

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 9/445 (2006.01)

G06F 11/14 (2006.01)



[12] 发明专利说明书

专利号 ZL 200510095938.7

[45] 授权公告日 2009 年 4 月 29 日

[11] 授权公告号 CN 100483347C

[22] 申请日 2005.8.23

[21] 申请号 200510095938.7

[30] 优先权

[32] 2004.8.23 [33] US [31] 10/924,712

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 H·艾瓦库拉 T·米苏

[56] 参考文献

US2003/0009640A1 2003.1.9

US6604123B1 2003.8.5

CN1483164A 2004.3.17

US2003/0177428A1 2003.9.18

US2003/0158842A1 2003.8.21

US2003/0145157A1 2003.7.31

审查员 刘光德

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 张政权

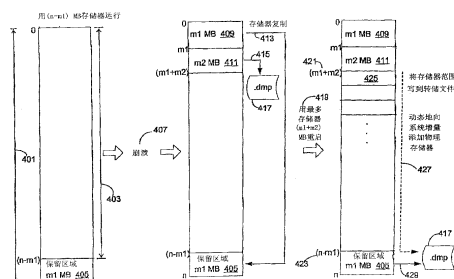
权利要求书 1 页 说明书 7 页 附图 8 页

[54] 发明名称

具有快速重启的存储器转储生成

[57] 摘要

一种用于将计算机的主存储器中的数据映像写到辅助存储设备的方法和系统将执行必需的重启的延迟最小化。在重启之前，物理存储器的一部分被复制到保留存储空间或辅助存储，或者存储器的一部分可被复制到保留存储空间且另一部分被复制到辅助存储中。在重启之后，存储器使用最初被限于在重启之前被复制的那部分存储器。剩余物理存储器被增量地写到存储设备，并作为存储器被动态地添加以供操作系统使用。如果重启是由于崩溃或其它问题，那么在扫描存储器之后确定该问题是已知问题，那么可以取消完全存储器转储的完成。



1. 一种调查要求重启计算机的问题的方法，所述方法包括：

将所述计算机的存储器的第一子集从所述存储器中的第一位置复制到所述存储器中的第二位置；

重启所述计算机；

将所述计算机上运行的操作系统限于所述存储器的第一子集；以及

关于所述存储器的一个或多个单元，对于每个单元：

扫描所述单元以确定所述问题的原因；以及

如果所述问题是已知问题，则取消到存储设备的存储器转储的完成，否则生成关于所述单元的存储器转储。

2. 如权利要求1所述的方法，其特征在于，生成关于所述单元的存储器转储还包括将所述单元中的数据写到所述存储设备，以及动态地添加所述单元以供所述操作系统使用。

3. 如权利要求1所述的方法，其特征在于，取消所述存储器转储的完成还包括使额外的物理存储器可供所述操作系统使用。

4. 如权利要求1所述的方法，其特征在于，复制所述存储器的第一子集还包括将所述存储器的第一子集复制到所述存储器中的所述第二位置中的保留存储器空间中。

5. 如权利要求4所述的方法，其特征在于，还包括，在重启所述计算机之后，将所述保留存储器空间中的数据写到所述存储设备中。

6. 如权利要求1所述的方法，其特征在于，复制所述存储器的第一子集还包括将所述存储器的第二子集复制到存储器转储文件中，并将所述存储器的第三子集写到所述存储设备中。

具有快速重启的存储器转储生成

技术领域

本发明一般涉及计算机系统，尤其涉及一种用于将主存储器数据映像复制到辅助存储，并同时执行必需的重启的延迟最小化的方法和系统。

背景技术

操作系统通常被配置成一旦发生系统崩溃和涉及挂起用户进程和服务的严重故障时即执行存储器转储。存储器转储包括将主存储器的内容以例如存储在硬盘或其它辅助存储介质上的文件的形式复制到辅助存储。在系统崩溃的情形中，存储器转储之后通常必须接着系统的重启。完全存储器转储对于涉及崩溃的问题的分析和纠正以及可靠系统的开发来说是必不可少的资源。

