

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/00 (2006.01)



[12] 发明专利说明书

专利号 ZL 200580033141.2

[45] 授权公告日 2009年12月2日

[11] 授权公告号 CN 100565495C

[22] 申请日 2005.9.29

[21] 申请号 200580033141.2

[30] 优先权

[32] 2004.9.30 [33] US [31] 10/956,482

[86] 国际申请 PCT/US2005/035200 2005.9.29

[87] 国际公布 WO2006/039492 英 2006.4.13

[85] 进入国家阶段日期 2007.3.29

[73] 专利权人 EMC 公司

地址 美国加利福尼亚州

[72] 发明人 R·帕拉波图

[56] 参考文献

US6591269B1 2003.7.8

审查员 李昕宇

[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 刘春元 刘红

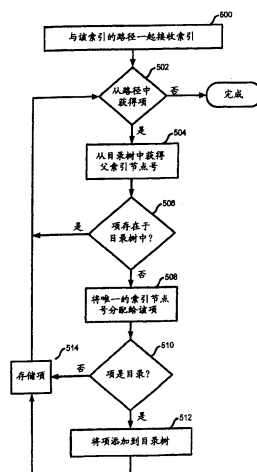
权利要求书 2 页 说明书 5 页 附图 6 页

[54] 发明名称

文件索引处理

[57] 摘要

根据一些实施例，用于处理索引的技术包括：接收索引，其中索引包括路径，并且其中路径包括元素；确定路径元素是否与标识符相关；并且如果路径元素被确定与标识符不相关，则使标识符与路径元素相关联。



1. 处理索引的方法，包括：
接收索引，其中索引包括路径，并且其中路径包括元素；
检查所存储的目录树用以确定路径元素是否与目录树中的标识符相关；
如果路径元素被确定与标识符不相关，则使标识符与路径元素相关联；
如果路径元素是目录，则使用与路径元素相关的标识符来将路径元素添加到目录树；和
在与目录树分开存储的平衡树中存储索引。
2. 根据权利要求1所述的方法，进一步包括把标识符存储到存储器。
3. 根据权利要求1所述的方法，进一步包括把父标识符存储到存储器，其中父标识符与路径元素的父相关。
4. 根据权利要求1所述的方法，进一步包括更新具有与路径元素相关的标识符的列表。
5. 根据权利要求4所述的方法，其中列表被存储在存储器中。
6. 根据权利要求1所述的方法，其中路径包括最后的元素，并且进一步包括：确定最后的元素是否是目录。
7. 根据权利要求6所述的方法，进一步包括：如果最后的元素是目录，则使标识符与最后的元素相关联。
8. 根据权利要求1所述的方法，其中使标识符相关联包括把标识符分配给路径元素。
9. 根据权利要求1所述的方法，进一步包括把索引存储到磁盘。
10. 根据权利要求1所述的方法，其中标识符是唯一的。
11. 根据权利要求1所述的方法，其中标识符是索引节点号。
12. 根据权利要求1所述的方法，进一步包括寻找标识符以及寻找具有父标识符的列表项，其中所述父标识符与所述标识符匹配。
13. 根据权利要求12所述的方法，其中具有匹配父标识符的列表项被标识为索引树的下一级。
14. 处理索引的系统，包括：
用于接收索引的装置，其中索引包括路径，并且其中路径包

