

(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
2017 年 7 月 13 日 (13.07.2017)



(10) 国际公布号
WO 2017/117734 A1

- (51) 国际专利分类号:
G06F 3/06 (2006.01)
- (21) 国际申请号: PCT/CN2016/070230
- (22) 国际申请日: 2016 年 1 月 6 日 (06.01.2016)
- (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: 徐君 (XU, Jun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。黄永兵 (HUANG, Yongbing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。王元钢 (WANG, Yuangang); 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: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市国贸大厦 15 楼西座 1521 室, Guangdong 518014 (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, IR, IS, JP, KE, KG, KN, KP, KR, 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)。

[见续页]

(54) Title: CACHE MANAGEMENT METHOD, CACHE CONTROLLER AND COMPUTER SYSTEM

(54) 发明名称: 一种缓存管理方法、缓存控制器以及计算机系统

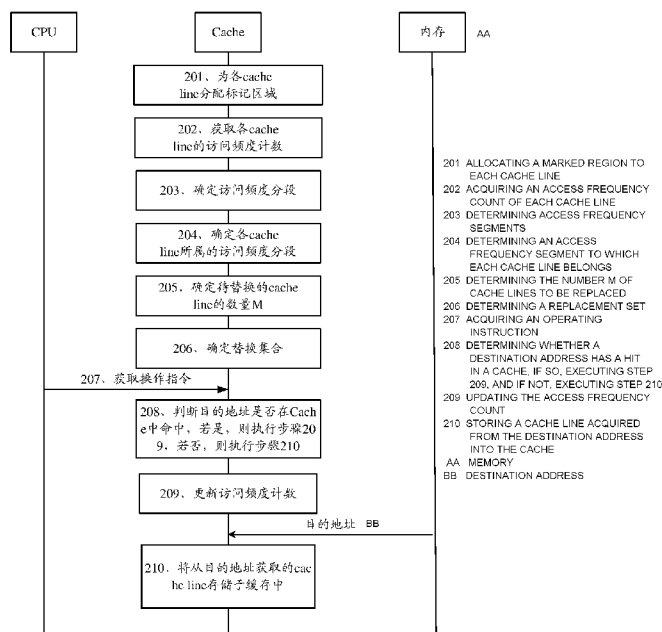


图 2

(57) Abstract: A cache management method, a cache controller and a computer system. In the method, the cache controller acquiring an operating instruction; when a destination address in the operating instruction does not hit any cache line in a cache of the computer system and the cache does not comprise an idle cache line, the cache controller selecting a cache line to be replaced from a replacement set, wherein the replacement set contains at least two cache lines; and the cache controller eliminating the cache line to be replaced from the cache and storing a cache line acquired from the destination address in the cache. By means of the present cache management method, system overheads can be reduced when a cache line is replaced, and the replacement efficiency of the cache line is improved.

(57) 摘要: 一种缓存管理方法、缓存控制器以及计算机系统。在该方法中, 缓存控制器获取操作指令, 当操作指令中的目的地址未命中计算机系统的缓存中的任意一个缓存行 cache line, 且缓存中未包括空闲的 cache line 时, 缓存控制器从替换集合中选择待替换的 cache line。其中, 替换集合中包含有至少两个 cache line。缓存控制器从缓存中淘汰待替换的 cache line 并将目的地址获取的 cache line 存储于缓存中。本缓存管理方法能够减少 cache line 替换时的系统开销, 提高 cache line 替换的效率。

WO 2017/117734 A1

根据细则 4.17 的声明:

— 关于申请人有权申请并被授予专利(细则 4.17(ii))

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种缓存管理方法、缓存控制器以及计算机系统

技术领域

5 本发明涉及通信领域，尤其涉及一种缓存管理方法、缓存控制器以及计算机系统。

背景技术

10 缓存（Cache）技术的主要思想是将常用的数据由存储介质（如磁盘）放置在 Cache 中，由于 Cache 的读写速度远高于磁盘，所以能够提高整个系统的访问效率。

一般而言，Cache 的性能相对较好，但是容量有限，仅能缓存部分被访问的数据。为了让 Cache 发挥出更好的效率，需要尽量让即将发生的读写操作涉及到的文件在 Cache 里命中。如何选择缓存在 Cache 中的数据（Cache 插入），如何选择替换出 Cache 的数据（Cache 替换），是 Cache 管理的两个主要方面。

15 现有技术提供了一种基于统计计数的 Cache 管理算法，为每个缓存行 cache line 增加一个计数器，用于统计一定时间间隔内访问 Cache line 的访问请求数，也就是用于统计 cache line 的访问次数。基于访问计数器，现有技术实现了两种 cache line 替换算法。第一种算法的时间间隔定义为相邻两次访问该 cache line；当访问计数器超过一定的阈值 Δthd 后，该 cache line 作为被替换目标。
20 第二种算法的时间间隔定义为自 cache line 被加入到 Cache 中开始到当前时间；同样，当访问计数器超过一定阈值 LT 后，该 cache line 作为被替换目标。

现有技术中，如果当前需要读写的数据未在 Cache 里命中 cache line，则控制器会从存储介质中提取出该数据，作为一个新的 cache line 准备放入 Cache
25 中，如果此时 Cache 中存储的 cache line 的数目已经达到了上限，则需要从原有的 cache line 中删除一个。

控制器会遍历原有的所有 cache line 的计数器，以获知每一个 cache line 的访问请求数或访问时间间隔，从而选取出一个访问请求数最少，或者是访问时间间隔最长的 cache line，并将该 cache line 删除，同时将新的 cache line 放
30 入 Cache 中。

因此，现有技术中，在进行 cache line 替换的过程中，控制器需要遍历所有的原有 cache line，并依据一定的计算规则选择一个待替换的 cache line，从而会造成较大的系统开销，并影响了 cache line 替换的效率。

5 发明内容

本发明实施例提供了一种缓存管理方法、缓存控制器以及计算机系统，能够有效减少 cache line 替换时的系统开销，提高 cache line 替换的效率。

本发明实施例第一方面提供了一种应用于计算机系统内的缓存管理方法。在该方法中，缓存控制器获取 CPU 发送的操作指令，其中，该操作指令中携带有目的地址，该目的地址为 CPU 待访问的内存中的地址。如果此时缓存控制器未在计算机系统的缓存中的任意一个缓存行 cache line 中找到与该目的地址匹配的 cache line，即该目的地址未命中，且此时在缓存中没有空闲的 cache line 存在，则缓存控制器从一个预先获得的替换集合中选择待替换的 cache line 进行替换，这里需要说明的是，该替换集合中包含有至少两个 cache line。缓存控制器将选择出的待替换的 cache line 从缓存中淘汰。之后，缓存控制器将从目的地址获取的 cache line 存储入缓存中以完成 cache line 的替换。

在本发明的实施例中，缓存控制器在目的地址未命中时，可以直接从替换集合中挑选出待替换的 cache line 进行替换，而与现有技术的每次未命中都需要遍历一遍各 cache line 找出替换 cache line 不同，有效减少了 cache line 替换时的系统开销，提高了效率。

在一个可能的设计中，缓存控制器在从替换集合中选择待替换的 cache line 之前，缓存控制器可以预先获取缓存中各 cache line 的访问频度计数，并依据预设的划分策略 cache line 的访问频度计数范围划分为多个访问频度分段。在依据各 cache line 的访问频度计数分别确定各 cache line 所属的访问频度分段之后，缓存控制器可以按照待替换的 cache line 的数量 M 从多个访问频度分段对应的 cache line 中选择待替换的 cache line，以得到替换集合。其中，M 为不小于 2 的整数。

可选的，本发明实施例中，缓存控制器也可以不划分多个访问频度分段，而直接从缓存中的各 cache line 中挑选出 M 个待替换的 cache line 得到替换集合。

本发明实施例中，缓存控制器将各 cache line 划分成多个分段，可以方便缓存控制器以段为单位选择所要替换的 cache line。避免在选择待替换的 cache line 时，需要大批量的一个个比较 cache line 的访问频度计数，减少工作开销。

在一个可能的设计中，缓存控制器按照待替换的 cache line 的数量 M 从多个访问频度分段对应的 cache line 中选择待替换的 cache line 包括：缓存控制器按照多个访问频度分段对应的访问频度计数范围从小到大的顺序，依次选择属于各访问频度分段中的 cache line，直至选择的 cache line 的数量等于 M 。