将数据从主存储器写到硬盘是相对缓慢的操作。在完全存储器转储的情形中，系统必须扫描存储器的所有内容并将内容写到辅助存储。因此，生成完全存储器转储的主要缺点是它使得系统承担的“停机时间”的长度，在此期间系统实际上无法用于其它目的。此停机时间取决于板上存储器大小，并且在需要系统重启时还取决于启动存储设备的速度。例如，将 16 千兆字节的存储器写到磁盘需要一个小时以上才能完成。对于具有 64 千兆字节存储器的计算机系统，生成完全存储器转储可能要花费 6 小时之久。

本说明书的附图 1 根据现有技术，提供了用于在系统崩溃后执行完全存储器转储的过程的简化图示。此图示旨在阐明以下所提供的对本发明的详细描述。在图 1 中，表示了时间顺序中的事件，其中时间从左至右增加。最初，系统使用其完全数量的存储器 101 来运行，此处为 n 千兆字节。发生系统崩溃事件 103。执行 n 千兆字节存储器 101 的完全存储器转储（箭头 105 所示），接着该计算机系统进行重启 109。

常规计算机中包括的物理存储器的量一直在稳定增长。此增长是归因于随机存取存储器（RAM）技术中经常的容量增进、64 位处理器技术的可用性、以及典型计算机程序对存储器使用的增长。因此，与完全存储器转储相关联的平均停机时

间也增加了。当前，企业服务器机器被配备成使用多达 32 到 64 千兆字节的板上 RAM。预期到 2007 年，低端和中等范围（mid-range）服务器将能够使用多达 128 千兆字节的 RAM。在此情形中，在崩溃后执行完全存储器转储，随之进行系统重启通常是不切实际的。

存在执行完全存储器转储的替换方法。例如，可以执行诸如操作系统内核空间或分配给特定进程的空间等部分主存储器的转储。但是，一般不可能预先知道崩溃是由于内核模式进程还是由于特定用户进程。此外，在某些涉及系统“冻结”的情形中，不可能生成进程专用的存储器转储。

发明内容

本发明针对提供一种用于以使执行所需重启的延迟最小化的方式将计算机主存储器中的数据映像写到辅助存储设备的方法和系统。在重启之前，对物理存储器的一部分进行复制。重启计算机，并且最初操作系统在其存储器使用方面仅限于在重启之前复制的存储器的该部分。剩余物理存储器被增量地写到存储设备，并动态地作为存储器添加以供操作系统使用。

根据本发明的一个方面，维护一保留的存储器空间。重启之前对物理存储器的一部分所进行的复制包括将该部分复制到所保留的空间。在重启以后，将所保留的存储器空间写到存储设备。

根据本发明的另一个方面，不使用任何保留的存储器空间。重启之前对物理存储器的一部分所进行的复制包括将该部分写到辅助存储设备。

根据本发明的又一个方面，重启之前所复制的物理存储器包括存储器中被复制到所保留的存储器空间的一部分，以及被写到辅助存储设备的另一部分。

本发明还针对一种用于调查诸如系统崩溃事件等要求重启的问题的方法。在重启之后，扫描存储器映像并增量地生成存储器转储。但是，如果在扫描期间判定该问题是已知问题，或者不要求完全存储器复制的完成，则可取消存储器转储的完成。

