



[12] 发明专利说明书

专利号 ZL 200510132879.6

[45] 授权公告日 2008年4月23日

[11] 授权公告号 CN 100383791C

[22] 申请日 2005.12.27

[21] 申请号 200510132879.6

[73] 专利权人 北京中星微电子有限公司

地址 100083 北京市海淀区学院路35号
世宁大厦15层

[72] 发明人 艾国

[56] 参考文献

US2004220941A1 2004.11.4

CN1530852A 2004.9.22

US5907842A 1999.5.25

CN1659546A 2005.8.24

CN1489738A 2004.4.14

审查员 王雪莲

[74] 专利代理机构 北京北翔知识产权代理有限公司

代理人 陈霁

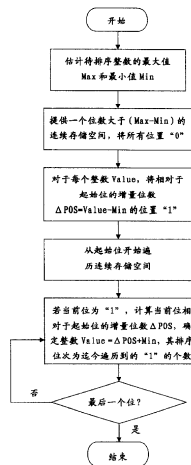
权利要求书2页 说明书6页 附图2页

[54] 发明名称

基于位的排序方法和装置

[57] 摘要

对一组互不相同的整数排序的方法，包括：估计该组整数中的最大值和最小值；提供位数不小于最大值与最小值之差的连续存储空间，将连续存储空间的所有位置为第一值，连续存储空间具有起始位；选择一整数作为参考值；对于该组整数中的每个整数，将连续存储空间中、相对于起始位的增量位数等于该整数相对于参考值之增量的位置为第二值，其中参考值的选择使得针对每个整数被置为第二值的位都在连续存储空间的范围；对于每个被置为第二值的位，确定该位相对于起始位的增量位数与参考值之和为该组整数中的一个整数，确定该整数的排序位次为该位在所有被置为第二值的位中所处的位次。并提供基于关键字对记录排序的方法。主要优点是复杂度仅为 $O(n)$ 。



1. 对一组互不相同的整数排序的方法，包括：
 - 估计所述一组互不相同的整数中的最大值和最小值；
 - 提供一个位数不小于最大值与最小值之差的连续存储空间，将所述连续存储空间的所有位置为第一值，所述连续存储空间具有起始位；
 - 选择一个整数作为参考值；
 - 对于所述一组整数中的每个整数，将所述连续存储空间中、相对于所述起始位的增量位数等于该整数相对于所述参考值之增量的位置为第二值，其中所述参考值的选择使得针对所述每个整数被置为第二值的位都在所述连续存储空间的范围内；以及，
 - 对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组整数中的一个整数，确定该整数的排序位次为该位在所有被置为第二值的位中所处的位次。
2. 根据权利要求1所述的方法，其特征在于，所述参考值是最小值。
3. 一种排序装置，用于对一组互不相同的整数排序，该装置包括：
 - 估计所述一组互不相同的整数中的最大值和最小值的装置；
 - 提供一个位数不小于最大值与最小值之差的连续存储空间并将所述连续存储空间的所有位置为第一值的装置，所述连续存储空间具有起始位；
 - 选择一个整数作为参考值的装置；
 - 映射装置，对于所述一组整数中的每个整数，将所述连续存储空间中、相对于所述起始位的增量位数等于该整数相对于所述参考值之增量的位置为第二值，其中所述参考值的选择使得针对所述每个整数被置为第二值的位都在所述连续存储空间的范围内；以及，
 - 还原装置，对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组整数中的一个整数，确定该整数的排序位次为该位在所有被置为第二值的位中所处的位次。
4. 根据权利要求3所述的装置，其特征在于，所述参考值是最小值。
5. 对一组记录排序的方法，所述一组记录中的每个记录具有能唯一识别该记录、且为整数的关键字，该方法包括：
 - 估计所述一组记录中的各个记录的关键字中的最大值和最小值；
 - 提供一个位数不小于最大值与最小值之差的连续存储空间，将所

述连续存储空间的所有位置为第一值，所述连续存储空间具有起始位；

选择一个整数作为参考值；

对于所述一组记录中的每个记录的关键字，将所述连续存储空间中、相对于所述起始位的增量位数等于该关键字相对于所述参考值之增量的位置为第二值，将该关键字所识别的记录与该位对应，其中所述参考值的选择使得针对所述每个记录的关键字被置为第二值的位都在所述连续存储空间的范围内；以及，