本发明实施例中，cache line 的访问频度计数越低代表该 cache line 的被访问次数越少，最近被访问的时间越晚，可以优先被替换。缓存控制器按照访问频度计数范围从小到大的顺序，依次选择属于各访问频度分段中的 cache line，可以选出最优的替换集合。

在一个可能的设计中，待替换的 cache line 的数目 M 是根据淘汰比例 R 与缓存中 Cache line 的总数确定的，其中， M 为淘汰比例 R 与 Cache line 总数的乘积。

在又一个可能的设计中，缓存控制器还可以周期性监控淘汰频率参数，该淘汰频率参数包括缓存中各 cache line 的缺失率和被遍历频率中的至少一个。当淘汰频率参数超过第一阈值时，此时缓存控制器将淘汰比例 R 调整为淘汰比例 R_1 ，该淘汰比例 R_1 会大于淘汰比例 R 。当淘汰频率参数小于第二阈值时，此时缓存控制器将淘汰比例 R 调整为淘汰比例 R_2 ，该淘汰比例 R_2 会小于淘汰比例 R ，其中，此处的第二阈值会小于第一阈值。

本发明实施例中，缓存控制器依据 cache line 的缺失率，即高未命中率，以及 cache line 的被遍历率动态调整淘汰比例 R ，以达到在 cache line 在高缺失率以及频繁被遍历的情况下，通过提高淘汰比例 R 来增加替换集合中待替换的 cache line 的数量，同时，还可以在 cache line 低缺失率以及较少被遍历的情况下，降低淘汰比例 R ，以减少替换集合中的待替换的 cache line 的数量，避免了频繁挑选 cache line 生成替换集合的开销。

在一个可能的设计中，缓存控制器从替换集合中选择待替换的 cache line 包括：缓存控制器按照替换集合中 cache line 的访问频度计数从小到大的顺序选择出待替换的 cache line，以达到最优的 cache line 的利用率。

在又一个可能的设计中，在获得替换集合之后，缓存控制器还可以监控属

于替换集合的 cache line 的访问频度计数，并从替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。换一种表达方式，在获得替换集合之后，缓存控制器可以继续监控替换集合中的 cache line 被命中的次数，若在预设时间内替换集合中的某个 cache line 的访问频度计数大于第三阈值时，缓存控制器可以认为该 cache line 被后续接收的操作指令中的目的地址命中，可以不淘汰。通过这种方式，可以更新替换集合中的 cache line，避免部分被访问的 cache line 被误淘汰。

在一个可能的设计中，缓存控制器可以根据各 cache line 的访问次数获取的各 cache line 的访问频度计数。缓存控制器也可以根据各 cache line 的访问次数以及访问时间获得各 cache line 的访问频度计数。

本发明第二方面提供了一种缓存控制器，该缓存控制器包括用于执行上述第一方面以及第一方面的各种可能的设计中所述的方法的模块。

本发明第三方面提供了一种计算机系统，该计算机系统包括处理器、缓存以及缓存控制器。该处理器用于发送操作指令。缓存控制器用于用于执行上述第一方面以及第一方面的各种可能的设计中所述的方法

本发明实施例中，缓存控制器获取操作指令，其中，操作指令中携带有目的地址，目的地址为操作指令待访问的内存中的地址，当目的地址未命中计算机系统的缓存中的任意一个缓存行 cache line，且缓存中未包括空闲的 cache line 时，则缓存控制器从替换集合中选择待替换的 cache line，其中，替换集合中包含有至少两个 cache line，缓存控制器从缓存中淘汰待替换的 cache line，缓存控制器将从目的地址获取的 cache line 存储于缓存中，因而在进行 cache line 替换的过程中，缓存控制器只需要从替换集合中挑选出待替换的 cache line 即可，而替换集合是预先选出的，有效提高了 cache line 替换的效率。

第四方面，本申请提供了一种计算机程序产品，包括存储了程序代码的计算机可读存储介质，所述程序代码包括的指令用于执行上述第一方面以及第一方面的各种可能的设计中所述的至少一种方法。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍。显而易见地，下面描述中的附图仅仅是本发明

的一些实施例。

图 1 为本发明实施例提供的一种计算机系统框架图；

图 2 为本发明实施例提供的一种缓存管理方法示意图；

图 3 为本发明实施例提供的一种缓存管理方法另示意图；

5 图 4 为本发明实施例提供的一种缓存控制器示意图；

图 5 为本发明实施例提供的一种缓存控制器示意图；

图 6 为本发明实施例提供的又一种计算机系统架构示意图。

具体实施方式

10 本发明实施例提供了一种缓存管理方法、缓存控制器以及计算机系统，能够有效减少 cache line 替换时的系统开销，提高 cache line 替换的效率。

本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等（如果存在）是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里
15 描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。

请参阅图 1，图 1 为本发明实施例提供的一种计算机系统架构图。如图 1 所示，Cache 是介于 CPU 和主存之间的小容量存储器，存取速度比主存快，接近 CPU。它能高速地向 CPU 提供指令和数据，提高程序的执行速度。Cache 技术是为了解决 CPU 和主存之间速度不匹配而采用的一项重要技术。

20 Cache 是主存的缓冲存储器，由高速的静态随机存取存储器 SRAM 组成，Cache 可以内置于中央处理器 CPU 中，也可以内置于内存 Memory 上，Cache 还可以以一个单独存在的实体外置于 CPU 与内存之间。Cache 的所有控制逻辑全部由内部缓存控制器实现。随着半导体器件集成度的不断提高，当前已出现了两级以上的多级 Cache 系统，如 L1 Cache（一级缓存）、L2 Cache（二级
25 缓存）以及 L3 Cache（三级级缓存）。

当前 Cache 的容量很小，它保存的内容只是主存内容的一个子集，且 CPU 与 Cache 之间的数据交换是以字（word）为单位的，而 Cache 与主存之间的数据交换则是以缓存行（cache line）为单位的。一个 cache line 由若干个定长的 word 组成。当 CPU 读取主存中的一个字时，该字的主存地址被发给 Cache 和
30 主存，此时，Cache 控制逻辑依据地址判断该字当前是否存在于 Cache 中：若

在, 该字立即被从 Cache 传送给 CPU; 若不在, 则用主存读周期把该字从主存读出送到 CPU, 同时把含有这个字的整个 cache line 从主存读出送到 Cache 中, 并采用一定的替换策略将 Cache 中的某一 cache line 替换掉, 替换算法由 Cache 管理逻辑电路来实现。

5 现有技术中, 在替换 cache line 的时候, 控制器会遍历原有的所有 cache line, 以获知每一个 cache line 的访问请求数或访问时间间隔, 从而选取出一个访问请求数最少, 或者是访问时间间隔最长的 cache line, 并将该 cache line 删除, 同时将新的 cache line 放入 Cache 中。现有技术中替换 cache line 时的系统开销较大, 且效率不高。

10 本发明实施例提供了一种缓存管理方法, 在本发明实施例提供的缓存管理方法中, 在每次替换 cache line 的过程中, 不需要遍历原有的所有 cache line, 也不需要计算选择待替换的 cache line, 从而能够有效的节省系统开销, 提高 cache line 替换的效率。为便于理解, 下面对本发明实施例中的具体流程对图 1 所示的计算机系统的缓存管理方法进行详细描述。可以理解的是, 图 2 所述的
15 缓存管理方法可以应用于图 1 中所示的 cache 中。请参阅图 2, 本发明实施例中缓存管理方法一个实施例包括:

201、缓存控制器为各 cache line 分配标记区域;

 本实施例中, 缓存控制器可以为每个 cache line 分配一个数据区域和标记区域, 该数据区域可以用于存放真正的文件数据, 该标记区域可以存放上述数
20 据区域所对应的文件号、偏移、长度信息, 或者数据地址信息, 具体此处不做限定。

 需要说明的是, 上述数据区域的每一个区域的存储空间大小都可以动态变化或者可以静态不动, 在实际应用中, 为了实现数据区域的存储空间大小动态变化, 缓存控制器可以在标记区域记录该数据区域的有效字符长度, 具体此处
25 不做限定。

