下资源隔离的系统的结构示意图；

图 4a 为本发明实施例二提供的支持多核架构下资源隔离的系统的运行时从核对外部设备进行访问的一实现示意图；

图 4b 为本发明实施例二提供的支持多核架构下资源隔离的系统的运行时从核对外部设备进行访问的另一实现示意图；

图 5 为本发明实施例三提供的支持多核架构下资源隔离的方法的流程图。

10

具体实施方式

实施例一

图 1a 为本发明实施例一提供的支持多核架构下资源隔离的系统的结构示意图。图 1a 所示的支持多核架构下资源隔离的系统 10 适用于基于大规模统一资源池的云存储系统。图 1b 为采用本发明实施例一的支持多核架构下资源隔离的系统的云存储系统的拓扑结构示意图。如图 1b 所示，云存储系统 100 包括：交换机 102、至少一个支持多核架构下资源隔离的系统 10、至少一个应用服务器 101。其中，应用服务器 101 与交换机 102 连接，支持多核架构下资源隔离的系统 10 与交换机 102 连接。

20 如图 1a 所示，本实施例的支持多核架构下资源隔离的系统 10 包括：处理器 11 和内存空间 12，其中：

处理器 11，包括多个处理核：一个启动时主核 111 以及至少一个启动时从核 112；

25 内存空间 12，包括共享内存段 123、启动时主核 111 的内存区段 121、启动时从核 112 的内存区段 122；共享内存段 123 中存储有运行信息，以及各处理核的运行时入口地址；所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像文件在各自的内存区段中的存储地址；

30 其中，处理器 11 将各处理核所需要运行的操作系统的内核镜像文件

分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址写入共享内存段 123 中；待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

具体地，如图 1a 所示，处理器 11 是包括多个处理核的多核处理器，例如可以采用 ARM 或 Intel ATOM 处理器，但是本发明并不以此为限。处理器 11 包括多个处理核，图 1a 示例性的示出处理器 11 包括启动时主核 111 以及一个启动时从核 112 的情形，然而，处理器 11 可以包括一个或多个启动时从核，在此不再赘述。启动时主核 111，是指由硬件预先确定的设置启动加载器 1111 的处理核，并且启动时主核 111 能够触发其他处理核启动并运行；启动时从核 112，是指处理器 11 在启动时，通过启动时主核 111 触发而启动的其他处理核。具体应用中，所述启动时主核内设置的启动加载器 1111，例如是 UBoot，可完成设备初始化，也可以将操作系统的内核镜像文件存储到需要运行该操作系统的处理核所的内存区段中。图 1a 示出了启动时主核 111、启动时从核 112 分别与操作系统 1、操作系统 2 的对应关系，其中，操作系统 1、操作系统 2 可以是 LINUX、Windows、Solaris、DOS、OS/2、UNIX、XENIX、Netware，或者同一操作系统的不同版本，并且，操作系统 1 和操作系统 2 可以为完全相同的操作系统，也可以为同一种类操作系统的不同版本，或者不同种类的操作系统，但本发明并不以此为限。

图 2 为图 1a 所示的支持多核架构下资源隔离的系统中的共享内存段的局部结构示意图。如图 1a 和图 2 所示，共享内存段 123 中存储有运行信息 123a、启动时从核 112 的运行时入口地址 123b，以及启动时主核 111 的运行时入口地址 123c；其中，运行信息 123a，记录了内存区段的规划信息，至少包括：各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址，即如下信息：（1）启动时主核 111 运行操作系统 1，为启动时主核 111 分配的内存区段 121 的大小和起始地址；（2）启动时从核 112 运行操作系统 2，启动时从核 112 的内存区段 122 的大小和起始地址；（3）共享内存段 123 的大小和起始地址。所述各内存区段相互独立，相互隔离，实现了为不同的操作系统分配相互独立的内存区段。如果未指定各内存区段的起始地址，将缺省从地址 0 开始，或者从前一个内存区

段的结束地址开始作为该内存区段的起始地址。运行信息地址 1231 是运行信息 123a 在共享内存段 123 中的存储地址，各处理核都需要遵循运行信息地址 1231 的约定，通过运行信息地址 1231 查询运行信息，可选的，在每个处理核存储的操作系统的内核镜像文件中分别指示各个处理核的运行信息地址，并且运行信息可以在系统运行过程中动态修改和调整。启动时从核 112 的运行时入口地址指针 1232 指向启动时从核 112 的运行时入口地址 123b，其中，运行时入口地址 123b，用以指示启动时从核 112 所需要运行的操作系统 2 的内核镜像文件在内存区段 122 中的存储地址。启动时主核 111 的运行时入口地址指针 1233 指向启动时主核 111 的运行时入口地址 123c，其中，运行时入口地址 123c，用以指示启动时主核 111 所需要运行的操作系统 1 的内核镜像文件在内存区段 121 中的存储地址。每个处理核的运行时入口地址指针是系统启动前在 UBoot 中为每个处理核分别预设的。

本发明实施例提供的支持多核架构下资源隔离的系统，通过核间操作系统隔离、内存区段隔离的方式，使得在多核处理器的不同处理核上运行的操作系统能够独立运行而互不影响，当单个处理核或其上运行的操作系统失效时，不会影响到其他处理核或其他处理核上运行的操作系统，从而，充分利用了多核处理器高集成度、综合成本低的优势，实现了多核处理器的故障域维持在单块硬盘，具有较高的可靠性。

在上述实施例的基础上，处理器 11 将各处理核所需要运行的操作系统的内核镜像文件存储到各处理核的内存区段中的实现方式，至少包括以下两种方式：

方式 1：启动加载器 1111 根据运行信息 123a，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中。