对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组记录中的一个记录的关键字，确定该关键字的排序位次为该位在所有被置为第二值的位中所处的位次，将该关键字与对应于该位的记录对应。

6. 根据权利要求5所述的方法，其特征在于，所述参考值是最小值。

7. 一种排序装置，用于对一组记录排序，所述一组记录中的每个记录具有能唯一识别该记录、且为整数的关键字，该装置包括：

估计所述一组记录中的各个记录的关键字中的最大值和最小值的装置；

提供一个位数不小于最大值与最小值之差的连续存储空间并将所述连续存储空间的所有位置为第一值的装置，所述连续存储空间具有起始位；

选择一个整数作为参考值的装置；

映射装置，对于所述一组记录中的每个记录的关键字，将所述连续存储空间中、相对于所述起始位的增量位数等于该关键字相对于所述参考值之增量的位置为第二值，将该关键字所识别的记录与该位对应，其中所述参考值的选择使得针对所述每个记录的关键字被置为第二值的位都在所述连续存储空间的范围内；以及，

还原装置，对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组记录中的一个记录的关键字，确定该关键字的排序位次为该位在所有被置为第二值的位中所处的位次，将该关键字与对应于该位的记录对应。

8. 根据权利要求7所述的装置，其特征在于，所述参考值是最小值。

基于位的排序方法和装置

技术领域

本发明涉及排序方法及装置。

背景技术

排序是一种基本且常用的算法。所谓排序，通常是指对文件中的记录进行整理，使之按关键字递增(或递减)的次序排列。排序对象，即文件，由一组记录组成。每个记录由若干个数据项(或域)组成，其中有一个数据项可用来标识记录，称为关键字项。用来作为排序运算依据的关键字项可以是数字类型，也可以是字符类型，其选取应根据具体情况而定，其值称为关键字。

随着计算机的发展，排序算法层出不穷。常用的有，例如，起泡排序、直接插入排序、简单选择排序、快速排序和堆排序等经典排序算法。大多数排序算法都涉及两个基本操作：比较关键字的大小；和，改变指向记录的指针或移动记录本身，其中后者的实现依赖于待排序记录的存储方式。大多数排序算法的时间开销主要来源于关键字之间的比较和记录的移动。对于有些排序算法，其执行时间不仅取决于问题的规模，还取决于实例中数据的状态。因此，在实际进行排序时，应根据具体情况选择合适的排序算法。

衡量算法性能的一个主要标准是时间复杂度，一般用 O 方法表示。例如，起泡排序的平均时间复杂度为 $O(n^2)$ ，快速排序的平均时间复杂度为 $O(n \lg n)$ 。

发明内容

本发明提出了一种基于位的排序方法和装置。

根据本发明的第一方面，提供了对一组互不相同的整数排序的方法，包括：

估计所述一组互不相同的整数中的最大值和最小值；

提供一个位数不小于最大值与最小值之差的连续存储空间，将所述连续存储空间的所有位置为第一值，所述连续存储空间具有起始位；

选择一个整数作为参考值；

对于所述一组整数中的每个整数，将所述连续存储空间中、相对于所述起始位的增量位数等于该整数相对于所述参考值之增量的位置为第二值，其中所述参考值的选择使得针对所述每个整数被置为第二值的位都在所述连续存储空间的范围内；以及，

对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组整数中的一个整数，确定该整数的排序位次为该位在所有被置为第二值的位中所处的位次。

优选的是，所述参考值是最小值。

根据本发明的第二方面，提供了一种排序装置，用于按照第一方面所述的方法对一组互不相同的整数排序。

根据本发明的第三方面，提供了对一组记录排序的方法，所述一组记录中的每个记录具有能唯一识别该记录、且为整数的关键字，该方法包括：

估计所述一组记录中的各个记录的关键字中的最大值和最小值；

提供一个位数不小于最大值与最小值之差的连续存储空间，将所述连续存储空间的所有位置为第一值，所述连续存储空间具有起始位；

选择一个整数作为参考值；

对于所述一组记录中的每个记录的关键字，将所述连续存储空间中、相对于所述起始位的增量位数等于该关键字相对于所述参考值之增量的位置为第二值，将该关键字所识别的记录与该位对应，其中所述参考值的选择使得针对所述每个记录的关键字被置为第二值的位都在所述连续存储空间的范围内；以及，