附图说明

图 1 根据现有技术，图示了在系统崩溃后生成完全存储器转储并接着进行重启。

图 2 根据本发明的一个实施例，图示了在系统崩溃后的存储器转储的生成，

其中使用了保留的存储器区域以允许在重启后执行存储器转储。

图 3 根据本发明的一个实施例，图示了在系统崩溃后的存储器转储的生成，其中在重启之前执行部分存储器转储，并且不使用保留的存储器区域。

图 4 根据本发明的一个实施例，图示了在系统崩溃后的存储器转储的生成，其中结合了图 2 和图 3 中所示的实施例的特征。

图 5 是根据本发明的一个实施例，示出对应于图 2 所示的实施例的过程的步骤的流程图。

图 6 是根据本发明的一个实施例，示出对应于图 3 所示的实施例的过程的步骤的流程图。

图 7 是根据本发明的一个实施例，示出对应于图 4 所示的实施例的过程的步骤的流程图。

图 8 是根据本发明的一个实施例，示出一种用于调查要求重启的问题的方法的步骤。

具体实施方式

本发明提供一种在要求重启的情况中减少与将大容量存储器复制到辅助存储的执行相关联的停机时间的技术。根据本发明，存储器转储的完成被推迟到发生系统重启之后，因而使得系统能够在诸如崩溃等事件后相对快速地重启。复制了初始存储器区域，因而释放了有限的存储器空间以供操作系统在重启之后使用。在重启之后，操作系统最初用此有限量的物理存储器进行操作。存在于操作系统控制之外的存储器空间被扫描并被写到存储器转储文件或类似文件。在每个离散的子范围扫描和写入之后，操作系统将对应量的物理存储器添加到系统。如果重启是由于系统崩溃或其它问题，则可以在重启之后当扫描存储器映像时立即开始调查问题的原因。如果判定该问题是由于已知问题或者具有简单的解决方法，则可以取消完全存储器转储的完成。

现在转到附图，图 2、3 和 4 示出本发明的各实施例。如图 1 中所示，这些附图按时间顺序描述时间，其中时间从左至右增加。应当注意，在此描述及附图中，将存储器的各段描述和描绘成位于诸如物理存储器完整范围的开头或结尾等物理存储器的特定区域仅仅是出于说明简单的目的。

如图 2 中所示，在本发明的一个实施例中，将计算机系统表示为配备了 n 兆字节的物理存储器 201，但是在崩溃事件之前，它用 $(n-m)$ 兆字节存储器 203 运行。

m 兆字节的存储器区域为“保留”空间 205。在一个实施例中，保留区域 205 的大小可以被配置成多达总板上存储器大小的一半。

在发生崩溃事件 207 之后，存储器 203 的最初 m 兆字节 209 被复制到（箭头 211 所示）保留区域 205。这一复制可相对快速地完成，因为其涉及将数据转出或转入主存储器。发生重启 213，并且系统被配置成使用存储器的 m 兆字节 209。该系统使用一种诸如在某些体系结构上设置私有页目录条目（PDE）等机制，以访问操作系统未识别的物理存储器范围。在一个使用私有 PDE 的实施例中，使用了至少为 4 兆字节的页大小。

在重启 213 之后，崩溃之前的存储器映像仍然保留在系统存储器中。如下文提及的，在某些实施例中，这要求 BIOS 中的改造。系统将保留区域 205 中的存储器数据写到存储器转储文件 215；写操作如箭头 225 所示。系统增量地扫描从 m MB 217 到(n-m) MB 219 范围内的存储器数据并将其写到存储器转储文件 215。存储器的扫描、写和添加由虚线箭头 223 所示。当扫描了一个存储器单元 221（例如，该范围内的每 256 MB 或 512 MB）并写到存储器转储文件 215，操作系统动态地将此现在可用的物理存储器添加到系统中。

某些在物理地址扩展（PAE）或类似的存储器寻址机制可用且由操作系统支持的体系结构上实现的实施例使用虚拟保留空间，并且将虚拟保留空间映射到物理保留空间 205。在分区的 NUMA 多处理器系统上实现的实施例中，根据系统的物理存储器布局，每个节点都有保留区域 205。相关拓扑信息可以在启动时间通过诸如由固件构造的静态资源仿射表等方式来确定。