方式 2：启动加载器 1111 根据运行信息 123a，将所述处理器的至少一个处理核所需要运行的操作系统的内核镜像文件存储到该处理核的内存区段中，该处理核运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据运行信息，将需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到其余处理核所的内存区段中。

具体地，在实现方式 2 中，启动加载器 1111 根据运行信息 123a，将

所述处理器的至少一个处理核所需要运行的操作系统的内核镜像文件存储到该处理器的内存区段中，例如，启动加载器 1111 根据运行信息 123a，将启动时主核 111 所需要运行的操作系统 1 的内核镜像文件存储到为启动时主核 111 分配的内存区段 121 中；接着，启动时主核 111 运行内存区段 5 121 中的操作系统 1 的内核镜像文件；随后，启动时主核 111 通过运行的操作系统 1，根据运行信息 123a 将启动时从核 112 所需要运行的操作系统 2 的内核镜像文件存储到启动时从核 112 的内存区段 122 中，从而将每个处理核需要运行的操作系统的内核镜像文件在各自的内存区段中展开。需要说明的是，本实施例示例性地描述了启动时主核 111 作为最先运行操作系统的处理核，但是本领域技术人员应该理解，最先启动运行操作系统的处理核并不只限于启动时主核 111，处理器 11 的任意处理核都可能通过例如预先指定的方式确定为最先启动运行操作系统的处理核。

本发明实施例提供的支持多核架构下资源隔离的系统，通过将每个处理核需要运行的操作系统的内核镜像文件在各自的内存区段中展开，使得不同的处理核的内存区段中存储各自需要运行的操作系统的内核镜像文件，并且各内存区段上的资源是相互隔离的。

在上述实施例的基础上，运行信息中记录的各处理核与各操作系统之间的对应关系，至少包括以下两种关系：

对应关系 1：由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件。可选的，构成所述处理核组的所有处理核共用同一段内存区段。

具体地，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件。构成所述处理核组的所有处理核可以共用同一段内存区段。处理核组的数目可以不止一个，并且各个处理核组各自运行的操作系统相互独立，为不同的操作系统分配相互独立的内存区段，使得单个操作系统的失效，不会影响到其他的处理核组或其上运行的操作系统，实现了多核处理器的故障域维持在单块硬盘。

对应关系 2：各处理核运行互不相同的操作系统的内核镜像文件。

可选的，根据运行信息，处理器 11 中可以存在处理核组和独立的单个处理核同时存在的情形。

在上述实施例的基础上，在处理核组或独立的单个处理核上运行着相互独立的操作系统，实现了核间操作系统隔离。鉴于多核处理器的 I/O 资源以及协处理器全局共享，当各处理核启动后，需要在各处理核中选择一个处理核进行统一 I/O 资源的仲裁控制。进一步，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行 I/O 资源的仲裁控制。

可选的，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

可选的，运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

下面通过一个具体的实施例对上述支持多核架构下资源隔离的系统的具体实现，进行详细的说明。

实施例二

图 3 为本发明实施例二提供的支持多核架构下资源隔离的系统的结构示意图。本实施例基于上述实施例来实现。如图 3 所示，本实施例的支持多核架构下资源隔离的系统 20 包括：处理器 21 和内存空间 22，其中：处理器 21，包括启动时主核 111、启动时从核 112、启动时从核 211、启动时从核 212，启动时主核 111 内设置有启动加载器 1111；并且，启动时主核 111 与启动时从核 112 分别运行操作系统 1 与操作系统 2，启动时从核 211 和启动时从核 212 组成处理核组 214，处理核组 214 运行同一个操作系统 3；启动时主核 111 对应内存区段 121，启动时从核 112 对应内存区段 122，处理核组 214 对应内存区段 221，该些信息均记录在运行信息中。

具体地，运行信息中记录的例如是以下的信息：（1）启动时主核 111 运行操作系统 1，为启动时主核 111 分配的内存区段 121 的大小和起始地址；（2）启动时从核 112 运行操作系统 2，启动时从核 112 的内存区段 122 的大小和起始地址；（3）处理核组 214 运行操作系统 3，处理核组 214 的内存区段 221 的大小和起始地址。在内存空间 22 中，内存区段 121、内存区段 122、内存区段 221 及共享内存段 123 相互之间，实现了内存区段隔离。

系统启动后，每个处理核处于自旋状态，自旋状态是指处理核不停空转，等待外部事件唤醒。启动加载器 1111 完成设备初始化之后；处理器 21 通过启动加载器 1111，根据运行信息中记录的各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址，使用启动加载模式或
5 下载模式，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理器的内存区段中，并将各处理核的运行时入口地址分别写入共享内存段 123 中。

在本实施例中，最先启动运行操作系统的处理核预设启动时主核 111，但并不以此为限，最先启动运行操作系统的处理核也可以预设处
10 理器 11 的其他处理核。

启动时主核 111 读取自身的运行时入口地址指针，检查是否有效物理地址，若是，则跳转到该运行时入口地址指针所指向的启动时主核 111 的运行时入口地址，运行操作系统 1 的内核镜像文件，从而运行操作系统 1。启动时主核 111 通过核间通信机制，唤醒待启动的处理核；各所述待启动
15 的处理核分别读取各自的运行时入口地址指针，检查是否有效物理地址，若是，则跳转到各自的运行时入口地址指针所指向的运行时入口地址，运行对应的操作系统的内核镜像文件，从而运行对应的操作系统。其中，各处理核在执行过程中，将跳转到运行信息地址读取运行信息，获知各自需要运行的操作系统、分配的内存区段的大小以及起始地址，从而，各处理核只
20 可以操作共享内存段 123，以及各自的内存区段。因此，本实施例的支持多核架构下资源隔离的系统 20，实现了在处理器的不同的处理核或处理核组上运行相互独立的操作系统，并且只在系统启动时，通过运行于启动时主核 111 上的启动加载器 1111 将各处理核或处理核组所需要运行的操作系统的内核镜像文件存储到各处理核或处理核组的内存区段中，之后，每个
25 处理核可以独立进行重启、休眠或关闭，而并不会影响其他处理核的正常运行，尤其在单个处理核或其操作系统失效时，也不会影响到其他处理核或在其上运行的操作系统，实现了核间操作系统隔离。