对于每个被置为第二值的位，确定该位相对于所述起始位的增量位数与所述参考值之和为所述一组记录中的一个记录的关键字，确定该关键字的排序位次为该位在所有被置为第二值的位中所处的位次，将该关键字与对应于该位的记录对应。

优选的是，所述参考值是最小值。

根据本发明的第四方面，提供了一种对一组记录排序的排序装置，用于实现第三方面所述的方法。

本发明的主要优点是时间复杂度仅为 $O(n)$ 。附加优点是，在某些情况下，可节省存储空间。

附图说明

仅以举例的方式，参照附图更详细地描述本发明的优选实施方案，其中：

图 1 示出了根据本发明的一个优选实施方案的流程图；

图 2a-2g 示出了利用图 1 的流程图对一个实例排序的过程；和

图 3 示意性图解了根据本发明对记录排序的过程。

具体实施方式

首先通过一个具体例子来说明本发明的基本原则。

假设某班一共有 15 名学生，其中六名学生要补考英语，他们的学号分别为 10、8、1、5、4、9，现在希望借助计算机根据他们学号的次序安排他们补考时的座位次序。为此，计算机要对这六名学生的学号进行排序。容易理解的是，待排序的学号均为整数、且互不相同。需要说明的是，该实例可能并未体现出本发明的一般应用，不过给出这个例子的目的仅在于清楚地描述本发明的基本原则。稍后将结合一般性的例子对本发明作进一步描述。

根据本发明的一个优选实施方案，参照图 1 所示的流程图进行排序。首先，估计待排序整数的最大值 Max 和最小值 Min。在这个例子中，根据该班共 15 名学生且学号一般会从 1 开始，可估计待排序整数的最大值为 $\text{Max}=15$ 、最小值为 $\text{Min}=1$ 。

然后，提供一个位数大于最大值与最小值之差的连续存储空间，将所有位置“0”。在这个例子中，最大值与最小值之差为 $(\text{Max}-\text{Min})=14$ 。鉴于计算机中的存储空间一般是以字节 (Byte) 为单位的，我们可确定一个 16 位 (Bit) 的连续存储空间，如图 2a 所示。

接下来，对于每个整数，将连续存储空间中、相对于起始位的增量位数等于该整数相对于最小值之增量的位置“1”。在这个例子中，对于第一个整数 10，其相对于最小值 Min 的增量为 $10-1=9$ ，于是将连续存储空间中相对于起始位的增量位数等于 9 的位，即第 $(1+9)=10$ 位，置“1”，如图 2b 所示；对于第二个整数 8，其相对于最小值的增量为 $8-1=7$ ，于是将连续存储空间中相对于起始位的增量位数等于 7 的位，即第 $(1+7)=8$ 位，置“1”，如图 2c 所示；依此类推，对于第三、四、五、六个整数：1、5、4、9，分别将第 $(1+(1-\text{Min}))=1$ 位、第 $(1+(5-\text{Min}))=5$ 位、第 $(1+(4-\text{Min}))=4$ 位、第 $(1+(9-\text{Min}))=9$ 位置“1”，如图 2d-2g 所示。

然后，从第一位开始遍历连续存储空间。若当前位为“1”，则将当前位相对于起始位的增量位数与最小值之和确定为所排序整数中的

一个整数，并将到目前为止遍历到的“1”的个数确定为该整数的排序位次。就这个例子而言，第一位是“1”且是遍历到的第一个“1”，第一位相对于起始位的增量位数是 $(1-1)=0$ ，可确定排好序后的第一个整数相对于最小值的增量为0，即第一个整数为 $(\text{Min}+0)=1$ ；第四位是“1”且是遍历到的第二个“1”，第四位相对于起始位的增量位数是 $(4-1)=3$ ，可确定排好序后的第二个整数相对于最小值的增量为3，即第二个整数为 $(\text{Min}+3)=4$ ；类似地，第5位、第8位、第9位、第10位均为“1”，可确定排好序后的第三、四、五、六个整数分别相对于最小值的增量位数为 $(5-1)=4$ 、 $(8-1)=7$ 、 $(9-1)=8$ 、 $(10-1)=9$ ，即第三、四、五、六个整数分别为 $(\text{Min}+4)=5$ 、 $(\text{Min}+7)=8$ 、 $(\text{Min}+8)=9$ 、 $(\text{Min}+9)=10$ 。这样，就得到了这6个整数即学号的排序结果：1、4、5、8、9、10。