 本实施例中, 缓存控制器可以为每个标记区域增加访问次数和访问时间计数器, 该访问次数计数器可以用于统计 cache line 被访问的次数, 该访问时间计数器可以用于记录 cache line 最近被访问的时间, 该时间可以用绝对时间值来标识, 也可以用时间周期数来标识, 具体此处不做限定。

30 需要说明的是, 缓存控制器可以为每个 cache line 统计访问次数, 还可以

同时为每个 cache line 统计访问次数和访问时间，具体此处不做限定。

本发明实施例中，Cache 可以内置于 CPU，也可以内置于内存，还可以为单独存在的实体，该实体可以包括缓存控制器和 cache line，该 cache line 的组成可以包括静态随机存取存储器 SRAM，具体此处不做限定。

5 202、缓存控制器获取各 cache line 的访问频度计数；

本实施例中，缓存控制器可以通过检测各 cache line 的标记区域得到各 cache line 的访问频度计数，该访问频度计数可以包含访问次数或访问时间中的至少一种，具体此处不做限定。例如，在一种情形下，各 cache line 的访问频度计数可以根据各 cache line 的访问次数获得。在这种情况下，可以将 cache line 的访问次数作为该 cache line 的访问频度计数。在又一种情形下，各 cache line 的访问频度计数还可以根据各 cache line 的访问次数以及访问时间获得。例如，在实际应用中，缓存控制器可以通过访问时间来调整访问频度计数，其一种实现方式可以是：缓存控制器将访问时间对应到时间片，其中一个时间片包含一小段时间，计算当前时间片与 cache line 最近访问时间所在的时间片的差值，根据时间片的差值来调整访问频度计数（例如：访问频度计数=当前访问次数-时间片差值*Delta，该 Delta 值可以根据经验值静态设定）。

需要说明的是，缓存控制器可以在各 cache line 被访问之后更新该 cache line 的访问次数或访问时间，在实际应用中，更新访问时间的的方式可以是记录当前访问时间，还可以是将上述访问时间计数器归零并重新开始计数，具体此处不做限定。

203、缓存控制器确定访问频度分段；

本实施例中，缓存控制器可以根据缓存中各 cache line 的访问频度计数以及预设的划分策略确定多个访问频度分段，每个访问频度分段可以对应不同的访问频度计数范围。

本实施例中，该预设的划分策略可以是缓存控制器先基于各 cache line 的访问频度计数确定参数上限值和参数下限值，该参数下限值可以为 0，该参数上限值可以为 access_times_max，需要说明的是，在实际应用中，该 access_times_max 可以为用户根据经验预先设置的值，还可以为缓存控制器通过统计各 cache line 的访问频度计数获得，具体此处不做限定。

需要说明的是，在本实施例中，若 cache line 的访问频度计数大于该

access_times_max, 则缓存控制器可以判定该 cache line 不需要被替换。

本实施例中预设的划分策略, 缓存控制器可以在确定参数上限和参数下限后将参数下限值 0 到参数上限值 access_times_max 的范围划分为 N 个访问频度分段, 该数值 N 可以为用户预先输入的数值, 可以理解的是, 在实际应用中, 每个访问频度分段的大小可以为定值 $\text{access\_times\_max}/N$, 还可以为随机值, 该随机值可以为缓存控制器随机生成的数值, 还可以为用户预先输入的数值, 具体此处不做限定。

需要说明的是, 在实际应用中, 每个访问频度分段的大小总和等于 access_times_max。

本实施例中, 缓存控制器还可以建立一个数据结构, 该数据结构用来记录每个访问频度分段内 cache line 的访问频度计数属于该分段的 cache line 的数量, 还可以为每个访问频度分段对应建立一个数据结构, 具体此处不做限定。

204、缓存控制器确定各 cache line 所属的访问频度分段;

本实施例中, 缓存控制器可以通过遍历各 cache line 确定各 cache line 的访问频度计数所属的目标访问频度计数范围, 确定各 cache line 所属目标访问频度计数范围对应的目标访问频度分段, 需要说明的是, 缓存控制器可以通过上述数据结构记录各访问频度分段内所包含的 cache line 数量来统计各访问频度分段的 cache line 数量。

需要说明的是, 在实际应用中, 缓存控制器每判断一个 cache line 属于一个访问频度分段, 可以对应调整该访问频度分段对应的数据结构, 可以理解的是, 在实际应用中, 缓存控制器每判断一个 cache line 属于一个访问频度分段, 该访问频度分段对应的数据结构可以增加数值 1。

205、缓存控制器确定待替换的 cache line 的数量 M;

本实施例中, 缓存控制器可以根据淘汰比例 R 预先选择 M 个可以待替换的 cache line, 该待替换的 cache line 可以为访问频度计数低的 cache line, 该待替换的 cache line 的数目 M 可以是淘汰比例 R 与 cache line 总数的乘积, 且 M 为不小于 2 的整数, 需要说明的是, 在实际应用中, 该淘汰比例 R 可以为定值, 也可以为缓存控制器对其动态调整, 具体此处不做限定。

本实施例中, 缓存控制器可以通过实时监控淘汰频率参数, 判断该淘汰频率参数是否超过了第一阈值, 若是, 缓存控制器可以将淘汰比例 R 调整为淘

淘汰比例 $R1$ ，若否，则缓存控制器可以进一步判断该淘汰频率参数是否小于第二阈值，若小于，则缓存控制器可以将淘汰比例 R 调整为淘汰比例 $R2$ ，需要说明的是，若不小于，则缓存控制器可以继续监控淘汰频率参数，该第一阈值和第二阈值可以为用户预先设置的值，该第一阈值小于第二阈值。

5 需要说明的是，在实际应用中，缓存控制器可以将淘汰比例 R 增加 A 得到淘汰比例 $R1$ ，可以将淘汰比例 R 减少 A 得到淘汰比例 $R2$ ，该数值 A 可以为用户预先设置的值。

本实施例中，该淘汰频率参数可以包括缓存中各 cache line 的缺失率和被遍历频率中的至少一个，且缓存控制器可以在每次调整完淘汰比例 R 后返回
10 继续监控淘汰频率参数。

206、缓存控制器确定替换集合；

本实施例中，缓存控制器可以按照访问频度计数范围对应的数值从小到大的顺序，依次将各访问频度分段中的 cache line 累计入替换集合，直至该替换集合中 cache line 的数目等于 M ，可以理解的是，在实际应用中，缓存控制器
15 累计各访问频度分段中的 cache line 可以包含以下步骤：

1) 缓存控制器可以将访问频度分段组中访问频度计数范围对应的数值最低的访问频度分段确定为累计访问频度分段，该访问频度分段组包含各访问频度分段；

2) 缓存控制器可以从访问频度分段组中去除累计访问频度分段；

20 3) 缓存控制器可以将该累计访问频度分段中的 cache line 累计入替换集合；

4) 缓存控制器可以判断替换集合中的候选 cache line 的数目是否等于 M ，若等于 M ，则可以结束流程，若小于 M ，则可以判断下一个将要累计入替换集合的访问频度分段是否为临界访问频度分段，若不是，则可以重复执行步骤
25 1) 至 4)，若是，则可以从该临界访问频度分段中随机选取 X 个 cache line 累计入替换集合，该 X 可以为 M 与已累计入替换集合的 cache line 的数目的差值，需要说明的是，该临界访问频度分段为所包含的 cache line 累计入替换集合之前，替换集合中的候选 cache line 小于 M ，且所包含的 cache line 累计入替换集合之后，替换集合中的候选 cache line 大于 M 的访问频度分段。

30 需要说明的是，若当前替换集合中累计值不超过 M ，则缓存控制器可以

继续遍历各访问频度分段，直至累计值大于或等于 M。

需要说明的是，在实际应用中，缓存控制器可以先将各访问频度分段累计入被淘汰的范围段集合，再可以通过遍历各 cache line，选出属于被淘汰的访问频度分段的 cache line，缓存控制器可以通过将该选出的 cache line 作为替换集合 cache line 来确定替换集合。

207、缓存控制器获取操作指令；

缓存控制器可以接收 CPU 发送的操作指令，该操作指令中可以携带有目的地址，该目的地址可以为操作指令待访问的内存中的地址。

208、缓存控制器判断目的地址是否在 Cache 中命中，若是，则执行步骤

10 209，若否，则执行步骤 210；

当 CPU 读取主存中的一个字时，该字的主存地址可以被发给 Cache 和主存，此时，缓存控制器可以依据该地址判断该字当前是否存在于 Cache 中，需要说明的是，在实际应用中，缓存控制器可以优先匹配除替换集合以外 cache line，也可以优先匹配替换集合中的 cache line，具体此处不做限定。