如图 3 中所示，在本发明的另一个实施例中，不使用任何保留的存储器区域。最初，系统用其全部 n 兆字节物理存储器 301 运行。在发生崩溃事件 303 之后，存储器 301 的最初 m 兆字节 305 被复制（如箭头 307 所示）到存储器转储文件 309。然后发生重启 311，且系统被配置成使用多达存储器的 m 兆字节 305。在重启 311 之后，除了存储在 m 兆字节 305 中、在重启 311 之前被写（307）到转储文件 309 的数据之外，崩溃前（pre-crash）的存储器映像仍保留在系统存储器中。系统增量地扫描从 m MB 313 到 n MB 315 范围内的存储器数据，并将其写到存储器转储文件 309。如前述实施例中所述，当扫描了一个存储器单元 317 并将其写到存储器转储文件 309，操作系统动态地将此现在可用的物理存储器添加到系统中。存储器的扫描、写和添加由虚线箭头 319 示出。

区别在于，图 3 中所示的实施例将比之前参考图 2 所描述的实施例需要更多

时间来完成存储器转储过程。在图3所示的实施例中,在发生崩溃后,部分存储器空间被写到磁盘。但是,此方法比图1中所示的现有技术方法需要少得多的时间,在现有技术方法中,在重启之前要将整个存储器空间写到磁盘。此外,图3中所示的实施例比图2中所示的一般较快的实施例更节约地使用存储器,因为不使用保留区域。

现在转到图4,示出包括图2和3中所示的实施例的特征的本发明的第三个实施例。所描绘的系统配备了 n 兆字节的存储器401,但最初它用存储器的 $(n-m_1)$ 兆字节403运行,并且它保持 m_1 兆字节的保留存储器区域405。在发生崩溃事件407之后,存储器403的最初 m_1 兆字节409被复制(如箭头413所示)到保留区域405,并且存储器403另外的 m_2 兆字节区域411被写到存储器转储文件417,如箭头415所示。随即发生重启419,且系统被配置成使用多达存储器的 (m_1+m_2) 兆字节409、411。在重启419之后,保留区域405中的数据被写到存储器转储文件417;写操作由箭头429表示。系统增量地扫描从 (m_1+m_2) MB 421到 $(n-m_1)$ MB 423范围内的存储器数据并将其写到存储器转储文件417。存储器的扫描、写和添加由虚线箭头427表示。当扫描了一个存储器单元425并将其写到存储器转储文件417时,操作系统动态地将此现在可用的物理存储器添加到系统中。

图5、6和7的流程图更明确地将前述的各个实施例示为可通过计算机实现的过程。图5示出与图2中所示的实施例相关联的方法。在步骤501,计算机用 m 兆字节的保留存储器区域启动。在崩溃(步骤503)之后,在步骤505,将 m 兆字节的存储器(在此例中,存储器的前 m 兆字节)复制到保留区域。在步骤507,计算机用多达 m 兆字节存储器重启。在步骤509,保留区域中包含的数据被写到存储器转储文件。在步骤511,操作系统扫描剩余物理存储器中的数据映像并将其写到转储文件(不包括保留区域、或被复制到保留区域的 m 兆字节区域)。系统增量地添加物理存储器(例如,每256或512兆字节所扫描和写入的存储器)。

图6示出与图3中所示的实施例相关联的方法,在此方法中不使用保留区域。在系统用其可用的 n 兆字节物理存储器启动之后,在步骤601发生崩溃。在步骤603,存储器的前 m 兆字节被写到存储器转储文件。在步骤605,系统用多达 m 兆字节的存储器重启。在步骤607,操作系统扫描剩余物理存储器中的数据映像并将其写到转储文件中。系统增量地添加物理存储器。

图7示出与图4中所示的实施例相关联的混合方法。在步骤701,用 m_1 兆字节保留区域启动计算机。在发生崩溃(步骤703)之后,在步骤705,存储器的前

m1 兆字节被复制到保留区域。在步骤 707, 存储器的另外 m2 兆字节被写到存储器转储文件。在步骤 709, 计算机用多达(m1+m2)兆字节的存储器重启。在步骤 711, 保留区域中所包含的 m1 兆字节数据被写到转储文件。在步骤 713, 系统扫描剩余物理存储器中的数据映像, 并将其写到转储文件中(不包括系统用来启动的(m1+m2)兆字节区域, 以及 m1 兆字节保留存储器区域)。系统增量地添加物理存储器。