可替换地，在这一遍历步骤中，可从连续存储空间的最后一位进行遍历，其它过程类似。容易理解的是，利用这种方式遍历，将得到这6名学生的学号的递减排序结果：10、9、8、5、4、1。

实际上，针对所有排序整数进行上述置“1”操作后，连续存储空间中存储的位流和用作置“1”操作参考的最小值一起包含了有关所有排序整数的值及排序次序的信息，即已经完成了排序。因此，根据具体情况，就排序而言，上述遍历步骤可以不是必要的，而是可以在后来需利用排序结果时进行，例如在查询、检索等进行。

需要说明的是，在这个优选实施方案中，通过置“1”将待排序整数映射到所提供的连续存储空间中的操作是相对所估计的最小值进行的，通常这也是优选的。然而，就本发明的实质而言，该操作也可相对除最小值以外的其它整数进行。当然，这个用作参考的整数应合适，以便能将待排序整数映射到所提供的连续存储空间的范围内，即被确定要置“1”的位不超出连续存储空间的范围，从而避免空间不够用。另外，对于所提供的连续存储空间，最初可将所有位置“1”，之后进行映射操作时将有关位置“0”。

下面通过对记录排序来对本发明进行描述。例如，一个学生管理系统的数据库中有 n 条记录，每条记录有一个唯一的关键字，这个关键字一般是从1开始的。由于经常对数据库进行修改、删除和添加，所以数据库中的关键字一般并不是顺序和连续的。但我们经常要查询某个关键字的记录，我们通常会按照关键字对数据库中的记录进行排序，以便之后能以很快的速度进行查询。

按照之前描述的方式，假如我们估计这个学校有 1000 名学生，则可推定关键字的最小值为 $\text{Min}=1$ ，最大值 $\text{Max}=1000$ 。可开辟一块包含至少 1000 个位的连续内存，将该连续内存全部置“0”。然后针对各个待排序记录的关键字将有关位置“1”。这样处理后得到的部分连续内存如图 3 的上部所示。

注意到，由于排序对象是记录，对关键字进行排序时，还应对关键字所对应的记录进行适当的操作，以记录它们之间的对应关系。在根据本发明的一种优选实施方案中，为实现这一点，可开辟另一块连续内存（为方便叙述，在下文中将该另一块连续内存称为第二连续内存，将前面提到的连续内存称为第一连续内存），用于存放记录的地址，并将第一和第二连续内存对应起来。目前，记录的地址一般是 32 位的。鉴于此，使第二连续内存的每 32 位依次与第一连续内存中的 1 个位对应：第二连续内存中的第一个 32 位与第一连续内存中的第一位对应；第二连续内存中的第二个 32 位与第一连续内存中的第二位对应；...；第二连续内存中的最后一个 32 位与第一连续内存中的最后一位对应。在开辟第二连续内存时，通过使第二连续内存的位数是第一连续内存的位数的 32 倍，即可实现两块连续内存之间的这种对应。

现在详细说明如何保持关键字与相应记录之间的对应关系。当将第一连续内存中的第 m 位置“1”时，将该置“1”操作所针对的特定关键字所对应的记录的地址存放在第二连续内存中的第 m 个 32 位中。例如，当将图 3 上部所示的第一连续内存中的第 4 个位置“1”时，将该置“1”操作所针对的关键字所对应的记录的地址存放在图 3 下部所示的第二连续内存中的第 4 个 32 位中。需要说明的是，图 3 中的箭头是示意性的，仅表示第一连续内存与第二连续内存之间存在对应关系。针对所有关键字进行上述置“1”操作后，从起始位开始遍历第一连续内存，可还原出各个记录的关键字。具体过程参考上文。需要附加说明的操作是，当基于第一连续内存中的某个被置为“1”的位还原出一个关键字时，还应根据第二连续内存中的、与该被置为“1”的位对应的一个 32 位中存放的地址，将该地址所指向的记录与所还原的关键字对应起来。这样，还原所有关键字并将各个关键字与相应的记录对应起来之后，可释放第一和第二连续内存。