括元素；

用于检查所存储的目录树用以确定路径元素是否与目录树中的标识符相关的装置；

用于如果路径元素被确定与标识符不相关则使标识符与路径元素相关联的装置；

用于如果路径元素是目录则使用与路径元素相关的标识符来将路径元素添加到目录树的装置；和

用于在与目录树分开存储的平衡树中存储索引的装置。

15. 根据权利要求 14 所述的系统，进一步包括用于把标识符存储到存储器的装置。

16. 根据权利要求 14 所述的系统，进一步包括用于把父标识符存储到存储器的装置，其中父标识符与路径元素的父相关。

17. 根据权利要求 14 所述的系统，进一步包括用于寻找标识符以及寻找具有父标识符的列表项的装置，其中所述父标识符与所述标识符匹配。

18. 根据权利要求 17 所述的系统，其中具有匹配父标识符的列表项被标识为索引树的下一级。

文件索引处理

技术领域

本发明涉及计算机环境，尤其是数据备份。

背景技术

惯例是在计算机环境中备份数据。当数据被备份到备份介质（例如磁带驱动器、硬驱动器）时，附加信息（这里被称为索引）也可以从正在备份其数据的计算机系统被传送。索引提供关于正在被备份的数据的信息，诸如备份数据如何被组织的。索引经常以随机顺序被接收，并且所接收的索引典型地需要被重组。对于大量的索引而言，在备份结束时对索引的组织可能花费很长时间并且耗费高的系统资源。应该值得期望的是能够以一种避免为重建索引所典型需要的极长的时间周期的方式来组织索引。

发明内容

本发明提供一种用于处理索引的方法，包括：

接收索引，其中索引包括路径，并且其中路径包括元素；

检查所存储的目录树用以确定路径元素是否与目录树中的标识符相关；

如果路径元素被确定与标识符不相关，则使标识符与路径元素相关联；

如果路径元素是目录，则使用与路径元素相关的标识符来将路径元素添加到目录树；和

在与目录树分开存储的平衡树中存储索引。

另外，本发明还提供一种用于处理索引的系统，包括：

处理器，所述处理器包含用于接收索引的装置，其中索引包括路径，并且其中路径包括元素；用于检查所存储的目录树用以确定路径元素是否与目录树中的标识符相关的装置；用于如果路径元素被确定与标识符不相关则使标识符与路径元素相关联的装置；用于如果路径元素是目录则使用与路径元素相关的标识符来将路径元素添加到目录树的装置；和用于在与目录树分开存储的平衡树中存储索引的装置；和

与处理器耦合的存储器，其中存储器为处理器提供指令。

附图说明

在以下的详细描述和附图中公开了本发明的不同实施例。

图 1 是适用于执行本发明的系统的框图。

图 2A - 2B 是根据用于处理索引的一些实施例的方法的流程图。

图 3 是根据一些实施例用于处理索引的方法的流程图。

图 4 是根据本发明的一些实施例的结合图 3 所讨论的具有由索引节点号所标识的索引项的索引树的例子。

图 5 是根据一些实施例的用于索引处理的方法的流程图。

具体实施方式

可以以各种方式实施本分明，包括如过程、装置、系统、合成物质 (composition of matter)、计算机可读介质 (诸如计算机可读存储介质) 或者计算机网络，其中程序指令经由光或者电子通信链路被发送。在该说明书中，这些实施或者本分明可以采取的任何其它形式可以被称为技术。通常，在本分明的范围内可以改变所公开的处理的步骤顺序。

下面，与示出本分明原理的附图一起提供本分明的一个或者多个实施例的详细描述。关于这些实施例描述本分明，但是本分明并不局限于任一实施例。本分明的范围只由权利要求限制并且本分明包含各种替代方案、修改和等价物。在下面的描述中阐明许多特定的细节，以便提供对本分明的彻底理解。为了实例的目的提供这些细节，并且可以在没有一些或者全部这些特定细节的情况下根据权利要求实施本分明。为了清楚的目的，在与本分明有关的技术领域中所公知的技术资料未被详细描述，使得本分明有必要地晦涩。

图 1 是适用于执行本发明的系统的框图。在本例中，数据服务器 100 被示出用以发送数据到备份介质 102。数据服务器 100 也发送索引到备份服务器 104。

图 2A-2B 是根据一些实施例的用于处理索引的方法的流程图。在本例中，索引以块 (piece) 的方式被接收。例如，用于特定数据块的索引也可以有 DIR 组件 (component) 和 NODE 组件，其中 DIR 包括索引节点 (inode) 号、父索引节点号和索引名称。在本实施例中，索引节点号是标识索引的唯一号码。索引的 NODE 组件也包括与 DIR 索引节点