一种可选的实现场景为：处理核独立重启时，只需要复位，读取对应的运行时入口地址指针获取对应的运行时入口地址，运行存储在该处理核
30 的内存区段中的操作系统的内核镜像文件，其不用依赖于其他处理核的操

作系统。

另一种可选的实现场景为：在系统运行过程中，对于某一处理核，其他处理核可以根据系统需求修改运行信息，该处理核获知运行信息改变后进行复位，重新启动时，通过读取运行信息地址中的新的运行信息，获知新的运行信息，并通过读取运行时入口地址指针中的运行时入口地址，运行重新指定的操作系统。例如，在系统启动后的某一时刻，多核处理器的处理核 a 已经占用了一段内存区段，并在其上运行着操作系统 A，处理核 b 处于自旋状态；处理核 c 可以根据系统运行状态对运行信息进行修改，在修改后的运行信息中重新指定了：处理核 a 和处理核 b 构成处理核组 T、并在处理核组 T 上运行同一个操作系统 B、分配给处理核组 T 的内存区段的大小，以及内存区段的起始地址。处理核 a 获知运行信息改变后，可选的，处理核 a 首先将正在处理核 a 上运行的操作系统 A 卸载，并释放占用的内存区段，处理核 a 重新启动。处理核组 T（处理核 a 和处理核 b）根据修改后的运行信息，运行操作系统 B，并占用一段新的内存区段。

处理器 21 的各处理核启动后，各处理核或处理核组上运行着相互独立的操作系统，实现了内存区段隔离、核间操作系统隔离。鉴于多核处理器的 I/O 资源以及协处理器全局共享，当各处理核启动后，需要在处理器 21 的各处理核中选择一个运行时主核。所述运行时主核是指处理器 21 的各处理核启动后，能够作为各处理核集中控制和仲裁的处理核；所述运行时主核负责统一进行 I/O 资源的仲裁控制。具体地，运行时主核，是由处理器 21 的各处理核根据预设的选举算法在各处理核中选举确定的。如图 3 所示，假设处理器 21 的启动时主核 111、启动时从核 112、启动时从核 211 和启动时从核 212 均处于运行状态，那么这 4 个处于运行状态的处理核构成一个集群，在该集群内，这 4 个处理核通过核间通信，运行选举算法，例如 PAXOS 算法，选举确定运行时主核，例如选举结果为：启动时主核 111 被选举为运行时主核，其他的处理核为运行时从核，需要说明的是：集群通过多核协商选举出的运行时主核并不只限于启动时主核，其可能是任一处理核，本实施例只是示例性的给出启动时主核作为选举出的运行时主核，但是本发明并不以此为限。当选举出的运行时主核失效时，处理核的集群重新选举一个新的运行时主核，运行时主核的唯一性由所使用的选

举算法来保证，例如 PAXOS 算法。

运行时主核负责统一进行 I/O 资源的仲裁控制，所有的 I/O 资源以及协处理器资源使用都需要经过运行时主核的认可授权。

当运行时从核需要访问外部设备 23 时，运行时从核向运行时主核发送资源使用请求；运行时主核根据该资源使用请求，基于系统资源使用情况和资源分配策略，在排他性使用资源的裁决后，向所述运行时从核发送资源请求响应；所述运行时从核接收到所述资源请求响应后，获得 I/O 资源的使用授权，开始对外部设备 23 进行 I/O 操作。

运行时主核在系统资源使用情况和资源分配策略允许时可以调用自身的本地物理驱动，对外部设备进行 I/O 操作；各运行时从核使用 I/O 资源对外部设备 23 进行访问的实现方式，至少可以包括以下两种方式：

方式 1：运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

方式 2：运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

图 4a 为本发明实施例二提供的支持多核架构下资源隔离的系统的运行时从核对外部设备进行访问的一实现示意图；图 4b 为本发明实施例二提供的支持多核架构下资源隔离的系统的运行时从核对外部设备进行访问的另一实现示意图。

如图 4a 所示，运行时从核 402 将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知运行时主核 401，以使运行时主核 401 调用自身的本地物理驱动 4011，对外部设备 23 进行所述 I/O 操作。运行时从核 402 对外部设备 23 进行 I/O 操作，是通过运行时主核 401 代理完成。运行时从核 402 可以通过共享内存方式，将需要本次 I/O 操作的数据放入共享内存段，并将与所述 I/O 操作相关的内存地址通知运行时主核 401；运行时主核 401 接收运行时从核 402 发送的通知信息，调用本地物理驱动 4011 将数据通过 I/O 资源发给外部设备 23，进行访问，例如运行时主核 401 对外部设备 23 进行 I/O 操作，并在访问结束后释放 I/O 资源及外部设备 23 的相关资源。

