

(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



(10) 国际公布号

WO 2021/244067 A1

(43) 国际公布日

2021 年 12 月 9 日 (09.12.2021)

(51) 国际专利分类号:

G06F 3/06 (2006.01)

(21) 国际申请号:

PCT/CN2021/076932

(22) 国际申请日:

2021 年 2 月 19 日 (19.02.2021)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202010505017.8 2020 年 6 月 5 日 (05.06.2020) CN

(71) 申请人: 苏州浪潮智能科技有限公司

(INSPUR SUZHOU INTELLIGENT TECHNOLOGY

CO., LTD.) [CN/CN]; 中国江苏省苏州市吴中

区吴中经济开发区郭巷街道官浦路 1 号

9 幢, Jiangsu 215000 (CN)。

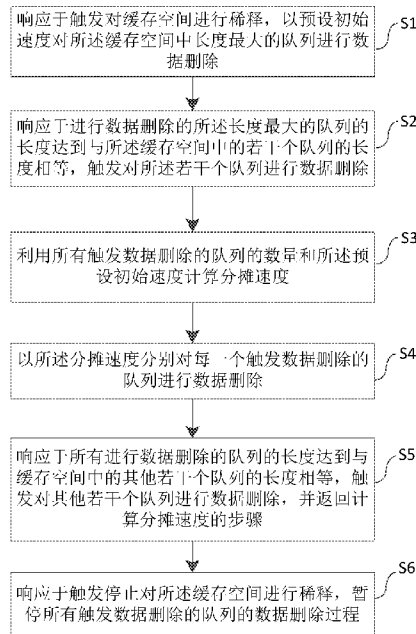
(72) 发明人: 丁增强 (DING, Zengqiang); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。袁博 (YUAN, Bo); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。

(74) 代理人: 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街 22 号赛特广场 7 层, Beijing 100004 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: METHOD FOR DILUTING CACHE SPACE, AND DEVICE AND MEDIUM

(54) 发明名称: 一种稀释缓存空间的方法、设备以及介质



- S1 In response to triggering the dilution of a cache space, perform, at a preset initial speed, data deletion on a queue with the maximum length in the cache space
- S2 In response to the length of the queue with the maximum length and subjected to the data deletion being equal to the lengths of a plurality of queues in the cache space, trigger the performance of data deletion on the plurality of queues
- S3 By means of the number of all queues, the data deletion of which is triggered, and the preset initial speed, calculate an allocation speed
- S4 Perform, at the allocation speed, data deletion on each queue, the data deletion of which is triggered
- S5 In response to the lengths of all queues subjected to the data deletion being equal to the lengths of a plurality of other queues in the cache space, trigger the performance of data deletion on the plurality of other queues, and returning to the step of calculating the allocation speed
- S6 In response to triggering the stopping of diluting the cache space, pause a data deletion process for all queues, the data deletion of which is triggered

图 1

(57) Abstract: A method for diluting a cache space, and a device and a medium. The method comprises: in response to triggering the dilution of a cache space, performing, at a preset initial speed, data deletion on a queue with the maximum length in the cache space (S1); in response to the length of the queue with the maximum length and subjected to the data deletion being equal to the lengths of a plurality of queues in the cache space, triggering the performance of data deletion on the plurality of queues (S2); by means of the number of all queues, the data deletion of which is triggered, and the preset initial speed, calculating an allocation speed (S3);

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,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, 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))。

performing, at the allocation speed, data deletion on each queue, the data deletion of which is triggered (S4); in response to the lengths of all queues subjected to the data deletion being equal to the lengths of a plurality of other queues in the cache space, triggering the performance of data deletion on the plurality of other queues, and returning to the step of calculating the allocation speed (S5); and in response to triggering the stopping of diluting the cache space, pausing a data deletion process (S6).

(57) 摘要: 一种稀释缓存空间的方法、设备以及介质, 包括: 响应于触发对缓存空间进行稀释, 以预设初始速度对缓存空间中长度最大的队列进行数据删除 (S1); 响应于进行数据删除的长度最大的队列的长度达到与缓存空间中的若干个队列的长度相等, 触发对若干个队列进行数据删除 (S2); 利用所有触发数据删除的队列的数量和预设初始速度计算分摊速度(S3); 以分摊速度分别对每一个触发数据删除的队列进行数据删除(S4); 响应于所有进行数据删除的队列的长度达到与缓存空间中的其他若干个队列的长度相等, 触发对其他若干个队列进行数据删除, 并返回计算分摊速度的步骤(S5); 响应于触发停止对缓存空间进行稀释, 暂停数据删除过程(S6)。

一种稀释缓存空间的方法、设备以及介质

本申请要求于 2020 年 06 月 05 日提交中国专利局、申请号为 202010505017.8、发明名称为“一种稀释缓存空间的方法、设备以及介质”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。