15 209、缓存控制器更新访问频度计数；

本实施例中，缓存控制器可以在目的地址在 Cache 中命中之后更新该 cache line 的访问频度计数，该更新方式可以为增加该 cache line 中标记区域的计数器的数值。

在实际应用中，缓存控制器还可以实时监控替换集合中 cache line 的访问频度计数，并且可以从替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。例如，该第三阈值可以为 1。换一种表达方式，在获得替换集合后，缓存控制器可以继续监控替换集合中的 cache line 被命中的次数，若在预设时间内替换集合中的某个 cache line 的访问频度计数大于第三阈值时，缓存控制器可以认为该 cache line 被后续接收的操作指令中的目的地址命中，可能还会被访问，从而可以不必被淘汰。通过这种方式，可以更新替换集合中的 cache line，避免部分被访问的 cache line 被误淘汰。

需要说明的是，缓存控制器在更新完 cache line 的访问频度计数之后完成步骤流程。

210、缓存控制器将从目的地址获取的 cache line 存储于缓存中。

30 本实施例中，缓存控制器可以在目的地址未在 cache 中命中之后从存储介

质中获取该目的地址，并可以将该目的地址替换写入从替换集合中选出的待替换 cache line，需要说明的是，在实际应用中，缓存控制器可以通过按照访问频度计数对应的数值从小到大的顺序，优先选择替换集合中访问频度计数小的 cache line 作为待替换的 cache line。

5 需要说明的是，缓存控制器在将目的地址替换入替换集合中待替换的 cache line 之后完成步骤流程。

在本发明实施例的一个具体场景中，Cache 中包含 10 个 cache line，缓存控制器可以获取到各 cache line 的访问频度计数，详见表 1。

cache line	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
访问频度计数	1	5	10	12	15	20	20	25	30	35

表 1

10 缓存控制器可以根据表 1 中各 cache line 的访问频度计数来确定一个替换集合，其过程如下：

缓存控制器确定访问频度计数下限值为 0，上限值为 30，其中，缓存控制器判定访问频度计数大于参数上限 30 的 cache line 不需要淘汰；

15 缓存控制器可以将访问频度计数 0~30 均匀划分成 3 个访问频度分段（即范围段#1 为 0~10，范围段#2 为 11~20，范围段#3 为 21~30）；

缓存控制器可以通过遍历各 cache line 获得各访问频度分段的 cache line 数量，详见表 2；

范围段	范围段#1	范围段#2	范围段#3
访问频度计数	0~10	11~20	21~30
cache line 数量	3	4	2
cache line	A1、A2、A3	A4、A5、B1、B2	B3、B4

表 2

20 当缓存控制器确定淘汰比例 R 为 50%时，缓存控制器可以计算出待替换的 cache line 的数目 M 为 5($M = \text{cache line 总数} \times \text{淘汰比例 } R = 10 \times 50\% = 5$)，其中，缓存控制器可以确定范围段#1 内 A1、A2、A3 cache line 都可以是待替换的 cache line，并且缓存控制器可以判定范围段#2 为临界访问频度分段，临界访

问频度分段范围段#2 中需要被淘汰的 cache line 数量为 2 (M -范围段#1cache line 数量=5-3=2);

参考表 1 和表 2, 缓存控制器可以按照访问频度计数范围对应的数值从小到大的顺序, 将需要被淘汰的 cache line 累计作为替换集合, 即依次选出 cache line A1、A2、A3、A4 以及 A5 作为替换集合。

当目的地址未在 Cache 中命中时, 缓存控制器可以优先从替换集合中挑选出访问频度计数最低的待替换的 cache line A1, 并将从存储介质中读取的该目的地址替换入该待替换的 cache line A1。

本发明实施例中, 缓存控制器获取操作指令, 其中, 操作指令中携带有目的地址, 目的地址为操作指令待访问的内存中的地址, 若目的地址未命中缓存中的任意一个缓存行 cache line, 且缓存中未包括空闲的 cache line, 则缓存控制器从替换集合中选择待替换的 cache line, 其中, 替换集合中包含有至少两个 cache line, 缓存控制器从缓存中删除待替换的 cache line, 缓存控制器将根据目的地址从内存中获取的 cache line 存储于缓存中, 因而在进行 cache line 替换的过程中, 缓存控制器只需要从替换集合中挑选出待替换的 cache line 即可, 而替换集合是预先选出的, 有效提高了 cache line 替换的效率。

与图 2 所示实施例不同的是, 本发明实施例中一个可选的方案是: 缓存控制器可以不划分 cache line 的访问频度分段, 而直接通过各 cache line 的访问频度计数确定替换集合。

请参阅图 3, 本发明实施例中缓存管理方法另一实施例包括:

本实施例中的步骤 301 至 302 与图 2 所示实施例中的步骤 201 至 202 相同, 此处不再赘述。

本实施例中的步骤 303 与图 2 所示实施例中的步骤 205 相同, 此处不再赘述。

304、缓存控制器确定替换集合;

本实施例中, 缓存控制器可以按照访问频度计数范围对应的数值从小到大的顺序, 依次将各 cache line 累计入替换集合, 直至该替换集合中的待替换的 cache line 的数目等于 M , 该累计入替换集合的 cache line 都为待替换的 cache line。

本实施例中的步骤 305 至 308 与图 2 所示实施例中的步骤 207 至 210 相同,

此处不再赘述。

上面对本发明实施例中的缓存管理方法进行了描述,下面对本发明实施例中的缓存控制器进行描述,图4描述的是与图2方法实施例所对应的装置实施例,请参阅图4,本发明实施例中缓存控制器一个实施例包括:

- 5 获取模块401,用于获取操作指令,其中,操作指令中携带有目的地址,目的地址为操作指令待访问的地址;

第一选择模块402,用于当目的地址未命中计算机系统的缓存中的任意一个缓存行cache line,且缓存中未包括空闲的cache line时,从替换集合中选择待替换的cache line,其中,替换集合中包含有至少两个cache line;

- 10 淘汰模块403,用于从缓存中淘汰待替换的cache line;

存储模块404,用于将从目的地址获取的cache line存储于缓存中。

本实施例中的缓存控制器还包括:

- 15 第一确定模块405,用于根据缓存中各cache line的访问频度计数以及预设的划分策略确定多个访问频度分段,每个访问频度分段对应不同的访问频度计数范围;

第二确定模块406,用于分别根据缓存中各cache line的访问频度计数确定各cache line所属的访问频度分段;

- 20 第二选择模块407,用于按照待替换的cache line的数量M从多个访问频度分段对应的cache line中选择待替换的cache line,以得到替换集合,其中,M为不小于2的整数。

本实施例中的第二选择模块407具体用于按照多个访问频度分段对应的访问频度计数范围从小到大的顺序,依次选择属于各访问频度分段中的cache line,直至选择的cache line的数量等于M。

本实施例中的缓存控制器还包括:

- 25 第三确定模块408,用于根据淘汰比例R与缓存中cache line的总数确定待替换的cache line的数量M,M为淘汰比例R与cache line总数的乘积;

本实施例中的缓存控制器还可以进一步包括:

监控模块409,用于监控淘汰频率参数,淘汰频率参数包括缓存中各cache line的缺失率和被遍历频率中的至少一个;

- 30 调整模块410,用于当淘汰频率参数超过第一阈值时,将淘汰比例R调整

为淘汰比例 R1, 淘汰比例 R1 大于淘汰比例 R;

该调整模块 410, 还用于当淘汰频率参数小于第二阈值时, 将淘汰比例 R 调整为淘汰比例 R2, 淘汰比例 R2 小于淘汰比例 R, 其中, 第二阈值小于第一阈值。

5 本实施例中的缓存控制器还包括:

监控模块 409 具体用于监控属于替换集合的 cache line 的访问频度计数;