号相匹配的索引节点号、以及索引属性(例如统计信息(stat info.))。统计信息的例子包括关于索引的信息,比如大小、类型、许可、创建和修改日期。

在本例中,索引组件被接收 200。这里所使用的索引组件是索引的一部分,例如 DIR 组件或 NODE 组件。然后确定索引组件的索引节点号是否已经被存储 202。在一实施例中,索引节点号被存储在磁盘上(on-disk)平衡树结构中,如 B+ 树。B+ 树结构可以以与搜索库(诸如 Legato 公司、EMC 的分公司的 WISS)相结合的方式被使用。在一些实施例中,索引节点号可以是标识索引的唯一号码并且该索引的所有索引组件都与该号码相关。

如果这个特定索引的索引节点号未被发现被存储,那么所接收的索引组件作为新的项被存储 208。例如,如果 NODE 组件被接收并且这个特定索引的索引节点号未被发现被存储于磁盘上的平衡树中,那么该所接收的索引组件作为新的项被存储在平衡树中。

如果这个特定索引的索引节点号被发现被存储,那么就确定被接收的索引组件是否是 DIR 204。如果不是 DIR,那么假设是 NODE 组件并且被写入与所有匹配索引节点号的匹配存储索引节点号(matching stored inode number)相同的项中 210。如果被接收到的索引组件是 DIR 204,那么确定匹配存储索引项是否包括 NODE 206。如果不包括,那么被接收到的 DIR 作为新的项被存储 212。

如果匹配存储索引项包括 NODE 206,那么确定匹配存储索引项中的一个是否是 DIR 220。如果带有匹配索引节点号的 DIR 组件还未被存储但是有 NODE 项,那么匹配的 NODE 项用该所接收到的 DIR 组件更新 224。然而,如果匹配存储索引项中的一个 DIR 220,那么 NODE 属性从匹配项之一中被复制并且在接收 DIR 组件时被存储在新的项中 222。

图 3 是根据一些实施例用于处理索引的方法的流程图。在本例中,索引的块已经被重构,作为索引项被存储在平衡树中,并且索引树正在被重建。当与图 4 结合起来被讨论时可以最好地理解图 3。

图 4 是根据一些实施例结合图 3 所讨论的具有由索引节点号所标识的索引的索引树的例子。在本例中,该树被示出包括具有顶端索引的索引结构,所述顶端索引具有索引节点号 2。具有索引节点号 3 和 4 的索引被示出为节点索引号 2 的子;并且索引节点号 5、6、7 和 8 为

索引节点号 3 和 4 的子。索引节点号 2 这里被称为索引节点号 3 和索引节点号 4 的父；并且索引节点号 3 这里被称为索引节点号 5 和索引节点号 6 的父。在本例中，索引节点号 2 可以是一目录，该目录包括具有索引节点号 3 的目录和具有索引节点号 4 的另一目录。索引节点号 3 是一目录，该目录包括带有索引节点号 5 的文件和带有索引节点号 6 的另一文件。

在图 3 所示的实施例中，最顶端的索引节点号从所存储的索引节点号中被寻找 300。例如，可以预先确定树中的最顶端的索引节点号是索引节点号 2。因此，可以在平衡树（例如 B+ 树）中执行搜寻来寻找索引节点号 2。

然后，其父是最顶端的索引节点号的索引项被寻找 302。例如，在图 4 所示的例子中，其父是最顶端的索引节点号的索引是索引节点号 3 和索引节点号 4。因此，可以在平衡树中基于父信息执行搜索来寻找具有等同于索引节点号 2 的父的索引项。在一些实施例中，父索引节点号被存储在索引的 DIR 组件中。

这些索引项被标识为索引树的下一级 304。然后确定是否存在具有作为父的这些索引节点号的项 306。如果不存在具有作为父的这些索引节点号的索引项，那么重构完成。然而，如果存在具有作为父的这些索引节点号的项，那么最新发现的索引项被标识为索引树的下一级 304。