5 技术领域

本发明涉及缓存领域，具体涉及一种稀释缓存空间的方法、设备以及存储介质。

背景技术

10 在大数据时代，生产中对数据处理的速度要求越来越苛刻，除了在引擎上的改进，缓存技术也无疑被越来越多的人重视，缓存技术的发展很大程度上改善了数据之间的交互速度，但缓存终究还是不能作为持久化磁盘来使用，因此缓存就存在一个数据过期的过程。现有的数据过期清理策略都是线性的使数据失效，很容易造成缓存雪崩现象的出现。

15

发明内容

有鉴于此，为了克服上述问题的至少一个方面，本发明实施例提出一种稀释缓存空间的方法，包括以下步骤：

20 响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除；

响应于进行数据删除的所述长度最大的队列的长度达到与所述缓存空间中的若干个队列的长度相等，触发对所述若干个队列进行数据删除；

利用所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度；

25 以所述分摊速度分别对每一个触发数据删除的队列进行数据删除；

响应于所有进行数据删除的队列的长度达到与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤；

响应于触发停止对所述缓存空间进行稀释，暂停所有触发数据删除的队列的数据删除过程。

在一些实施例中，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

- 5 根据缓存数据过期策略确定所述缓存空间中待进行数据删除的多个队列；

对所述多个队列中长度最大的队列进行数据删除。

在一些实施例中，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

- 10 检测所述缓存空间已缓存的数据量是否达到阈值或者判断是否接收到用户主动发出的稀释所述缓存空间的指令；

响应于所述缓存空间已缓存的数据量达到阈值或接收到所述用户主动发出的稀释所述缓存空间的指令，触发对所述缓存空间进行稀释。

- 15 在一些实施例中，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

对触发数据删除的队列中的数据进行随机标记；

对所述随机标记后的数据进行归并删除。

- 20 在一些实施例中，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

确定触发数据删除的队列中的数据的删除优先级，以根据所述删除优先级对数据进行标记；

对所述标记后的数据进行归并删除。

- 25 基于同一发明构思，根据本发明的另一个方面，本发明的实施例还提供了一种计算机设备，包括：

至少一个处理器；以及

存储器，所述存储器存储有可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时执行以下步骤：

响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除；

响应于进行数据删除的所述长度最大的队列的长度达到与所述缓存空间中的若干个队列的长度相等，触发对所述若干个队列进行数据删除；

5 利用所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度；

以所述分摊速度分别对每一个触发数据删除的队列进行数据删除；

响应于所有进行数据删除的队列的长度达到与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤；

10

响应于触发停止对所述缓存空间进行稀释，暂停所有触发数据删除的队列的数据删除过程。

在一些实施例中，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

15 根据缓存数据过期策略确定所述缓存空间中待进行数据删除的多个队列；

对所述多个队列中长度最大的队列进行数据删除。

在一些实施例中，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

20 检测所述缓存空间已缓存的数据量是否达到阈值或者判断是否接收到用户主动发出的稀释所述缓存空间的指令；

响应于所述缓存空间已缓存的数据量达到阈值或接收到所述用户主动发出的稀释所述缓存空间的指令，触发对所述缓存空间进行稀释。

25 在一些实施例中，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

对触发数据删除的队列中的数据进行随机标记；

对所述随机标记后的数据进行归并删除。

基于同一发明构思，根据本发明的另一个方面，本发明的实施例还提

供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时执行如上所述的任一种稀释缓存空间的方法的步骤。

- 5 本发明具有以下有益技术效果之一：本发明提出的方案，能够通过将待删除数据分摊给各个符合条件的队列，使得原始缓存数据能够最大限度的保存下来，进而使得数据缓存失效宽泛化，而不是集中失效某一队列或某一部分数据，能最大限度的降低数据雪崩现象的发生，有效避免生产环境中的宕机事件。

10 附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的实
15 施例。

图 1 为本发明的实施例提供的稀释缓存空间的方法的流程示意图；

图 2 为本发明的实施例提供的计算机设备的结构示意图；

图 3 为本发明的实施例提供的计算机可读存储介质的结构示意图。

20 具体实施方式

为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明实施例进一步详细说明。

- 需要说明的是，本发明实施例中所有使用“第一”和“第二”的表述均是为了区分两个相同名称非相同的实体或者非相同的参量，可见“第一”
25 “第二”仅为了表述的方便，不应理解为对本发明实施例的限定，后续实施例对此不再一一说明。

在本发明的实施例中，当数据需要进行缓存时，通过缓存空间对数据

进行缓存，每一批次（例如，可以根据时间范围，例如 5s 之内的数据）作为一个单位发给缓存空间，在缓存空间中将不同批次的数据放入不同的缓存队列中，由于队列的大小由缓存的数据量决定，这样每个批次缓存的数据量不同，所以每个队列大小不一。