淘汰模块 403 具体用于从替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。

10 本实施例中, 获取模块 401 获取操作指令, 其中, 操作指令中携带有目的地址, 目的地址为操作指令待访问的地址, 当目的地址未命中计算机系统的缓存中的任意一个缓存行 cache line, 且缓存中未包括空闲的 cache line 时, 第一选择模块 402 从替换集合中选择待替换的 cache line, 其中, 替换集合中包含有至少两个 cache line, 淘汰模块 403 从缓存中淘汰待替换的 cache line, 存储模块 404 将从目的地址获取的 cache line 存储于缓存中, 因而在进行 cache line
15 替换的过程中, 第一选择模块 402 只需要从替换集合中挑选出待替换的 cache line 即可, 而替换集合是预先选出的, 有效提高了 cache line 替换的效率。

上面对图 2 方法实施例所对应的装置实施例进行了描述, 下面对图 3 所示方法实施例的装置部分实施例进行描述, 请参阅图 5, 本发明实施例中缓存控制器另一实施例包括:

20 获取模块 501, 用于获取操作指令, 其中, 操作指令中携带有目的地址, 目的地址为操作指令待访问的地址;

第一选择模块 502, 用于当目的地址未命中计算机系统的缓存中的任意一个缓存行 cache line, 且缓存中未包括空闲的 cache line 时, 从替换集合中选择待替换的 cache line, 其中, 替换集合中包含有至少两个 cache line;

25 淘汰模块 503, 用于从缓存中淘汰待替换的 cache line;

存储模块 504, 用于将从目的地址获取的 cache line 存储于缓存中。

本实施例中的缓存控制器还包括:

第二选择模块 505, 用于按照待替换的 cache line 的数量 M 从缓存中的 cache line 中选择待替换的 cache line, 以得到替换集合, 其中, M 为不小于 2
30 的整数。

本实施例中的第二选择模块 505 具体用于按照各 cache line 对应的访问频度计数范围从小到大的顺序, 依次选择出 cache line, 直至选择的 cache line 的数量等于 M。

本实施例中的缓存控制器还包括:

- 5 确定模块 506, 用于根据淘汰比例 R 与缓存中 cache line 的总数确定待替换的 cache line 的数量 M, M 为淘汰比例 R 与 cache line 总数的乘积;

本实施例中的缓存控制器还可以进一步包括:

监控单元 507, 用于监控淘汰频率参数, 淘汰频率参数包括缓存中各 cache line 的缺失率和被遍历频率中的至少一个;

- 10 调整模块 508, 用于当淘汰频率参数超过第一阈值时, 将淘汰比例 R 调整为淘汰比例 R1, 淘汰比例 R1 大于淘汰比例 R;

该调整模块 508, 还用于当淘汰频率参数小于第二阈值时, 将淘汰比例 R 调整为淘汰比例 R2, 淘汰比例 R2 小于淘汰比例 R, 其中, 第二阈值小于第一阈值。

- 15 本实施例中的缓存控制器还包括:

监控模块 506 具体用于监控属于替换集合的 cache line 的访问频度计数;

淘汰模块 503 具体用于从替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。

- 20 本实施例中, 获取模块 501 获取操作指令, 其中, 操作指令中携带有目的地址, 目的地址为操作指令待访问的地址, 当目的地址未命中计算机系统的缓存中的任意一个缓存行 cache line, 且缓存中未包括空闲的 cache line 时, 选择模块 502 从替换集合中选择待替换的 cache line, 其中, 替换集合中包含有至少两个 cache line, 淘汰模块 503 从缓存中淘汰待替换的 cache line, 存储模块 504 将从目的地址获取的 cache line 存储于缓存中, 因而在进行 cache line 替换的过程中, 选择模块 502 只需要从替换集合中挑选出待替换的 cache line 即可, 而替换集合是预先选出的, 有效提高了 cache line 替换的效率。
- 25

上面从模块化功能实体的角度对本发明实施例中的缓存控制器进行描述, 下面对本发明实施例中的计算机系统进行描述, 请参阅图 6, 本发明实施例中的计算机一个实施例包括: 处理器 601、缓存 602、缓存控制器 603 和内存 604。其中, 缓存 602 用于缓存内存 604 中的部分数据。

30

在本发明的一些实施例中，处理器 601 用于发送操作指令；

缓存控制器 603 用于：

获取操作指令，其中，操作指令中携带有目的地址，目的地址为操作指令待访问的地址；

- 5 当目的地址未命中计算机系统的缓存 602 中的任意一个缓存行 cache line，且缓存中未包括空闲的 cache line 时，从替换集合中选择待替换的 cache line，其中，替换集合中包含有至少两个 cache line；

从缓存中淘汰待替换的 cache line；

将从目的地址获取的 cache line 存储于缓存 602 中。

- 10 在本发明的一些实施例中，缓存控制器 603 还用于执行以下步骤：

根据缓存中各 cache line 的访问频度计数以及预设的划分策略确定多个访问频度分段，每个访问频度分段对应不同的访问频度计数范围；

分别根据缓存中各 cache line 的访问频度计数确定各 cache line 所属的访问频度分段；

- 15 按照待替换的 cache line 的数量 M 从多个访问频度分段对应的 cache line 中选择待替换的 cache line，以得到替换集合，其中， M 为不小于 2 的整数。

在本发明的一些实施例中，缓存控制器 603 还用于执行以下步骤：

按照多个访问频度分段对应的访问频度计数范围从小到大的顺序，依次选择属于各访问频度分段中的 cache line，直至选择的 cache line 的数量等于 M 。

- 20 在本发明的一些实施例中，缓存控制器 603 还用于执行以下步骤：

监控淘汰频率参数，淘汰频率参数包括缓存中各 cache line 的缺失率和被遍历频率中的至少一个；

当淘汰频率参数超过第一阈值时，将淘汰比例 R 调整为淘汰比例 R_1 ，淘汰比例 R_1 大于淘汰比例 R ；

- 25 当淘汰频率参数小于第二阈值时，将淘汰比例 R 调整为淘汰比例 R_2 ，淘汰比例 R_2 小于淘汰比例 R ，其中，第二阈值小于第一阈值。

在本发明的一些实施例中，缓存控制器 603 还用于执行以下步骤：

监控属于替换集合的 cache line 的访问频度计数；

从替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache

- 30 line。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

本发明实施例还提供一种实现访问请求处理方法的计算机程序产品，包括
5 存储了程序代码的计算机可读存储介质，所述程序代码包括的指令用于执行前述任意一个方法实施例所述的方法流程。本领域普通技术人员可以理解，前述的存储介质包括：U 盘、移动硬盘、磁碟、光盘、随机存储器（Random-Access Memory，RAM）、固态硬盘（Solid State Disk，SSD）或者其他非易失性存储器（non-volatile memory）等各种可以存储程序代码的非短暂性的
10 （non-transitory）机器可读介质。

需要说明的是，本申请所提供的实施例仅仅是示意性的。所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。在本发明实施例、权利要求以及附图中揭示的特征可以独立存在
15 也可以组合存在。在本发明实施例中以硬件形式描述的特征可以通过软件来执行，反之亦然。在此不做限定。

权 利 要 求

1、一种应用于计算机系统上的缓存管理方法，其特征在于，包括：

缓存控制器获取操作指令，其中，所述操作指令中携带有目的地址，所述目的地址为所述操作指令待访问的地址；

5 当所述目的地址未命中所述计算机系统的缓存中的任意一个缓存行 cache line，且所述缓存中未包括空闲的 cache line 时，所述缓存控制器从替换集合中选择待替换的 cache line，其中，所述替换集合中包含有至少两个 cache line；
所述缓存控制器从所述缓存中淘汰所述待替换的 cache line；
所述缓存控制器将根据所述目的地址获取的 cache line 存储于所述缓存
10 中。

2、根据权利要求 1 所述的缓存管理方法，其特征在于，所述方法还包括：

所述缓存控制器根据所述缓存中各 cache line 的访问频度计数以及预设的划分策略确定多个访问频度分段，每个访问频度分段对应不同的访问频度计数范围；

15 所述缓存控制器分别根据所述缓存中各 cache line 的访问频度计数确定各 cache line 所属的访问频度分段；

所述缓存控制器按照待替换的 cache line 的数量 M 从所述多个访问频度分段对应的 cache line 中选择待替换的 cache line，以得到所述替换集合，其中，所述 M 为不小于 2 的整数。