图 5 是根据一些实施例的用于处理索引的方法的流程图。在本例中，索引以单一块而不是以接收索引的多个块的方式从数据服务器被传送到备份服务器。这里，索引节点号不被接收，而是索引的路径被接收。例如，索引 d 可以与路径 a/b/c/d 一起被接收，其中 c 是 d 的父，b 是 c 的父，a 是 b 的父。另外，另一索引 d 与其路径一起可以被接收，其中索引 d 不同于索引 d。具有相同名称的文件可以出现在不同的目录下，例如 a/c/b/d，其中这些文件不具有相同的路径。因此，按名称（例如 d）的搜索在本例中是无用的。在一些实施例中，在路径中为项产生伪索引节点号。在一些实施例中，该索引节点号是唯一的。目录树可以在存储器中被产生，所述目录树包括索引名称、其被分配的索引节点号、以及其父被分配的索引节点号，使得可以通过搜索索引节点号或父索引节点号来执行快速搜索。

在图 5 所示的例子中，索引与该索引的路径（例如 a/b/c/d）一起被接收 500。确定路径项是否从路径中被找回 502。如果路径项不从路径中被找回，那么处理完成。如果路径项（例如“b”）从路径中被找回，那么父索引节点号从目录树中被找回 504。目录树包括目录列表。在一些实施例中，目录树被存储在存储器中而索引树被存储在磁盘上。然后确定路径项是否处于目录树中 506。如果路径项已经存在于目录树中，那么下一项从所接收到的路径中被找回 502。在一些实施例中，路径项从左到右、从父到子被找回，在路径 a/b/c/d 的例子中例如 a，然后 b，然后 c，然后 d。

如果路径项不存在于目录树中 506，那么在该例中唯一的索引节点号被分配给该路径项 508。确定路径项是否是目录 510。在一些实施例中，文件不被保存在目录树中而是目录被保存。目录具有至少一个子而文件没有。如果路径项是目录，那么该路径项被添加到目录树 512。在一些实施例中，目录名称、其被分配的索引节点号和其父被分配的索引节点号被存储在目录树中。例如，如果路径 a/b/c/d 已经为索引“d”被接收，并且“c”已经被分配索引节点号 4，并且“d”被分配索引节点号 5，那么“d”的父索引节点号是 4。不管路径项被加到目录树 512 还是路径项是目录 510，索引被存储在磁盘上的平衡树中 514。如果一个是可用的，那么下一路径项被找回 502。

其后，在一些实施例中根据与图 3 所例示的方法相类似的方法，索引树可以被重建。在该实施例中，搜索存储在存储器的目录树而不是平衡树，以找到索引节点号和父索引节点号。