如图 4b 所示, 运行时从核 402 调用自身的本地物理驱动 4021, 对外部设备 23 进行 I/O 操作。运行时从核 402 需要对外部设备 23 进行访问时, 通过控制仲裁方式, 由运行时主核 401 批准运行时从核 402 对 I/O 资源的使用; 运行时从核 402 根据运行时主核 401 发送的资源请求响应, 采用直通方式对为其分配的外部设备进行访问, 包括对外部设备 23 进行 I/O 的直通操作, 通过调用本地的物理驱动 4021 直接进行 I/O 资源的操作, 也即各处理核采用直通方式对外部设备 23 进行 I/O 操作, 操作完成后释放 I/O 资源的控制给运行时主核 401, 即, 通知运行时主核 401 已完成对外部设备 23 的 I/O 操作; 运行时主核 401 根据所述通知, 释放 I/O 资源及外部设备 23 的相关资源, 以便将资源释放给其他处理核使用。

可选的, 运行时主核与运行时从核之间维持状态同步信息。当运行时主核发现有某个处理核无法联系, 若此时, 运行时主核已经给该无法联系的处理核分配了外部设备资源, 则运行时主核将收回该外部设备资源。本发明上述实施例提供的支持多核架构下资源隔离的系统, 通过运行时主核负责统一进行 I/O 资源的仲裁控制, 实现了各个处理核使用 I/O 资源对外部设备进行访问。需要说明的是, 本发明上述实施例提供的支持多核架构下资源隔离的系统, 也用于实现各个处理核对公共功能部件进行访问, 公共功能部件至少可以包括协处理器资源。共享式的协处理器资源的仲裁应用的实现方式具体参见上述实施例的描述, 在此不再赘述。

本发明实施例提供的支持多核架构下资源隔离的系统, 通过在各处理核中选举运行时主核来负责统一进行 I/O 资源的仲裁控制, 实现 I/O 资源的隔离。在各处理核或处理核组上的运行的操作系统互相独立运行的情况下, 实现了多核处理器中统一 I/O 能力以及协处理器的调度控制, 保证了多核处理器的故障域维持在单块硬盘。

实施例三

图 5 为本发明实施例三提供的支持多核架构下资源隔离的方法的流程图。本实施例的方法应用于包括处理器和内存空间的支持多核架构下资源隔离的系统, 其中: 所述处理器包括多个处理核, 所述多个处理核中包括一个启动时主核以及至少一个启动时从核; 所述内存空间包括共享内存段, 以及各处理核的内存区段; 所述共享内存段中存储有运行信息, 以及

各处理核的运行时入口地址；所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像文件在各自的内存区段中的存储地址。如图 5 所示，本实施例的方法包括：

5 501、处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址分别写入所述共享内存段中。

502、待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

10 可选的，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行输入/输出 I/O 资源的仲裁控制。

可选的，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

15 可选的，运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

可选的，所述启动时主核内设置有启动加载器；所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：所述启动加载器根据所述运行信息，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中；

20 可选的，所述启动时主核内设置有启动加载器；所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：所述启动加载器根据所述运行信息，在所述处理器的至少一个处理核内存区段中存储所需要运行的操作系统的内核镜像文件，运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据所述运行信息，将所述处理器的需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到各所述其余处理核的内存区段中。

25 可选的，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件，并且，所述构成所述处理核组的所有处理核共用同一段内存

区段；或者，各处理核运行互不相同的操作系统内核镜像文件。

5 本发明实施例提供的支持多核架构下资源隔离的方法，通过核间操作系统隔离、内存区段隔离、I/O 资源隔离的方式，使得在多核处理器的不同处理核上运行的操作系统能够独立运行而互不影响，从而，充分利用了多核处理器高集成度、综合成本低的优势，实现了多核处理器的故障域维持在单块硬盘，具有较高的可靠性。

10 本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

15 最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

权 利 要 求 书

1、一种支持多核架构下资源隔离的系统，其特征在于，包括：

处理器，所述处理器包括多个处理核，所述多个处理核中包括一个启动时主核以及至少一个启动时从核；

5 内存空间，所述内存空间包括共享内存段，以及各处理核的内存区段；所述共享内存段中存储有运行信息，以及各处理核的运行时入口地址；所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像文件在各自对应的内存区段中的存储地址；
10 址；

其中，所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址分别写入所述共享内存段中；待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

15 2、根据权利要求 1 所述的支持多核架构下资源隔离的系统，其特征在于，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行输入/输出 I/O 资源的仲裁控制。

3、根据权利要求 2 所述的支持多核架构下资源隔离的系统，其特征在于，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。
20 述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

4、根据权利要求 2 所述的支持多核架构下资源隔离的系统，其特征在于，运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

5、根据权利要求 1-4 任一项所述的支持多核架构下资源隔离的系统，
25 其特征在于，所述启动时主核内设置有启动加载器；

所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

所述启动加载器根据所述运行信息，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中。

30 6、根据权利要求 1-4 任一项所述的支持多核架构下资源隔离的系统，

其特征在于，所述启动时主核内设置有启动加载器；

所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

所述启动加载器根据所述运行信息，在将所述处理器的至少一个处理核的内存区段中存储所需要运行的操作系统的内核镜像文件存储到对应的内存区段中，运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据所述运行信息，将所述处理器的需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到各所述其余处理核对应的内存区段中。

7、根据权利要求 1-6 任一项所述的支持多核架构下资源隔离的系统，其特征在于，各处理核运行互不相同的操作系统的内核镜像文件；或者，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件，并且，所述构成所述处理核组的所有处理核共用同一段内存区段。

8、一种支持多核架构下资源隔离的方法，其特征在于，应用于包括处理器和内存空间的支持多核架构下资源隔离的系统，其中：所述处理器包括多个处理核，所述多个处理核中包括一个启动时主核以及至少一个启动时从核；所述内存空间包括共享内存段，以及各处理核的内存区段；所述共享内存段中存储有运行信息，以及各处理核的运行时入口地址；所述运行信息包括各处理核与各操作系统之间的对应关系、各内存区段的大小以及起始地址；所述各处理核的运行时入口地址，用于指示各处理核所需要运行的操作系统的内核镜像文件在各自的内存区段中的存储地址；