20 3、根据权利要求 2 所述的缓存管理方法，其特征在于，所述缓存控制器按照待替换的 cache line 的数量 M 从所述多个访问频度分段对应的 cache line 中选待替换的 cache line 包括：

所述缓存控制器按照所述多个访问频度分段对应的访问频度计数范围从小到大的顺序，依次选择属于各访问频度分段中的 cache line，直至选择的 cache
25 line 的数量等于所述 M 。

4、根据权利要求 2 或 3 所述的缓存管理方法，其特征在于，所述待替换的 cache line 的数目 M 是根据淘汰比例 R 与所述缓存中 Cache line 的总数确定的，其中，所述 M 为所述淘汰比例 R 与所述 Cache line 总数的乘积；

所述方法还包括：

30 所述缓存控制器监控淘汰频率参数，所述淘汰频率参数包括所述缓存中各

cache line 的缺失率和被遍历频率中的至少一个;

当所述淘汰频率参数超过第一阈值时,所述缓存控制器将所述淘汰比例 R 调整为淘汰比例 R_1 ,所述淘汰比例 R_1 大于所述淘汰比例 R ;

当所述淘汰频率参数小于第二阈值时,所述缓存控制器将所述淘汰比例 R 调整为淘汰比例 R_2 ,所述淘汰比例 R_2 小于所述淘汰比例 R ,其中,所述第二阈值小于所述第一阈值。

5、根据权利要求 1-4 任意一项所述的缓存管理方法,其特征在于,所述各 cache line 的访问频度计数包括:

所述各 cache line 的访问频度计数根据各 cache line 的访问次数获得;或

10 所述各 cache line 的访问频度计数根据各 cache line 的访问次数以及访问时间获得。

6、根据权利要求 1-4 任意一项所述的方法,其特征在于,还包括:

所述缓存控制器监控属于所述替换集合的 cache line 的访问频度计数;

15 所述缓存控制器从所述替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。

7、一种缓存控制器,其特征在于,包括:

获取模块,用于获取操作指令,其中,所述操作指令中携带有目的地址,所述目的地址为所述操作指令待访问的地址;

20 第一选择模块,用于当所述目的地址未命中所述计算机系统的缓存中的任意一个缓存行 cache line,且所述缓存中未包括空闲的 cache line 时,从替换集合中选择待替换的 cache line,其中,所述替换集合中包含有至少两个 cache line;

淘汰模块,用于从所述缓存中淘汰所述待替换的 cache line;

存储模块,用于将从所述目的地址获取的 cache line 存储于所述缓存中。

25 8、根据权利要求 7 所述的缓存控制器,其特征在于,所述缓存控制器还包括:

第一确定模块,用于根据所述缓存中各 cache line 的访问频度计数以及预设的划分策略确定多个访问频度分段,每个访问频度分段对应不同的访问频度计数范围;

30 第二确定模块,用于分别根据所述缓存中各 cache line 的访问频度计数确

定各 cache line 所属的访问频度分段;

第二选择模块,用于按照待替换的 cache line 的数量 M 从所述多个访问频度分段对应的 cache line 中选择待替换的 cache line,以得到所述替换集合,其中,所述 M 为不小于 2 的整数。

5 9、根据权利要求 8 所述的缓存控制器,其特征在于,所述第二选择模块具体用于:按照所述多个访问频度分段对应的访问频度计数范围从小到大的顺序,依次选择属于各访问频度分段中的 cache line,直至选择的 cache line 的数量等于所述 M 。

10 10、根据权利要求 8 或 9 所述的缓存控制器,其特征在于,所述缓存控制器还包括:

第三确定模块,用于根据淘汰比例 R 与所述缓存中 cache line 的总数确定所述待替换的 cache line 的数量 M ,所述 M 为所述淘汰比例 R 与所述 cache line 总数的乘积。

15 11、根据权利要求 10 所述的缓存控制器,其特征在于,所述缓存控制器还包括:

监控模块,用于监控淘汰频率参数,所述淘汰频率参数包括所述缓存中各 cache line 的缺失率和被遍历频率中的至少一个;

调整模块,用于当所述淘汰频率参数超过第一阈值时,将所述淘汰比例 R 调整为淘汰比例 $R1$,所述淘汰比例 $R1$ 大于所述淘汰比例 R ;或

20 当所述淘汰频率参数小于第二阈值时,将所述淘汰比例 R 调整为淘汰比例 $R2$,所述淘汰比例 $R2$ 小于所述淘汰比例 R ,其中,所述第二阈值小于所述第一阈值。

12、根据权利要求 7 所述的缓存控制器,其特征在于:

25 所述监控模块还用于监控属于所述替换集合的 cache line 的访问频度计数;

所述淘汰模块还用于从所述替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache line。

13、一种计算机系统,其特征在于,包括处理器、缓存以及缓存控制器:

所述处理器,用于发送操作指令;

30 所述缓存控制器用于:

获取操作指令，其中，所述操作指令中携带有目的地址，所述目的地址为所述操作指令待访问的地址；

当所述目的地址未命中所述计算机系统的缓存中的任意一个缓存行 cache line，且所述缓存中未包括空闲的 cache line 时，从替换集合中选择待替换的
5 cache line，其中，所述替换集合中包含有至少两个 cache line；

从所述缓存中淘汰所述待替换的 cache line；

将从所述目的地址获取的 cache line 存储于所述缓存中。

14、根据权利要求 13 所述的计算机系统，其特征在于，所述缓存控制器还用于：

10 根据所述缓存中各 cache line 的访问频度计数以及预设的划分策略确定多个访问频度分段，每个访问频度分段对应不同的访问频度计数范围；

分别根据所述缓存中各 cache line 的访问频度计数确定各 cache line 所属的访问频度分段；

按照待替换的 cache line 的数量 M 从所述多个访问频度分段对应的 cache
15 line 中选择待替换的 cache line，以得到所述替换集合，其中，所述 M 为不小于 2 的整数。

15、根据权利要求 14 所述的计算机系统，其特征在于，所述缓存控制器还用于：

按照所述多个访问频度分段对应的访问频度计数范围从小到大的顺序，依
20 次选择属于各访问频度分段中的 cache line，直至选择的 cache line 的数量等于所述 M。

16、根据权利要求 14 或 15 所述的计算机系统，其特征在于，所述缓存控制器还用于：

25 监控淘汰频率参数，所述淘汰频率参数包括所述缓存中各 cache line 的缺失率和被遍历频率中的至少一个；

当所述淘汰频率参数超过第一阈值时，将所述淘汰比例 R 调整为淘汰比例 R1，所述淘汰比例 R1 大于所述淘汰比例 R；

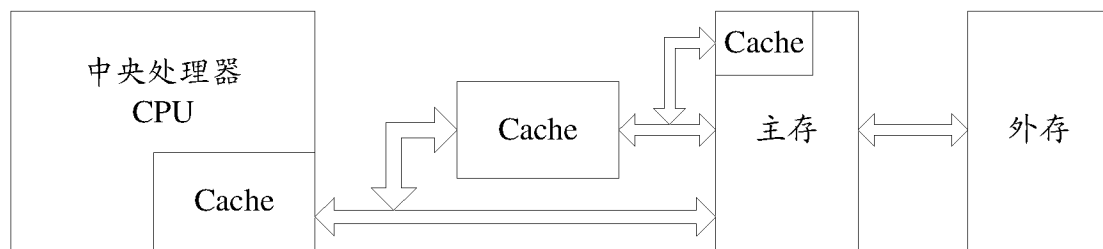
当所述淘汰频率参数小于第二阈值时，将所述淘汰比例 R 调整为淘汰比例 R2，所述淘汰比例 R2 小于所述淘汰比例 R，其中，所述第二阈值小于所述
30 第一阈值。