尽管为了清楚理解的目的已经稍详细地描述了前述的实施例，但是本发明不局限于所提供的细节。有许多实现本发明的可替代方式。公开的实施例是示例性的并且不是限制性的。

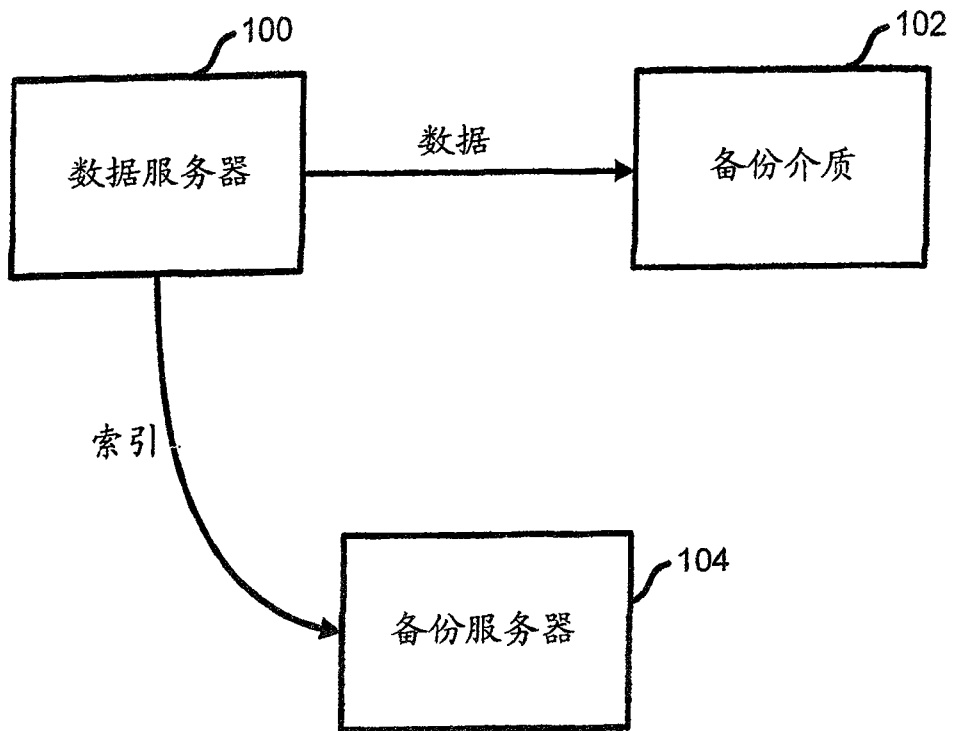


图 1

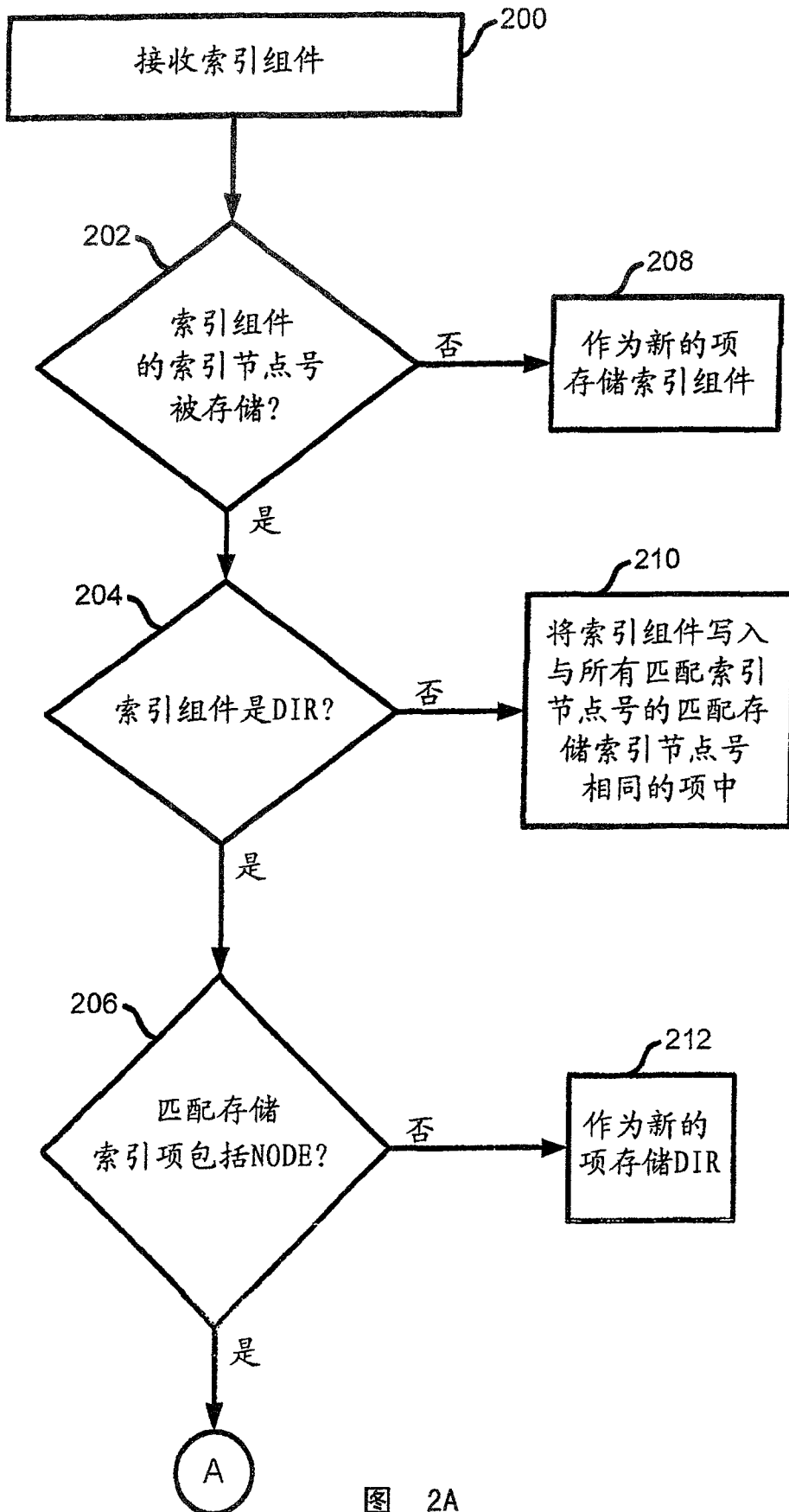


图 2A