- 5 根据本发明的一个方面，本发明的实施例提出一种稀释缓存空间的方法，如图 1 所示，其可以包括步骤：S1，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除；S2，响应于进行数据删除的所述长度最大的队列的长度达到与所述缓存空间中的若干个队列的长度相等，触发对所述若干个队列进行数据删除；S3，利用
10 所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度；S4，以所述分摊速度分别对每一个触发数据删除的队列进行数据删除；S5，响应于所有进行数据删除的队列的长度达到与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤；S6，响应于触发停止对所述缓存空间进行稀释，暂停所有
15 有触发数据删除的队列的数据删除过程。

本发明提出的方案，能够通过将待删除数据分摊给各个符合条件的队列，使得原始缓存数据能够最大限度的保存下来，进而使得数据缓存失效宽泛化，而不是集中失效某一队列或某一部分数据，能最大限度的降低数据雪崩现象的发生，有效避免生产环境中的宕机事件。

- 20 在一些实施例中，在步骤 S1 中，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

根据缓存数据过期策略确定所述缓存空间中待进行数据删除的多个队列；

- 25 对所述多个队列中长度最大的队列进行数据删除。

具体的，缓存过期策略可以是 FIFO（First Input First Output，指先进先出算法）、LFU（Least Frequently Used，最近最少使用算法）、LRU（Least Recently Used，最近最少使用算法），其中 FIFO 指最先进入缓存的数据在

缓存空间不够情况下(超出最大元素限制时)会被首先清理出去; LFU 指一直以来最少被使用的元素会被清理掉, 要求缓存的元素有一个 hit 属性, 在缓存空间不够的情况下, hit 值最小的将会被清出缓存; LRU 指最近最少使用的, 缓存的元素有一个时间戳, 当缓存容量满了, 而又需要腾出地方来
5 缓存新的元素的时候, 那么现有缓存元素中时间戳离当前时间最远的元素将被清出缓存。

这样, 通过缓存过期策略可以确定优先进行清理的多个队列, 从而在多个队列中确定长度最大的队列进行数据删除。

需要说明的是, 也可以在缓存空间全部的队列中选择一个长度最大的
10 队列进行数据删除, 也即可以先不通过缓存过期策略确定优先进行清理的多个队列, 直接对所有的队列进行清理。

在一些实施例中, 在步骤 S1 中, 响应于触发对缓存空间进行稀释, 以预设初始速度对所述缓存空间中长度最大的队列进行数据删除, 进一步包括:

15 检测所述缓存空间已缓存的数据量是否达到阈值或者判断是否接收到用户主动发出的稀释所述缓存空间的指令;

响应于所述缓存空间已缓存的数据量达到阈值或接收到所述用户主动发出的稀释所述缓存空间的指令, 触发对所述缓存空间进行稀释。

具体的, 可以通过多种途径触发对缓存空间进行清理, 包括但不限于通
20 过设置缓存空间的阈值, 当缓存空间中缓存的数据量达到阈值时, 触发对缓存空间的稀释清理; 或者, 当用户想手动触发清理缓存空间时, 可以直接发出清理缓存空间的指令, 当缓存空间接收到相应的指令后, 则触发对缓存空间的稀释清理。

在一些实施例中, 在步骤 S2 中, 响应于进行数据删除后的所述长度最
25 大的队列的长度与所述缓存空间中的若干个队列的长度相等, 触发对所述若干个队列进行数据删除, 具体的, 当长度最大的队列在进行数据删除后的长度减小到与缓存空间中 1 个或更多个的队列的长度相等时, 则触发对该 1 个或更多个的队列进行数据删除。

5 在一些实施例中，在步骤 S3 中，利用所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度，具体的，所有进行数据删除的队列的数据删除速度的总和即为初始速度，这样当有其他队列触发数据删除后，此时长度最大的队列的数据删除速度由于被其他队列进行分摊，因此数据删除的速度变慢。

需要说明的是，分摊速度可以通过预设初始速度除以所有触发数据删除的队列的数量得到。

10 在一些实施例中，在步骤 S5 中，响应于所有进行数据删除后的队列的长度与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤，具体的，在步骤 S3 中，在有新的队列触发数据删除后，当所有进行数据删除后的队列的长度与所述缓存空间中的未进行数据删除的其他 1 个或更多个队列的长度相等时，则触发对该未进行数据删除的其他 1 个或更多个的队列进行数据删除。