如图 8 的流程图所示, 本发明的各个实施例提供一种调查诸如系统崩溃等要求重启的问题的方法。根据本发明, 在某些情况下, 可以避免重启之后的完全诊断存储器转储。在图 8 的步骤 801, 发生崩溃事件。在步骤 803, 如上所述, m 兆字节的存储器被复制到保留存储器空间、到存储器转储文件, 或者 m1 兆字节子集被复制到保留存储器空间, 且 m2 兆字节子集被写到转储文件。在步骤 805, 操作系统使用多达 m 兆字节存储器重启系统。如果使用保留存储器空间, 则保留空间被写到转储文件(步骤 807)。在步骤 809, 开始扫描存储器映像。如果在扫描期间, 确定需要重启的问题是已知问题或很容易解决的问题(判定框 811), 则可取消完全存储器转储的完成(例如, 由系统管理员取消)(步骤 815)。否则如上述增量地生成完全存储器转储(步骤 813)。因此, 通过使用此技术, 不用等到生成完全存储器转储之后再调查问题, 而是可以在重启后立即开始。对于问题为已知问题的那些情形, 重启后系统以有限存储器和较慢性能运行的时间段相应地被最小化了。

从以上描述可以看出, 提供了一种新颖并且实用的发明, 允许在发生崩溃或类似事件、并且需要对具有很大物理存储器大小的系统执行完全存储器转储的情况下快速重启。在重启之后, 系统相对运行较慢, 因为它用减少的存储器进行操作, 并且因为它正在生成存储器转储。但是, 本发明向系统管理员提供了在停机事件和重启后初始系统性能间设置合乎需要的平衡的能力, 这是现有技术中不可用的选择。例如, 如果系统被配置成使用一半板上存储器, 那么在崩溃和重启之后, 系统用期望的存储器大小进行操作, 同时立即生成存储器转储。如果使用保留区域, 则可以设置保留区域的大小以容纳重启后所需要的服务。