17、根据权利要求 13 所述的计算机系统，其特征在于，所述缓存控制器还用于：

监控属于所述替换集合的 cache line 的访问频度计数；

从所述替换集合中淘汰在预设时间内访问频度计数大于第三阈值的 cache
5 line。

— 1/5 —

**图 1**

- 2/5 -

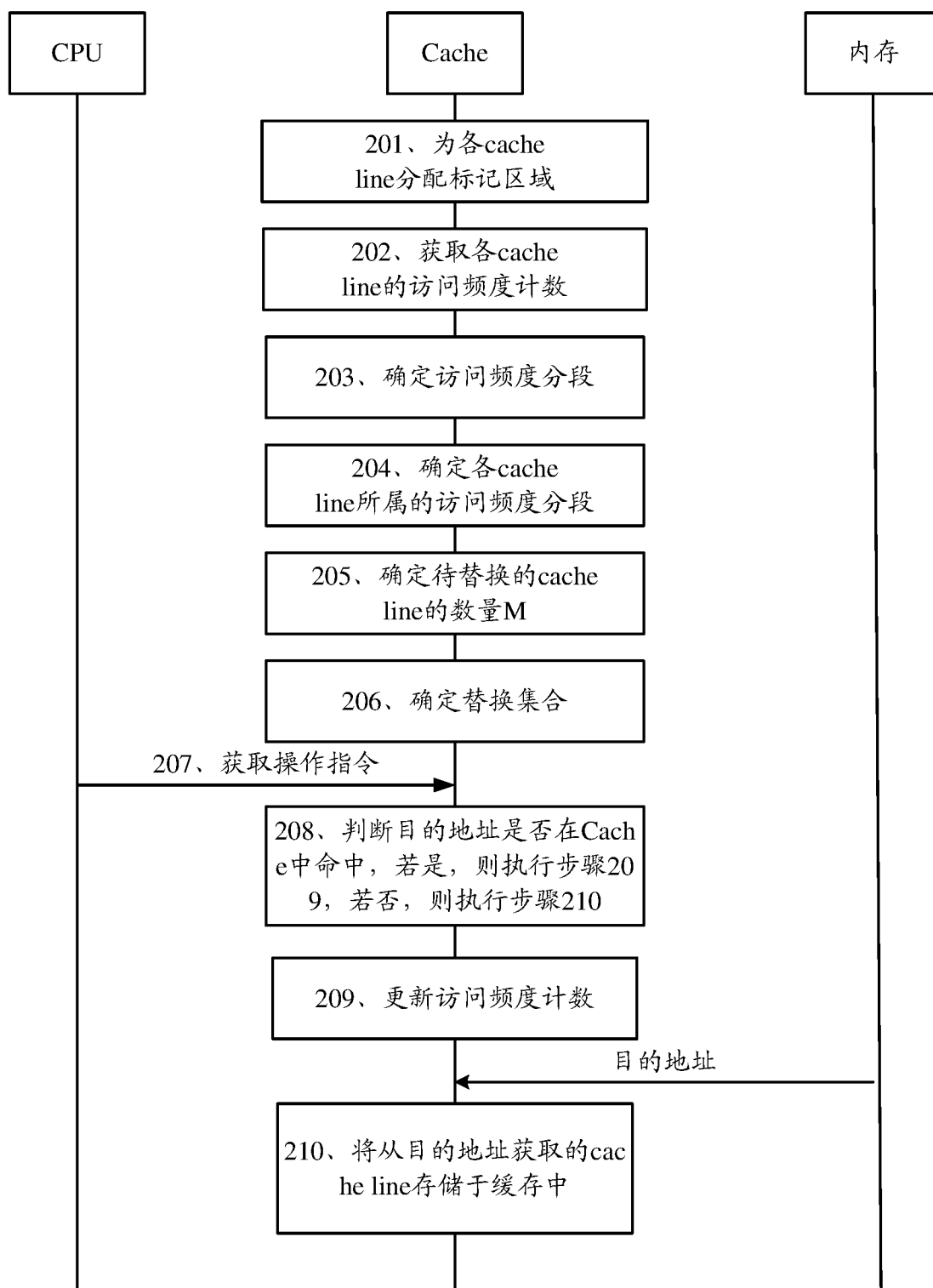


图 2

— 3/5 —

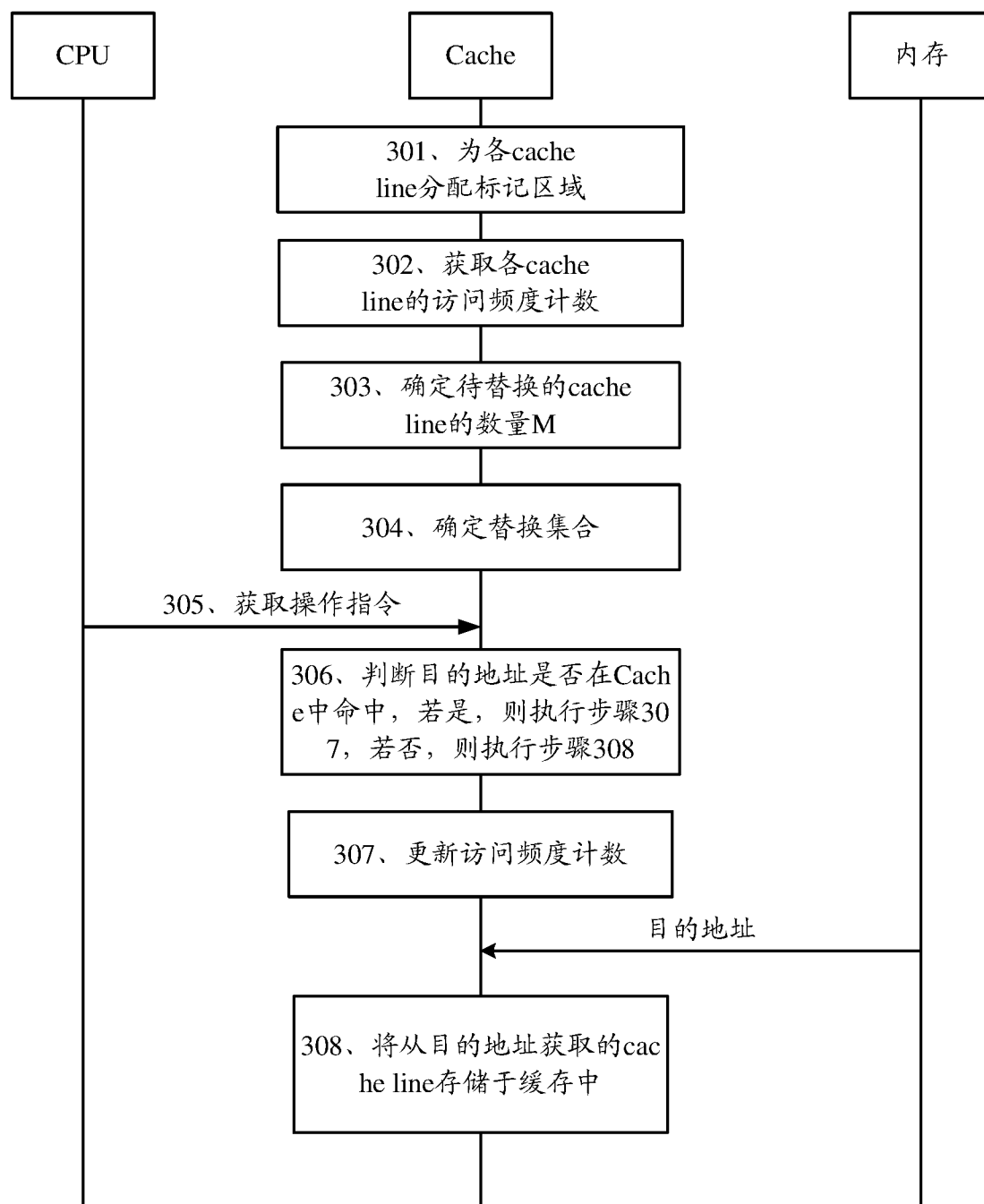


图 3

—4/5—

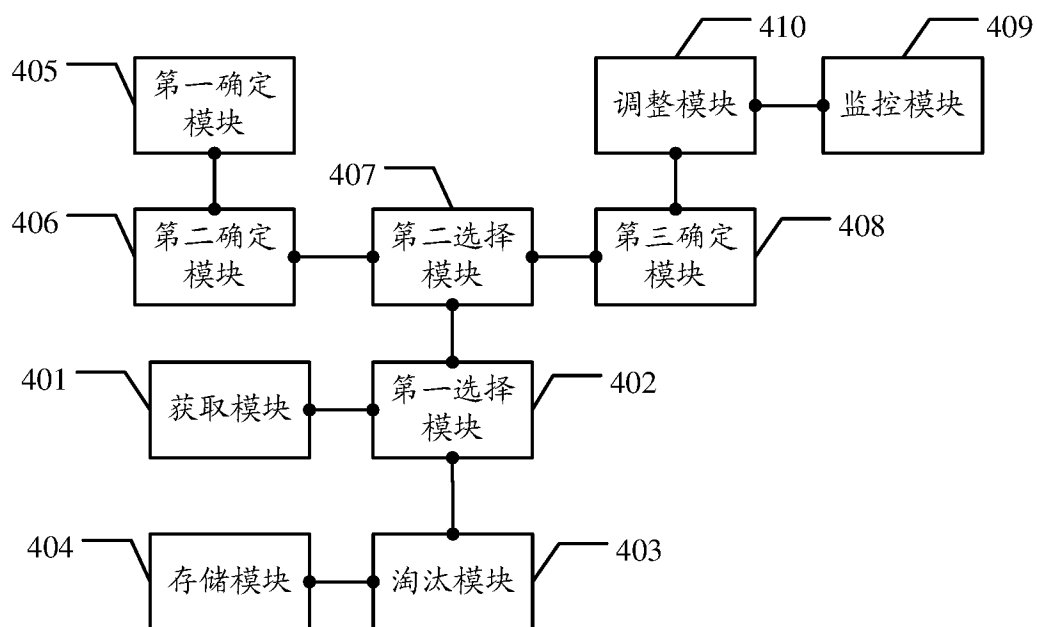


图 4

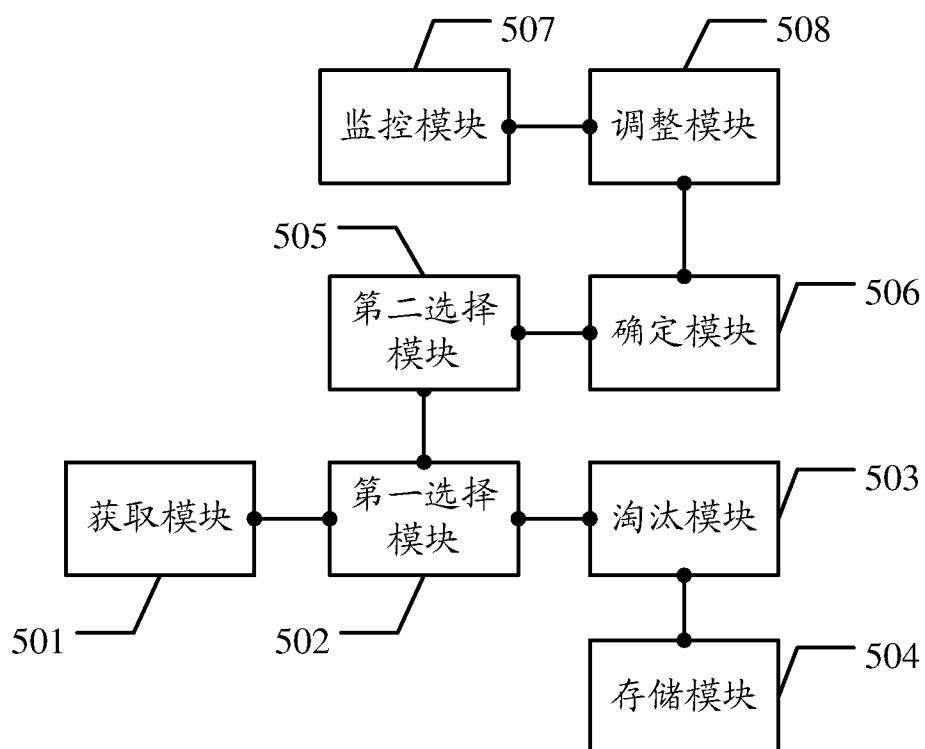


图 5

— 5/5 —

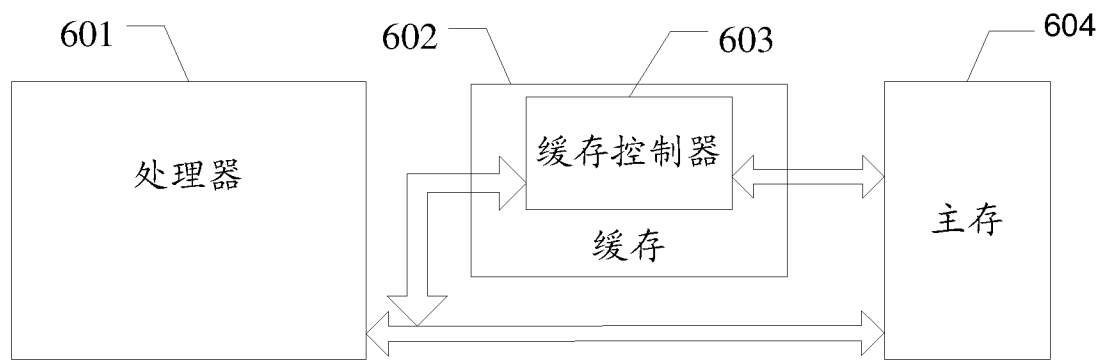


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/070230

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06 (2006.01), 1

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)

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)

CNPAT, WPI, EPODOC, CNKI, IEEE, GOOGLE: section, cache, cache line, choose, visit, frequency, time, group, instead

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105094686 A (HUAWEI TECHNOLOGIES CO., LTD. et al.), 25 November 2015 (25.11.2015), description, paragraphs [0027]-[0030], [0039]-[0054] and [0063]-[0080], and figure 2	1, 5, 7, 13
A	CN 103246616 A (INSUR ELECTRONIC INFORMATION INDUSTRY CO., LTD.), 14 August 2013 (14.08.2013), the whole document	1-17
A	CN 103873562 A (CHEZHI HULIAN (BEIJING) SCIENCE & TECHNOLOGY CO., LTD.), 18 June 2014 (18.06.2014), the whole document	1-17
A	CN 104077242 A (HUAWEI TECHNOLOGIES CO., LTD. et al.), 01 October 2014 (01.10.2014), the whole document	1-17
A	CN 104123243 A (HONGFUJIN PRECISION INDUSTRY (SHENZHEN) CO., LTD. et al.), 29 October 2014 (29.10.2014), the whole document	1-17
A	US 7546417 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 09 June 2009 (09.06.2009), the whole document	1-17

☐ 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 21 September 2016 (21.09.2016)	Date of mailing of the international search report 10 October 2016 (10.10.2016)