15 例如，当触发缓存空间的数据稀释清理后，缓存空间中的最长队列 Q1 开始以预设初始速度进行数据删除，并在最长队列 Q1 的长度等于队列 Q2 的长度后，触发对队列 Q2 的进行数据删除，则以预设初始速度/2 的速度分别对队列 Q1 和 Q2 进行数据删除，并在队列 Q1 和 Q2 的长度等于队列 Q3 的长度后，触发对队列 Q3 的进行数据删除，则以预设初始速度/3 的速度分别对队列 Q1、Q2 和 Q3 进行数据删除，依次类推。当然，所有进行数据删除后的队列的长度也可以与所述缓存空间中多个未进行数据删除的队列的长度相同，即缓存空间中的最长队列 Q1 开始以预设初始速度进行数据删除后，最长队列 Q1 的长度可以等于队列 Q2、Q3 的长度，此时则以预设初始速度/3 的速度分别对队列 Q1、Q2 和 Q3 进行数据删除。这样 Q1 的数据稀释速度被其他队列得到分摊，其稀释速度减慢，从而达到一种
20 25 非线性的数据清除。

在一些实施例中，在步骤 S6 中，响应于触发停止对所述缓存空间进行稀释，暂停所有触发数据删除的队列的数据删除过程，具体的，当对队列进行数据删除后，缓存空间中缓存的数据量小于安全阈值时，触发停止对

缓存空间的稀释清理；或者，当用户想手动触发停止清理缓存空间时，可以直接发出停止清理缓存空间的指令，当缓存空间接收到相应的指令后，则触发停止对缓存空间的稀释清理。

5 在一些实施例中，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

对触发数据删除的队列中的数据进行随机标记；

对所述随机标记后的数据进行归并删除。

10 具体的，每个触发数据清理的队列，可以随机标记待删除数据（标记的速度即为预设初始速度或分摊速度），然后对标记的数据进行归并删除，释放缓存空间，改变队列长度。

在一些实施例中，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

15 确定触发数据删除的队列中的数据的删除优先级，以根据所述删除优先级对数据进行标记；

对所述标记后的数据进行归并删除。

20 具体的，每个触发数据清理的队列，可以通过该队列中数据属性（例如 hit 属性）确定删除顺序，然后根据删除顺序进行标记待删除数据（标记的速度即为预设初始速度或分摊速度），然后对标记的数据进行归并删除，释放缓存空间，改变队列长度。

25 本发明提出的方案，能够通过将待删除数据分摊给各个符合条件的队列，使得原始缓存数据能够最大限度的保存下来，进而使得数据缓存失效宽泛化，而不是集中失效某一队列或某一部分数据，能最大限度的降低数据雪崩现象的发生，有效避免生产环境中的宕机事件。

基于同一发明构思，根据本发明的另一个方面，如图 2 所示，本发明的实施例还提供了一种计算机设备 201，包括：

至少一个处理器 220；以及

存储器 210，存储器 210 存储有可在处理器上运行的计算机程序 211，处理器 220 执行程序时执行如上的任一种稀释缓存空间的方法的步骤。

基于同一发明构思，根据本发明的另一个方面，如图 3 所示，本发明的实施例还提供了一种计算机可读存储介质 301，计算机可读存储介质 301 存储有计算机程序指令 310，计算机程序指令 310 被处理器执行时执行如上的任一种稀释缓存空间的方法的步骤。

最后需要说明的是，本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关硬件来完成，的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，的存储介质可为磁碟、光盘、只读存储记忆体（ROM，Read-Only Memory）或随机存储记忆体（RAM，Random Access Memory）等。上述计算机程序的实施例，可以达到与之对应的前述任意方法实施例相同或者相类似的效果。

此外，典型地，本发明实施例公开的装置、设备等可为各种电子终端设备，例如手机、个人数字助理（PDA，Personal Digital Assistant）、平板电脑（PAD，Portable android device）、智能电视等，也可以是大型终端设备，如服务器等，因此本发明实施例公开的保护范围不应限定为某种特定类型的装置、设备。本发明实施例公开的客户端可以是以电子硬件、计算机软件或两者的组合形式应用于上述任意一种电子终端设备中。

此外，根据本发明实施例公开的方法还可以被实现为由 CPU（central processing unit，中央处理器）执行的计算机程序，该计算机程序可以存储在计算机可读存储介质中。在该计算机程序被 CPU 执行时，执行本发明实施例公开的方法中限定的上述功能。