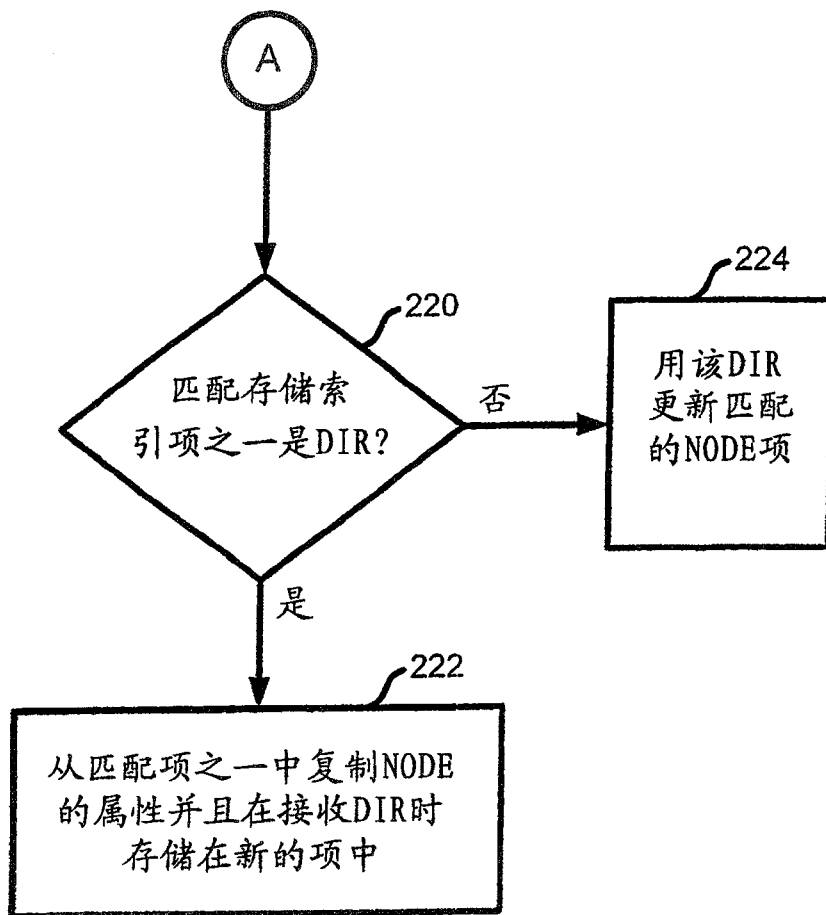


图 2B

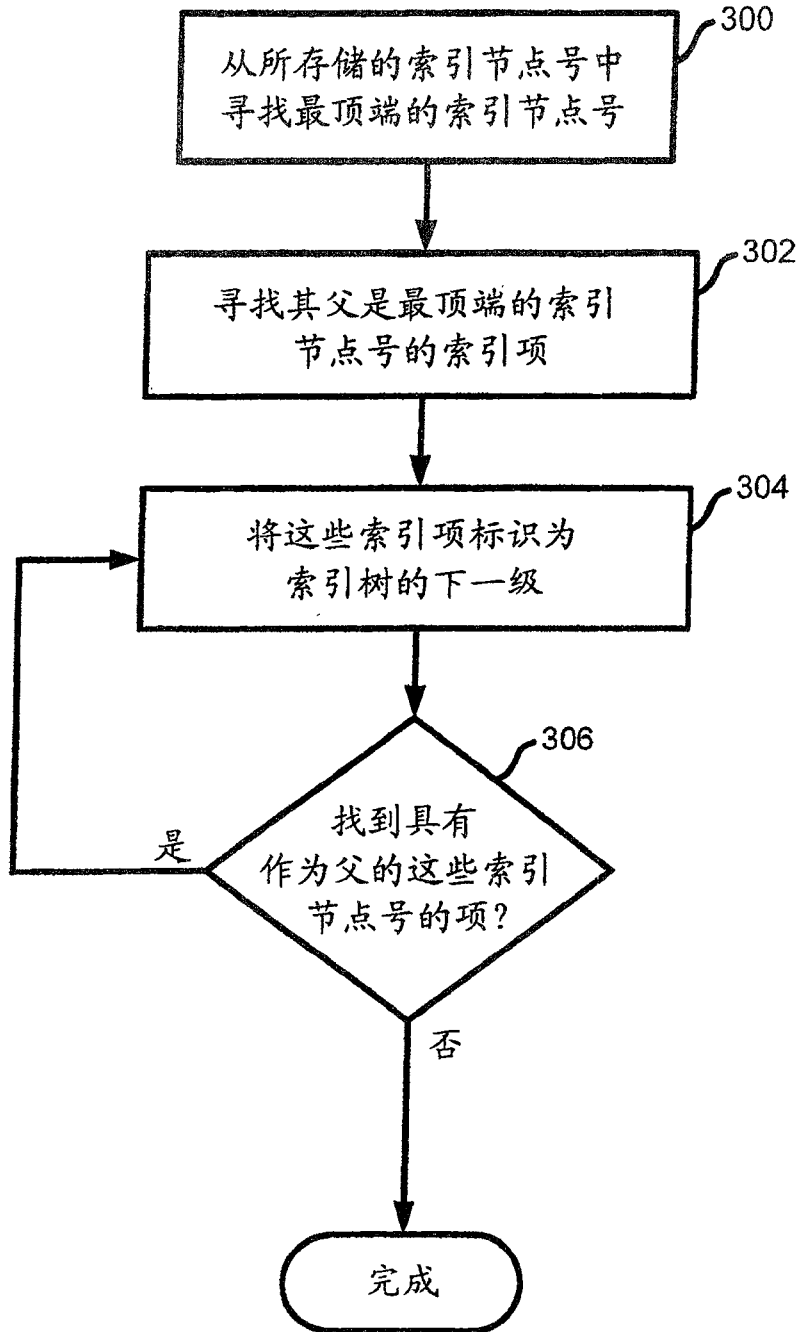


图 3