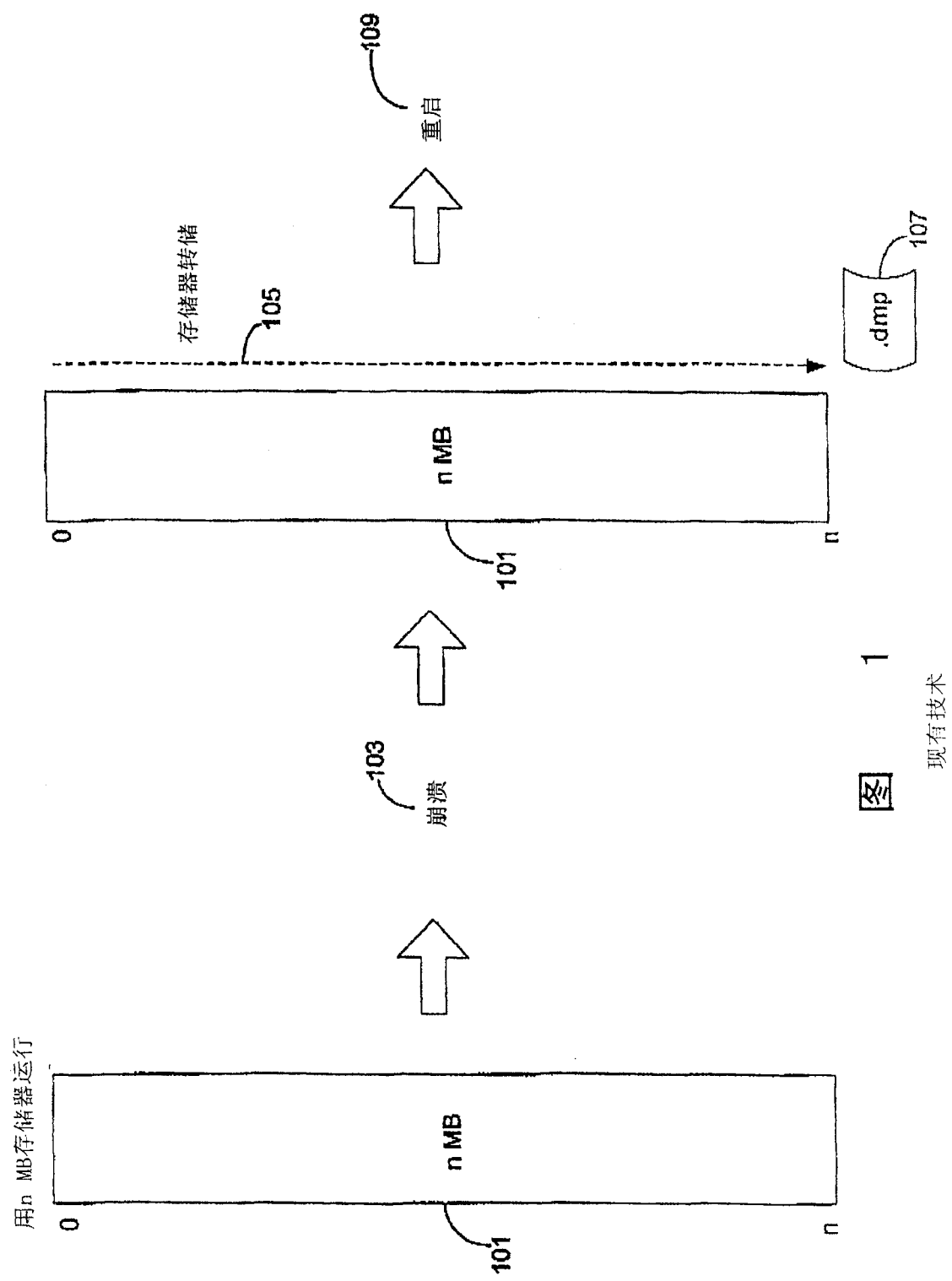
本发明的各个实施例在启动时间执行通常嵌入在板上 ROM 或闪存上的基本输入/输出系统(BIOS)固件等的计算机系统上实现。在常规系统中, BIOS 包括在加载操作系统之前测试和初始化计算机硬件的加电自检(POST)作为其第一个阶段。POST 一般包括系统 RAM 的读/写测试。因此, 在计算机的热重启或复位的情形中, 物理存储器的数据内容被盖写。因此, 本发明的实施例包括这样的实施例, 其中以在崩溃或需要存储器转储的其它事件后的重启期间不进行存储器测试的方

式来配置 BIOS，从而存储器的内容在整个启动期间都被保持，因而允许将存储器转储的完成延迟到重启之后。本发明的其它方面可以通过对常规 BIOS 定义的改变来达成。

本发明可以通过各种通用或专用计算系统来执行，或可包括或位于各种通用或专用计算系统内部，各种通用或专用计算系统包括，但不限于服务器、工作站或个人计算机，并包括单处理器和多处理器机器。在多处理器系统的情形中，如上所述，NUMA 以及 UMA 体系结构是合适的。此类计算系统的一般细节及其各种配置对本领域技术人员来说是很基本的。本发明尤其适用于具有相对大量主存储器以及系统内部或外部的一个或多个辅助存储设备（例如经由高速光纤通道连接的外部存储）的计算机系统。

以上将本发明描述成在发生崩溃事件并且出于诊断目的需要完全存储器转储的设置中实施。但是，本发明可以更宽泛地适用于需要将大量存储器写到辅助存储设备、并且需要重启、但是将重启延迟到存储器写完成之后是不合需要的其它情况。例如，可将数据库从硬盘映射到主存储器，以获取数据库管理系统较佳的性能。但是，在必须快速重启系统的情况下，必须将数据库存储器映像保存到磁盘上。本发明提供一种能将数据写到磁盘而没有不合需要的重启延迟的方法和机制。