此外，上述方法步骤以及系统单元也可以利用控制器以及用于存储使得控制器实现上述步骤或单元功能的计算机程序的计算机可读存储介质实现。

此外，应该明白的是，本文的计算机可读存储介质（例如，存储器）可以是易失性存储器或非易失性存储器，或者可以包括易失性存储器和非

易失性存储器两者。作为例子而非限制性的，非易失性存储器可以包括只读存储器（ROM）、可编程 ROM（PROM，Programmable Read-Only Memory）、电可编程 ROM（EPROM，Erasable Programmable Read-Only Memory）、电可擦写可编程 ROM（EEPROM，Electrically Erasable Programmable read only memory）或快闪存储器。易失性存储器可以包括随机存取存储器（RAM），该 RAM 可以充当外部高速缓存存储器。作为例子而非限制性的，RAM 可以以多种形式获得，比如同步 RAM（SRAM，Static Random Access Memory）、动态 RAM（DRAM，Dynamic Random Access Memory）、同步 DRAM（SDRAM，Synchronous Dynamic Random Access Memory）、双数据速率 SDRAM（DDR SDRAM，Double Data Rate Synchronous Dynamic Random Access Memory）、增强 SDRAM（ESDRAM，Enhanced Synchronous Dynamic Random Access Memory）、同步链路 DRAM（SLDRAM，Sync Link Dynamic Random Access Memory）、以及直接 Rambus RAM（RDRAM，Rambus Direct RAM）。所公开的方面的存储设备意在包括但不限于这些和其它合适类型的存储器。

本领域技术人员还将明白的是，结合这里的公开所描述的各种示例性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或两者的组合。为了清楚地说明硬件和软件的这种可互换性，已经就各种示意性组件、方块、模块、电路和步骤的功能对其进行了一般性的描述。这种功能是被实现为软件还是被实现为硬件取决于具体应用以及施加给整个系统的设计约束。本领域技术人员可以针对每种具体应用以各种方式来实现的功能，但是这种实现决定不应被解释为导致脱离本发明实施例公开的范围。

结合这里的公开所描述的各种示例性逻辑块、模块和电路可以利用被设计成用于执行这里功能的下列部件来实现或执行：通用处理器、数字信号处理器（DSP，Digital Signal Processor）、专用集成电路（ASIC，Application Specific Integrated Circuit）、现场可编程门阵列（FPGA，Field Programmable Gate Array）或其它可编程逻辑器件、分立门或晶体管逻辑、分立的硬件组

件或者这些部件的任何组合。通用处理器可以是微处理器，但是可替换地，处理器可以是任何传统处理器、控制器、微控制器或状态机。处理器也可以被实现为计算设备的组合，例如，DSP 和微处理器的组合、多个微处理器、一个或多个微处理器结合 DSP 和/或任何其它这种配置。

5 结合这里的公开所描述的方法或算法的步骤可以直接包含在硬件中、由处理器执行的软件模块中或这两者的组合中。软件模块可以驻留在 RAM 存储器、快闪存储器、ROM 存储器、EPROM 存储器、EEPROM 存储器、寄存器、硬盘、可移动盘、CD-ROM (Compact Disc Read-Only Memory, 只读光盘)、或本领域已知的任何其它形式的存储介质中。示例性的存储介
10 质被耦合到处理器，使得处理器能够从该存储介质中读取信息或向该存储介质写入信息。在一个替换方案中，存储介质可以与处理器集成在一起。处理器和存储介质可以驻留在 ASIC 中。ASIC 可以驻留在用户终端中。在一个替换方案中，处理器和存储介质可以作为分立组件驻留在用户终端中。

 在一个或多个示例性设计中，功能可以在硬件、软件、固件或其任意
15 组合中实现。如果在软件中实现，则可以将功能作为一个或多个指令或代码存储在计算机可读介质上或通过计算机可读介质来传送。计算机可读介质包括计算机存储介质和通信介质，该通信介质包括有助于将计算机程序从一个位置传送到另一个位置的任何介质。存储介质可以是能够被通用或专用计算机访问的任何可用介质。作为例子而非限制性的，该计算机可读
20 介质可以包括 RAM、ROM、EEPROM、CD-ROM 或其它光盘存储设备、磁盘存储设备或其它磁性存储设备，或者是可以用于携带或存储形式为指令或数据结构的所需程序代码并且能够被通用或专用计算机或者通用或专用处理器访问的任何其它介质。此外，任何连接都可以适当地称为计算机可读介质。例如，如果使用同轴线缆、光纤线缆、双绞线、数字用户线路
25 (DSL, Digital Subscriber Line) 或诸如红外线、无线电和微波的无线技术来从网站、服务器或其它远程源发送软件，则上述同轴线缆、光纤线缆、双绞线、DSL 或诸如红外线、无线电和微波的无线技术均包括在介质的定义。如这里所使用的，磁盘和光盘包括压缩盘 (CD, Compact Disk)、激

光盘、光盘、数字多功能盘（DVD，Digital Video Disc）、软盘、蓝光盘，其中磁盘通常磁性地再现数据，而光盘利用激光光学地再现数据。上述内容的组合也应当包括在计算机可读介质的范围内。

5 以上是本发明公开的示例性实施例，但是应当注意，在不背离权利要求限定的本发明实施例公开的范围的前提下，可以进行多种改变和修改。根据这里描述的公开实施例的方法权利要求的功能、步骤和/或动作不需以任何特定顺序执行。此外，尽管本发明实施例公开的元素可以以个体形式描述或要求，但除非明确限制为单数，也可以理解为多个。