所述支持多核架构下资源隔离的方法包括：

所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，并将各处理核的运行时入口地址分别写入所述共享内存段中；

待启动的处理核分别获取各自的运行时入口地址，并运行存储在各自的内存区段中的操作系统的内核镜像文件。

9、根据权利要求 8 所述的支持多核架构下资源隔离的方法，其特征在于，所述处理器的各处理核根据预设的选举算法在各处理核中选举确定运行时主核；所述运行时主核，用于进行输入/输出 I/O 资源的仲裁控制。

10、根据权利要求 9 所述的支持多核架构下资源隔离的方法，其特征在于，运行时从核将需要 I/O 操作的数据放入所述共享内存段，并将与所述 I/O 操作相关的内存地址通知所述运行时主核，以使所述运行时主核调用自身的本地物理驱动，对外部设备进行所述 I/O 操作。

5 11、根据权利要求 9 所述的支持多核架构下资源隔离的方法，其特征在于，运行时从核调用自身的本地物理驱动，对外部设备进行 I/O 操作。

12、根据权利要求 8-11 任一项所述的支持多核架构下资源隔离的方法，其特征在于，所述启动时主核内设置有启动加载器；

10 所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

所述启动加载器根据所述运行信息，将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中。

13、根据权利要求 8-11 任一项所述的支持多核架构下资源隔离的方法，其特征在于，所述启动时主核内设置有启动加载器；

15 所述处理器将各处理核所需要运行的操作系统的内核镜像文件分别存储到各处理核的内存区段中，包括：

20 所述启动加载器根据所述运行信息，在所述处理器的至少一个处理核的内存区段中存储所需要运行的操作系统的内核镜像文件，运行所述操作系统的内核镜像文件；所述运行操作系统的处理核根据所述运行信息，将所述处理器的需要运行操作系统的其余处理核所对应的操作系统的内核镜像文件分别存储到各所述其余处理核的内存区段中。

25 14、根据权利要求 8-13 任一项所述的支持多核架构下资源隔离的方法，其特征在于，各处理核运行互不相同的操作系统的内核镜像文件；或者，由至少两个处理核构成的处理核组运行相同的操作系统的内核镜像文件，并且，所述构成所述处理核组的所有处理核共用同一段内存区段。