Name and mailing address of the ISA/CN: 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 LAN, Juan Telephone No.: (86-10) 010-62413709

International application No.
PCT/CN2016/070230

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2016/070230

A. 主题的分类

G06F 3/06 (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)

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))

CNPAT, WPI, EPDOC, CNKI, IEEE, GOOGLE: 缓存, 替换, 选择, 访问, 频率, 频度, 分段, 分组, 次数, cache, cache line, choose, visit, frequency, time, group, instead

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
X	CN 105094686 A (华为技术有限公司 等) 2015年 11月 25日 (2015 - 11 - 25) 说明书第[0027]-[0030], [0039]-[0054], [0063]-[0080]段, 附图2	1, 5, 7, 13
A	CN 103246616 A (浪潮电子信息产业股份有限公司) 2013年 8月 14日 (2013 - 08 - 14) 全文	1-17
A	CN 103873562 A (车智互联北京科技有限公司) 2014年 6月 18日 (2014 - 06 - 18) 全文	1-17
A	CN 104077242 A (华为技术有限公司 等) 2014年 10月 1日 (2014 - 10 - 01) 全文	1-17
A	CN 104123243 A (鸿富锦精密工业深圳有限公司 等) 2014年 10月 29日 (2014 - 10 - 29) 全文	1-17
A	US 7546417 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2009年 6月 9日 (2009 - 06 - 09) 全文	1-17

☐ 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

2016年 9月 21日

国际检索报告邮寄日期

2016年 10月 10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

受权官员

蓝娟

传真号 (86-10) 62019451

电话号码 (86-10) 010-62413709

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/070230

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	105094686	A	2015年 11月 25日	WO	2015169245	A1	2015年 11月 12日
CN	103246616	A	2013年 8月 14日	无			
CN	103873562	A	2014年 6月 18日	无			
CN	104077242	A	2014年 10月 1日	无			
CN	104123243	A	2014年 10月 29日	无			
US	7546417	B1	2009年 6月 9日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)