10 应当理解的是，在本文中使用的，除非上下文清楚地支持例外情况，单数形式“一个”旨在也包括复数形式。还应当理解的是，在本文中使用的“和/或”是指包括一个或者一个以上相关联地列出的项目的任意和所有可能组合。

上述本发明实施例公开实施例序号仅仅为了描述，不代表实施例的优劣。

15 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

20 所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本发明实施例公开的范围（包括权利要求）被限于这些例子；在本发明实施例的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，并存在如上的本发明实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。因此，凡在本发明实施例的精神和原则之内，所做的任何省略、修改、等同替换、改进等，均应包含
25 在本发明实施例的保护范围之内。

权 利 要 求

1、一种稀释缓存空间的方法，其特征在于，包括以下步骤：

响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除；

5 响应于进行数据删除的所述长度最大的队列的长度达到与所述缓存空间中的若干个队列的长度相等，触发对所述若干个队列进行数据删除；

利用所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度；

以所述分摊速度分别对每一个触发数据删除的队列进行数据删除；

10 响应于所有进行数据删除的队列的长度达到与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤；

响应于触发停止对所述缓存空间进行稀释，暂停所有触发数据删除的队列的数据删除过程。

15 2、如权利要求 1 所述的方法，其特征在于，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

根据缓存数据过期策略确定所述缓存空间中待进行数据删除的多个队列；

20 对所述多个队列中长度最大的队列进行数据删除。

3、如权利要求 1 所述的方法，其特征在于，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

25 检测所述缓存空间已缓存的数据量是否达到阈值或者判断是否接收到用户主动发出的稀释所述缓存空间的指令；

响应于所述缓存空间已缓存的数据量达到阈值或接收到所述用户主动发出的稀释所述缓存空间的指令，触发对所述缓存空间进行稀释。

4、如权利要求 1 所述的方法，其特征在于，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个

触发数据删除的队列进行数据删除，进一步包括：

对触发数据删除的队列中的数据进行随机标记；

对所述随机标记后的数据进行归并删除。

- 5 5、如权利要求 1 所述的方法，其特征在于，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

确定触发数据删除的队列中的数据的删除优先级，以根据所述删除优先级对数据进行标记；

对所述标记后的数据进行归并删除。

- 10 6、一种计算机设备，包括：

至少一个处理器；以及

存储器，所述存储器存储有可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时执行以下步骤：

- 15 响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除；

响应于进行数据删除的所述长度最大的队列的长度达到与所述缓存空间中的若干个队列的长度相等，触发对所述若干个队列进行数据删除；

利用所有触发数据删除的队列的数量和所述预设初始速度计算分摊速度；

- 20 以所述分摊速度分别对每一个触发数据删除的队列进行数据删除；

响应于所有进行数据删除的队列的长度达到与所述缓存空间中的其他若干个队列的长度相等，触发对所述其他若干个队列进行数据删除，并返回计算分摊速度的步骤；

- 25 响应于触发停止对所述缓存空间进行稀释，暂停所有触发数据删除的队列的数据删除过程。

7、如权利要求 6 所述的设备，其特征在于，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

根据缓存数据过期策略确定所述缓存空间中待进行数据删除的多个队

列;

对所述多个队列中长度最大的队列进行数据删除。

- 8、如权利要求 6 所述的设备，其特征在于，响应于触发对缓存空间进行稀释，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除，进一步包括：

检测所述缓存空间已缓存的数据量是否达到阈值或者判断是否接收到用户主动发出的稀释所述缓存空间的指令；

响应于所述缓存空间已缓存的数据量达到阈值或接收到所述用户主动发出的稀释所述缓存空间的指令，触发对所述缓存空间进行稀释。

- 9、如权利要求 6 所述的设备，其特征在于，以预设初始速度对所述缓存空间中长度最大的队列进行数据删除或以所述分摊速度分别对每一个触发数据删除的队列进行数据删除，进一步包括：

对触发数据删除的队列中的数据进行随机标记；

对所述随机标记后的数据进行归并删除。

- 10、一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时执行如权利要求 1-5 任意一项所述的方法的步骤。