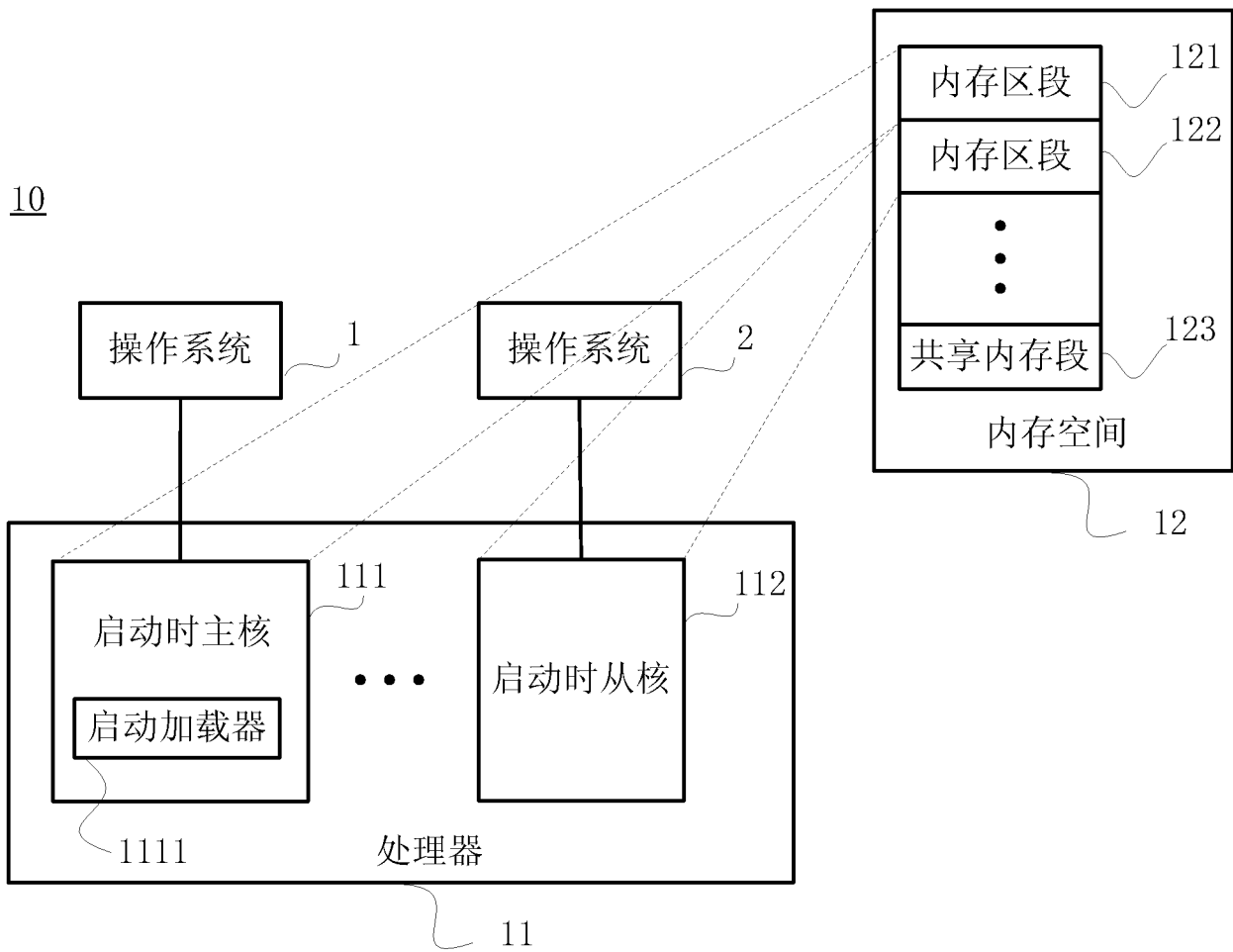


图 1a

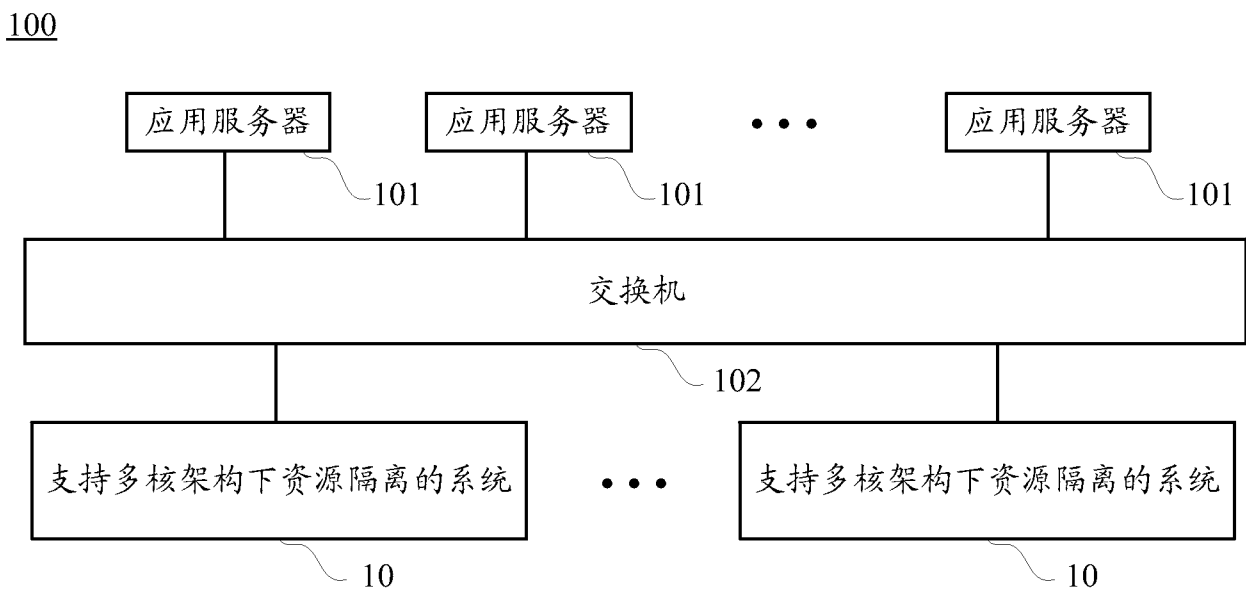


图 1b

2/4

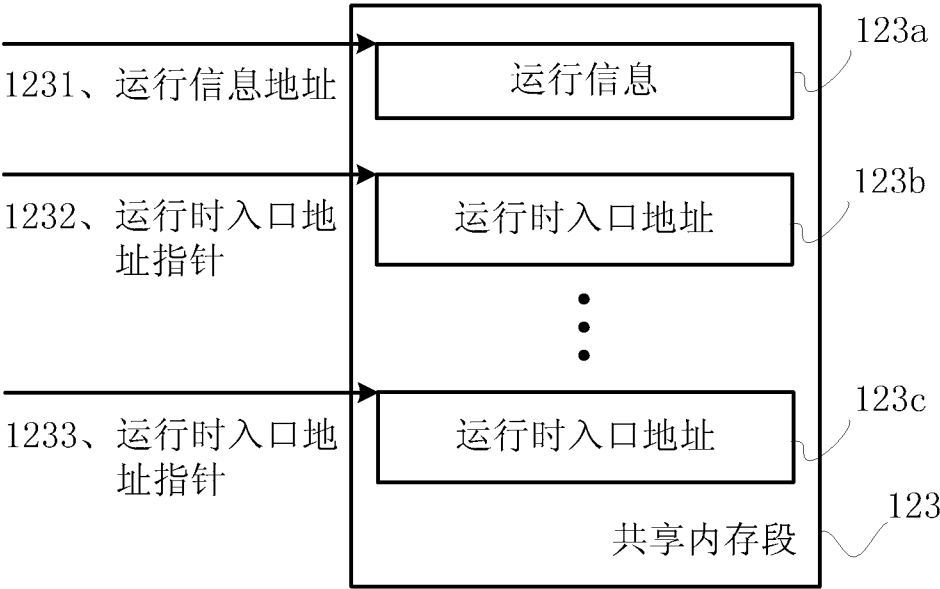


图 2

20

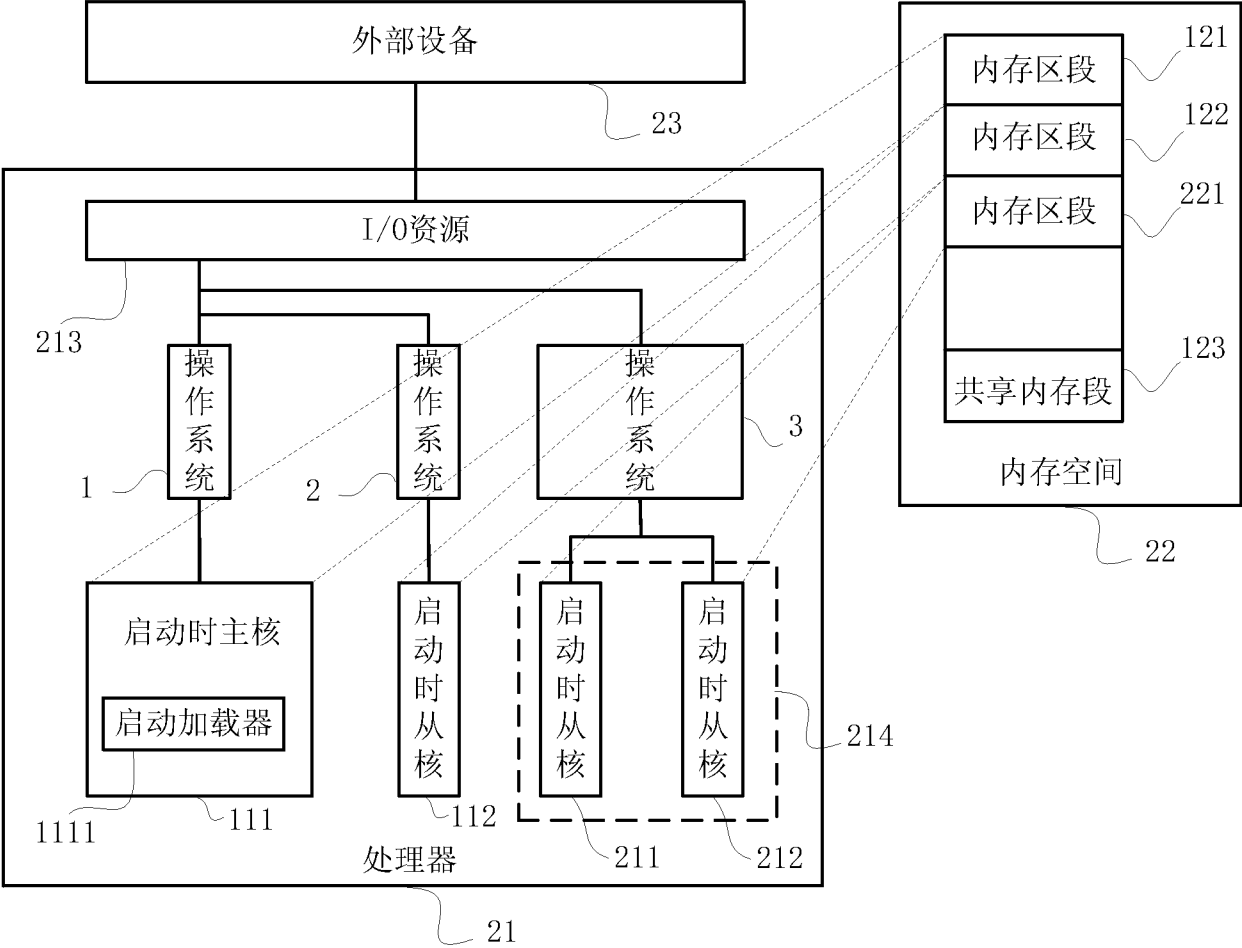


图 3

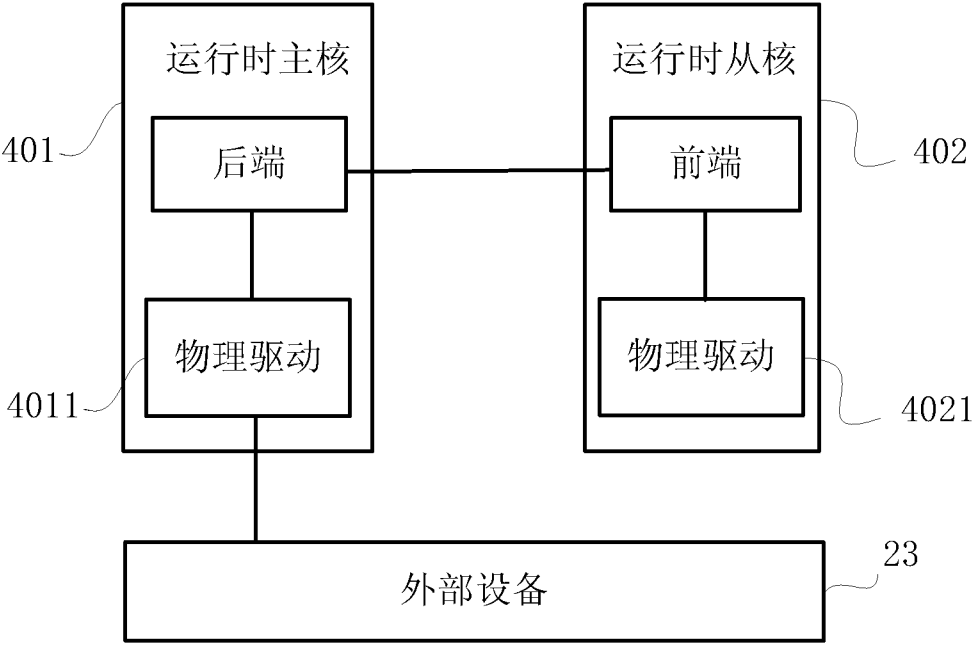


图 4a

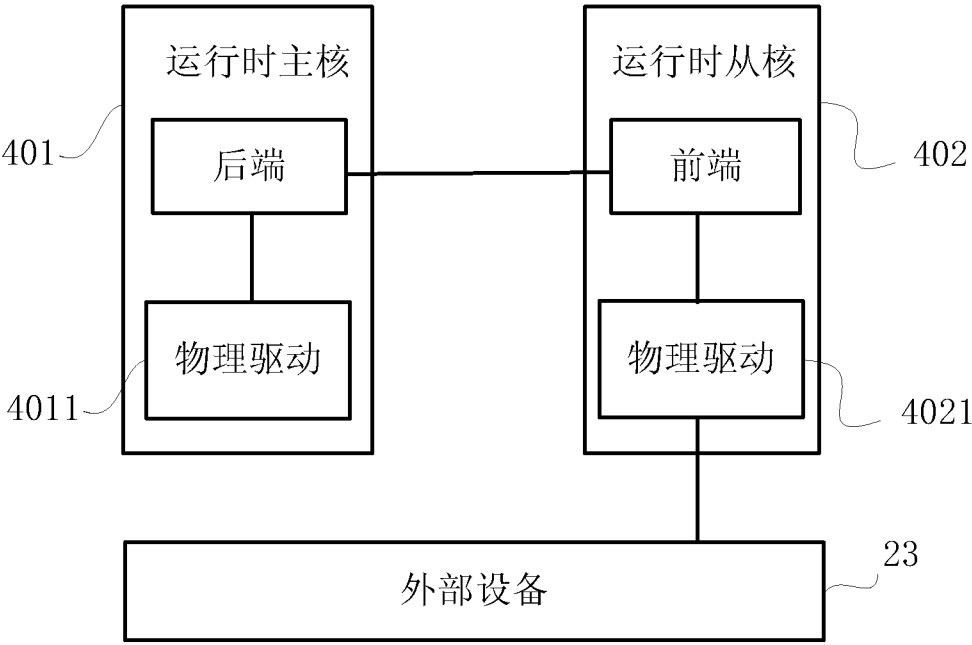


图 4b

4/4

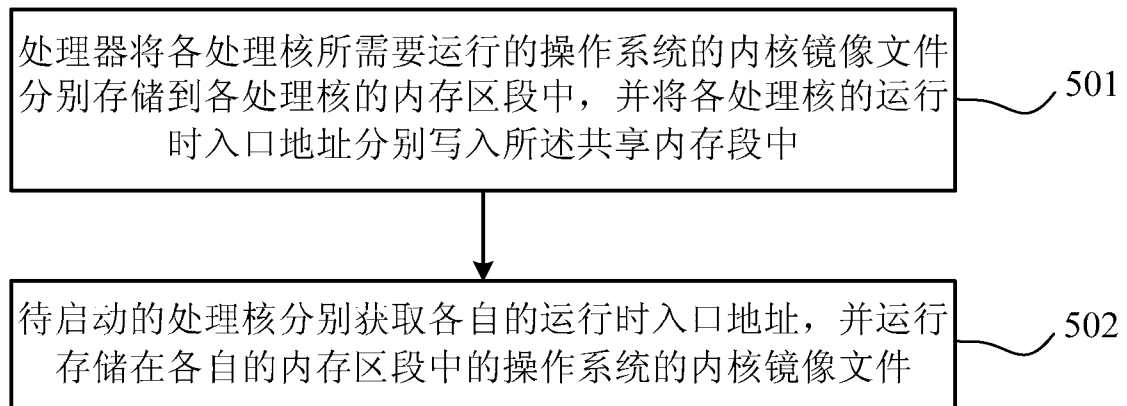


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/076329

A. CLASSIFICATION OF SUBJECT MATTER

G06F 15/167 (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)

IPC: G06F 15/-

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, DWPI, CNKI: multi, processor, kernel, share, memory, cache, image file, OS, resource, isolat???

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103020002 A (THE PLA INFORMATION ENGINEERING UNIVERSITY) 03 April 2013 (03.04.2013) description, paragraphs [0026]-[0037]	1-14
A	CN 102541805 A (SHENYANG GAOJING NUMERICAL CONTROL TECHNOLOGY CO., LTD.) 04 July 2012 (04.07.2012) the whole document	1-14