可替换地，可将开辟的第二连续内存用来直接存储记录的内容。

基于本发明的原则，还可设想许多其他应用。下面给出一个例子。目前，在计算机中，整数一般需要用 32 位表示。假设要存储一组互不相同的整数，而该组整数的平均间隔较小，例如，小于 32。在这种情况下，可利用本发明以位流的形式存储这组整数。具体而言，以上文描述的方式，将待存储的各个整数依次映射到一连续存储空间中，得到一个位流，并记录作为该映射操作参考的参考值。这样位流和参考值就共同包含了有关这些整数的信息。以这种方式存储的好处是能节省存储空间。

尽管描述了本发明的优选实施方案，但通过阅读和掌握本发明的原则和教导，本领域的技术人员可对这里公开的实施方案进行各种改型。因而，本发明的范围由附在这里的权利要求限定。

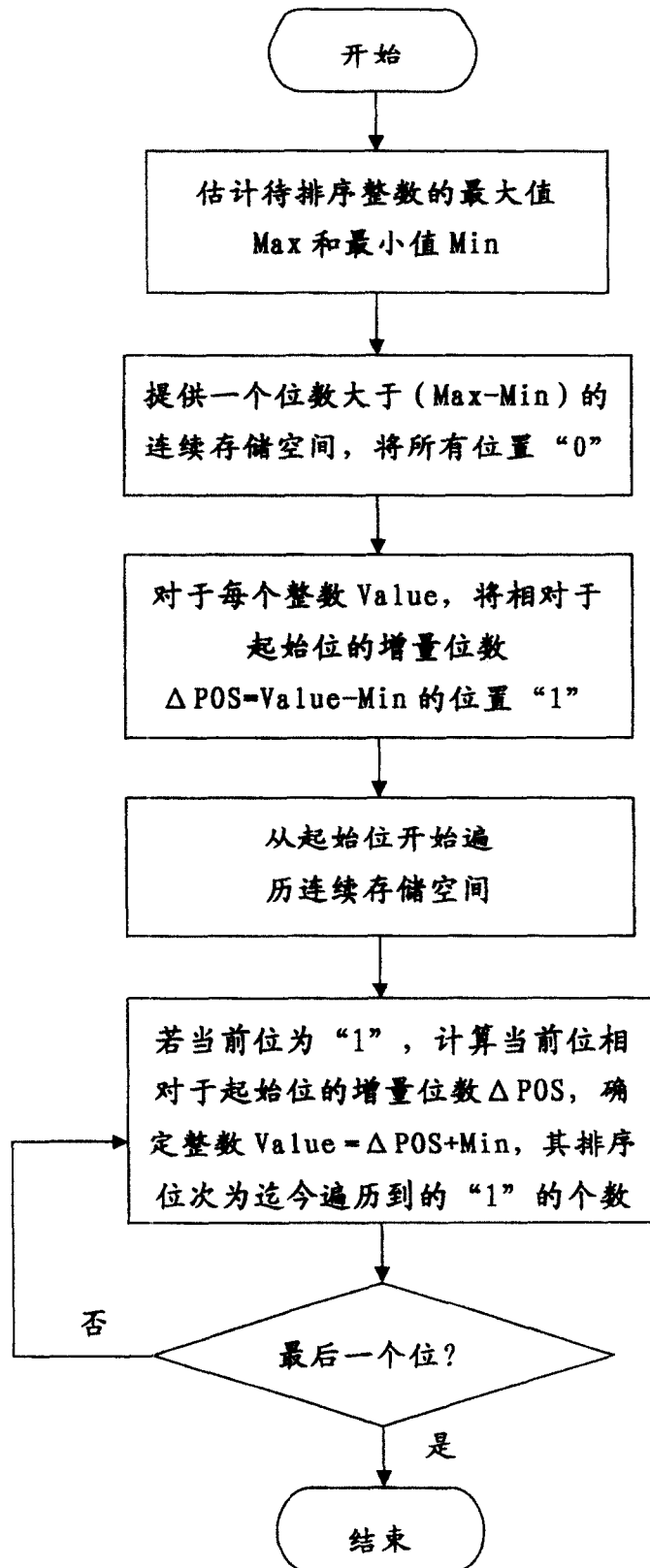