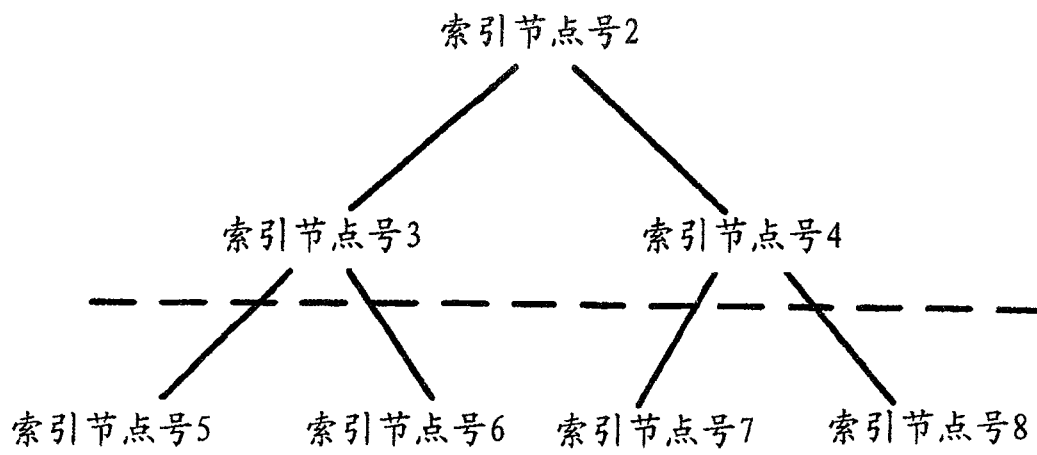


图 4

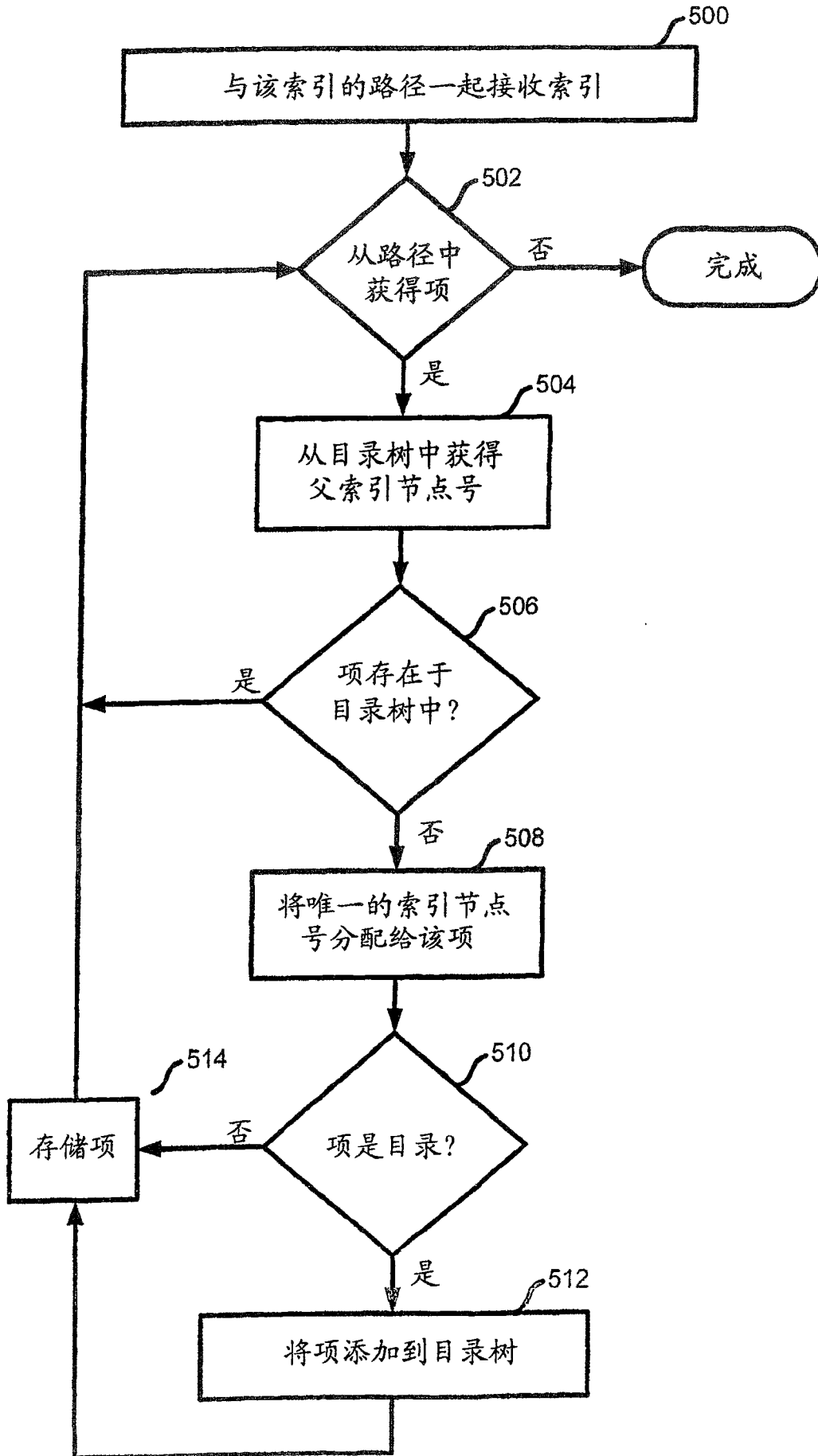


图 5