A	CN 101149728 A (INSTITUTE OF COMPUTING TECHNOLOGY) 26 March 2008 (26.03.2008) the whole document	1-14
A	CN 103092788 A (HUAWEI TECHNOLOGIES CO., LTD.) 08 May 2013 (08.05.2013) the whole document	1-14

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 28 January 2014 (28.01.2014)	Date of mailing of the international search report 13 March 2014 (13.03.2014)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer LI, Yan Telephone No. (86-10) 82245955

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/CN2013/076329

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103020002 A	03.04.2013	None	
CN 102541805 A	04.07.2012	None	
CN 101149728 A	26.03.2008	None	
CN 103092788 A	08.05.2013	None	

国际检索报告

国际申请号

PCT/CN2013/076329

A. 主题的分类

G06F 15/167 (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)

IPC: G06F 15/-

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, DWPI, CNKI, 多, 处理器, 内核, 共享, 内存, 镜像文件, 操作系统, 资源, 隔离
multi, processor, kernel, share, memory, cache, image file, OS, resource, isolat???

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	CN 103020002 A (中国人民解放军信息工程大学) 03.4 月 2013(03.04.2013) 说明书第[0026]-[0037]段	1-14
A	CN 102541805 A (沈阳高精数控技术有限公司) 04.7 月 2012 (04.07.2012) 全文	1-14
A	CN 101149728 A (中国科学院计算技术研究所) 26.3 月 2008 (26.03.2008) 全文	1-14
A	CN 103092788 A (华为技术有限公司) 08.5 月 2013 (08.05.2013) 全文	1-14

☐ 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

28.1 月 2014 (28.01.2014)

国际检索报告邮寄日期

13.3 月 2014 (13.03.2014)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局
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

李琰

电话号码: (86-10) 82245955

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2013/076329

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 103020002 A	03.04.2013	无	
CN 102541805 A	04.07.2012	无	
CN 101149728 A	26.03.2008	无	
CN 103092788 A	08.05.2013	无	