图 1

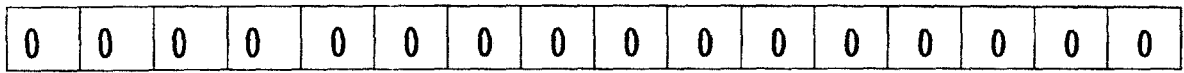


图 2a

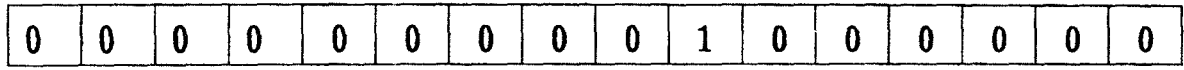


图 2b

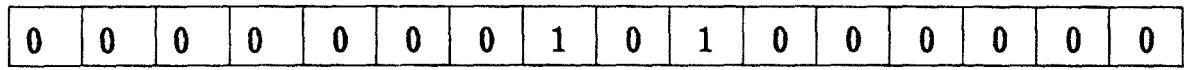


图 2c



图 2d

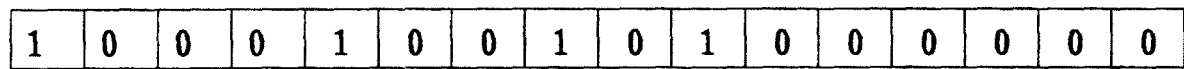


图 2e

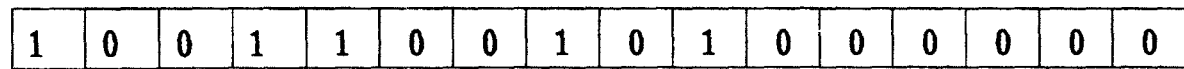


图 2f

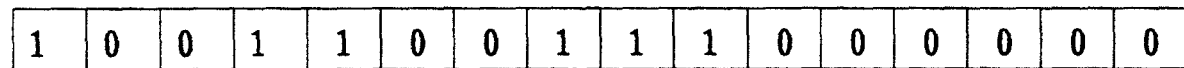


图 2g

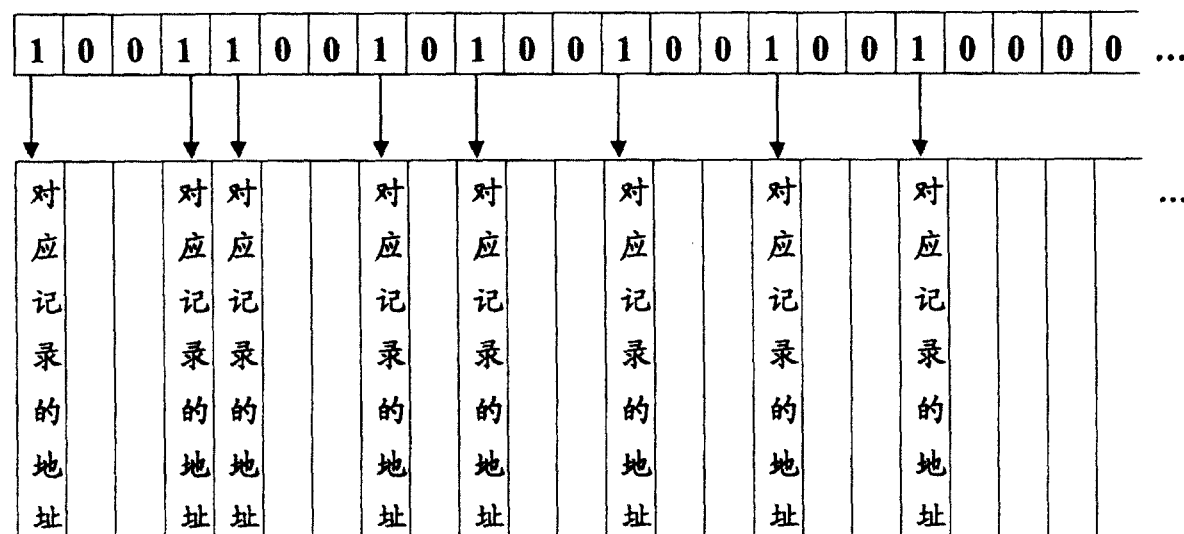


图 3