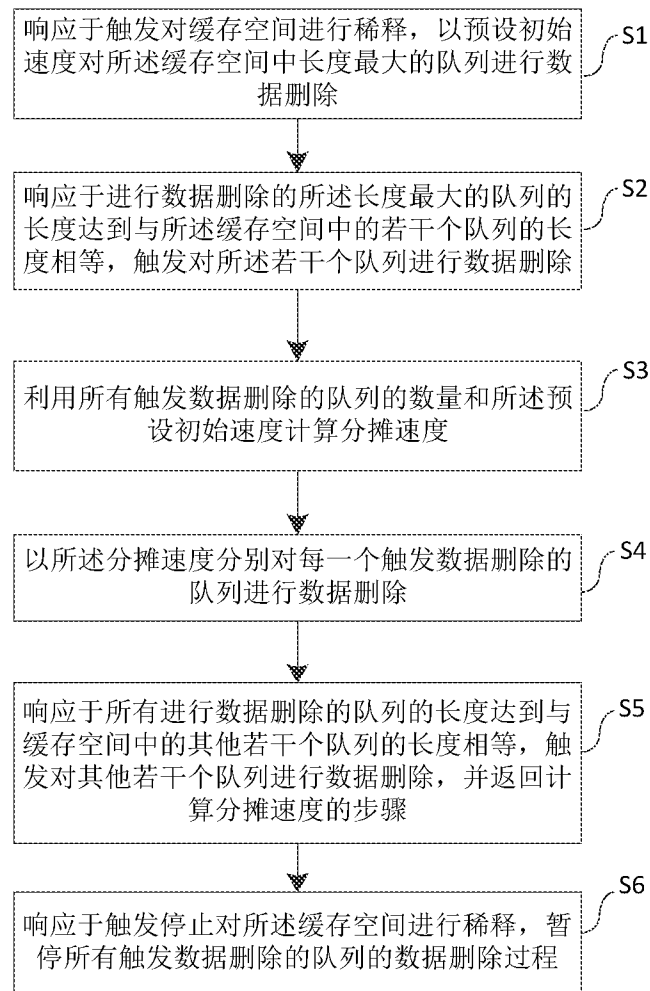


图 1

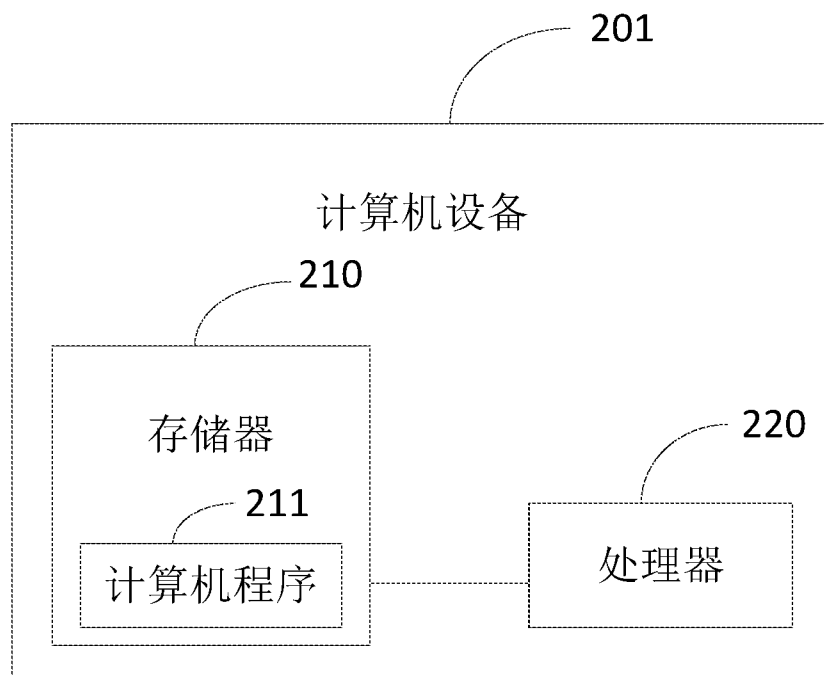


图 2

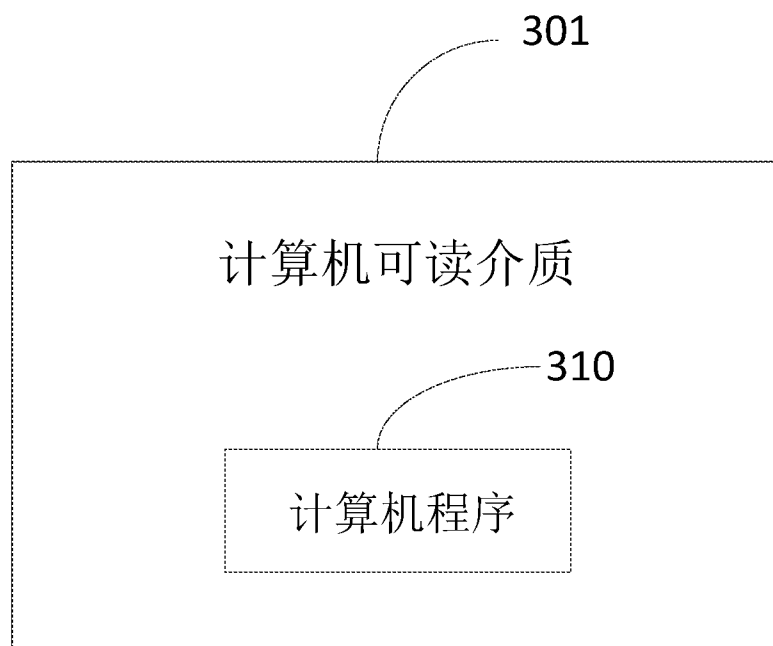


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/076932

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06(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)

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, EPODOC, WPI, CNKI, ISI: 缓存, 稀释, 雪崩, 分摊, 均摊, 分配, 均衡, 队列, 擦除, 删除, 速度, 失效, 过期, cache, dilute, avalanche, share, proportion, balance, queue, erase, delete, speed, expired, time, limit

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111736769 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 02 October 2020 (2020-10-02) claims 1-10	1-10
A	CN 106021445 A (NUBIA TECHNOLOGY CO., LTD.) 12 October 2016 (2016-10-12) description paragraphs [0005]-[0064], [0251]	1-10
A	CN 110119487 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 13 August 2019 (2019-08-13) entire document	1-10
A	CN 109491928 A (SHENZHEN LEXIN SOFTWARE TECHNOLOGY CO., LTD.) 19 March 2019 (2019-03-19) entire document	1-10
A	US 2016342359 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 24 November 2016 (2016-11-24) entire document	1-10



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

08 May 2021

Date of mailing of the international search report

18 May 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/076932

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111736769	A	02 October 2020	None			
CN	106021445	A	12 October 2016	CN	106021445	B	15 October 2019
CN	110119487	A	13 August 2019	None			
CN	109491928	A	19 March 2019	None			
US	2016342359	A1	24 November 2016	US	2015309749	A1	29 October 2015
				US	9098717	B2	04 August 2015
				US	2008263271	A1	23 October 2008
				US	9933959	B2	03 April 2018
				US	9542109	B2	10 January 2017

国际检索报告

国际申请号

PCT/CN2021/076932

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, EPDOC, WPI, CNKI, ISI: 缓存, 稀释, 雪崩, 分摊, 均摊, 分配, 均衡, 队列, 擦除, 删除, 速度, 失效, 过期, cache, dilute, avalanche, share, proportion, balance, queue, erase, delete, speed, expired, time, limit																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 111736769 A (苏州浪潮智能科技有限公司) 2020年 10月 2日 (2020 - 10 - 02) 权利要求1-10</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 106021445 A (努比亚技术有限公司) 2016年 10月 12日 (2016 - 10 - 12) 说明书第[0005]-[0064]、[0251]段</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 110119487 A (华南理工大学) 2019年 8月 13日 (2019 - 08 - 13) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 109491928 A (深圳乐信软件技术有限公司) 2019年 3月 19日 (2019 - 03 - 19) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 2016342359 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2016年 11月 24日 (2016 - 11 - 24) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 111736769 A (苏州浪潮智能科技有限公司) 2020年 10月 2日 (2020 - 10 - 02) 权利要求1-10	1-10	A	CN 106021445 A (努比亚技术有限公司) 2016年 10月 12日 (2016 - 10 - 12) 说明书第[0005]-[0064]、[0251]段	1-10	A	CN 110119487 A (华南理工大学) 2019年 8月 13日 (2019 - 08 - 13) 全文	1-10	A	CN 109491928 A (深圳乐信软件技术有限公司) 2019年 3月 19日 (2019 - 03 - 19) 全文	1-10	A	US 2016342359 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2016年 11月 24日 (2016 - 11 - 24) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 111736769 A (苏州浪潮智能科技有限公司) 2020年 10月 2日 (2020 - 10 - 02) 权利要求1-10	1-10																		
A	CN 106021445 A (努比亚技术有限公司) 2016年 10月 12日 (2016 - 10 - 12) 说明书第[0005]-[0064]、[0251]段	1-10																		
A	CN 110119487 A (华南理工大学) 2019年 8月 13日 (2019 - 08 - 13) 全文	1-10																		
A	CN 109491928 A (深圳乐信软件技术有限公司) 2019年 3月 19日 (2019 - 03 - 19) 全文	1-10																		
A	US 2016342359 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2016年 11月 24日 (2016 - 11 - 24) 全文	1-10																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 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 </td> <td> “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 </td> </tr> </table>			* 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																
* 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 2021年 5月 8日		国际检索报告邮寄日期 2021年 5月 18日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局 (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		授权官员 欧晓丹 电话号码 86-(10)-53961416																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/076932

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	111736769	A	2020年 10月 2日	无	
CN	106021445	A	2016年 10月 12日	CN 106021445	B 2019年 10月 15日
CN	110119487	A	2019年 8月 13日	无	
CN	109491928	A	2019年 3月 19日	无	
US	2016342359	A1	2016年 11月 24日	US 2015309749	A1 2015年 10月 29日
				US 9098717	B2 2015年 8月 4日
				US 2008263271	A1 2008年 10月 23日
				US 9933959	B2 2018年 4月 3日
				US 9542109	B2 2017年 1月 10日