本文描述了本发明较佳的实施例，包括发明人所知的用于实现本发明的最佳模式。考虑到可应用本发明原理的许多可能的实施例，应当认识到，本文所描述的实施例仅旨在说明性的，而不应将其理解为限制本发明的范围。本领域技术人员应认识到，可在安排和细节上修改所描述的实施例，而不会偏离本发明的精神。因此，如本文中所描述的本发明构想所有此类实施例都落入所附权利要求书及其等效技术方案的范围之内。



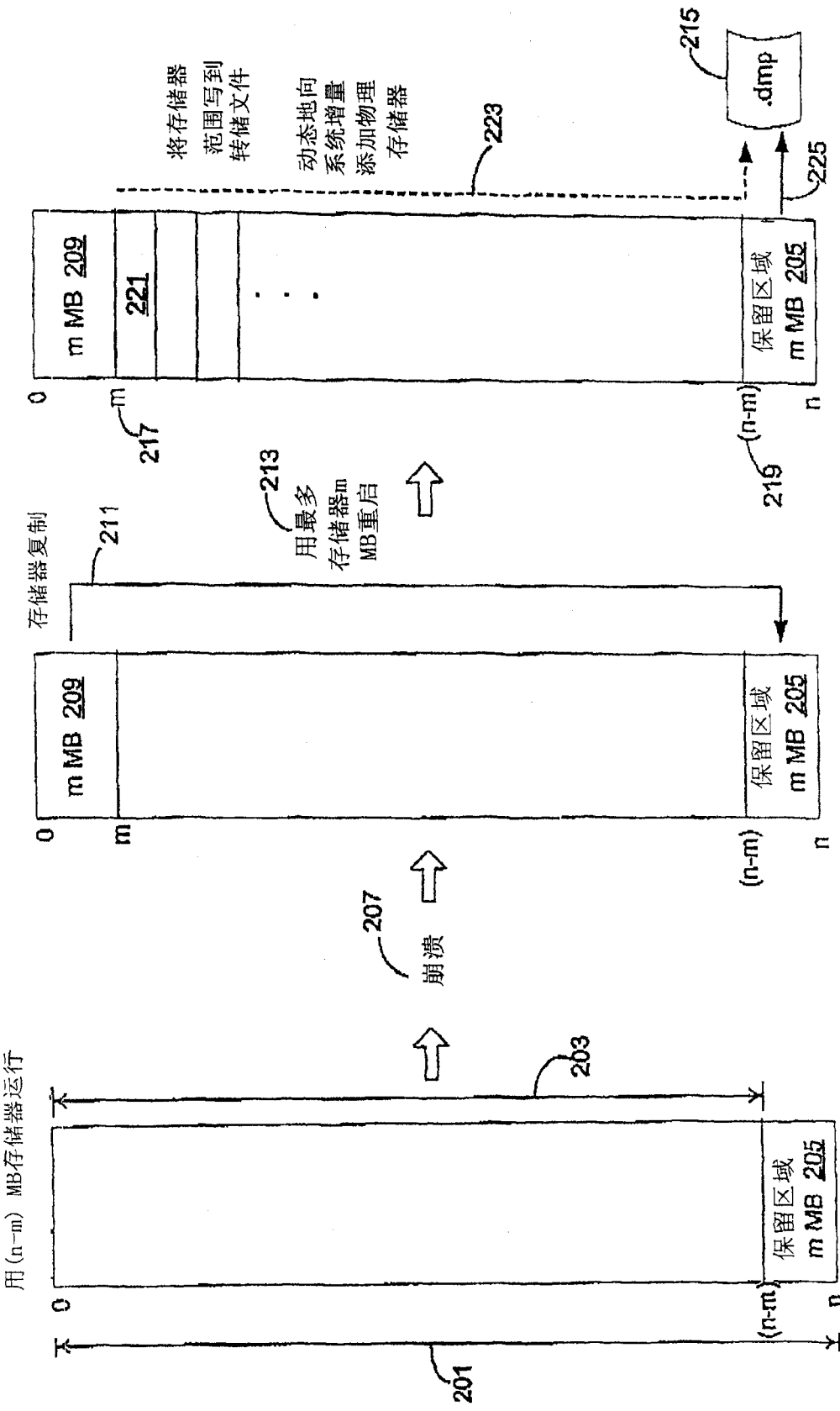


图 2

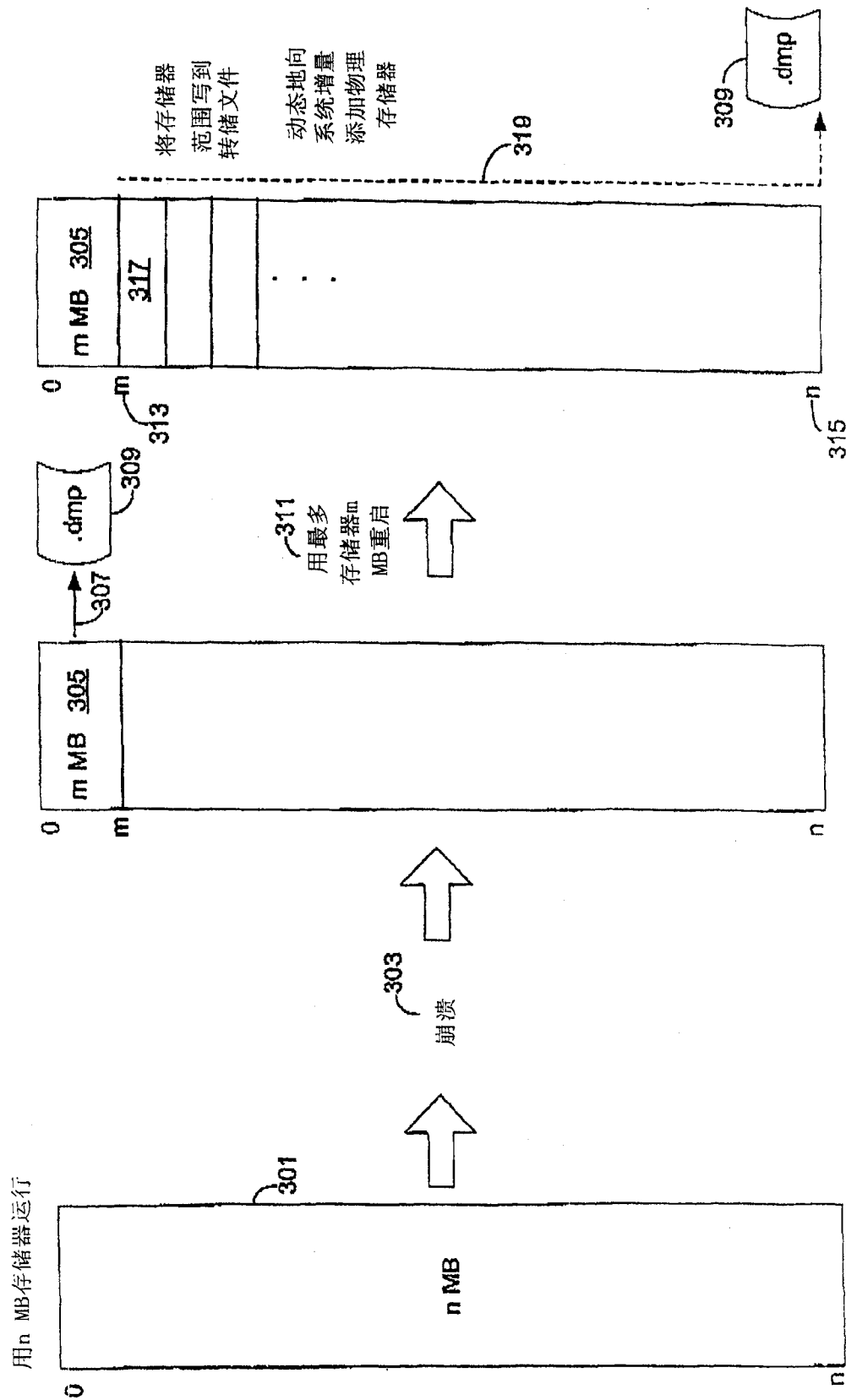


图 3

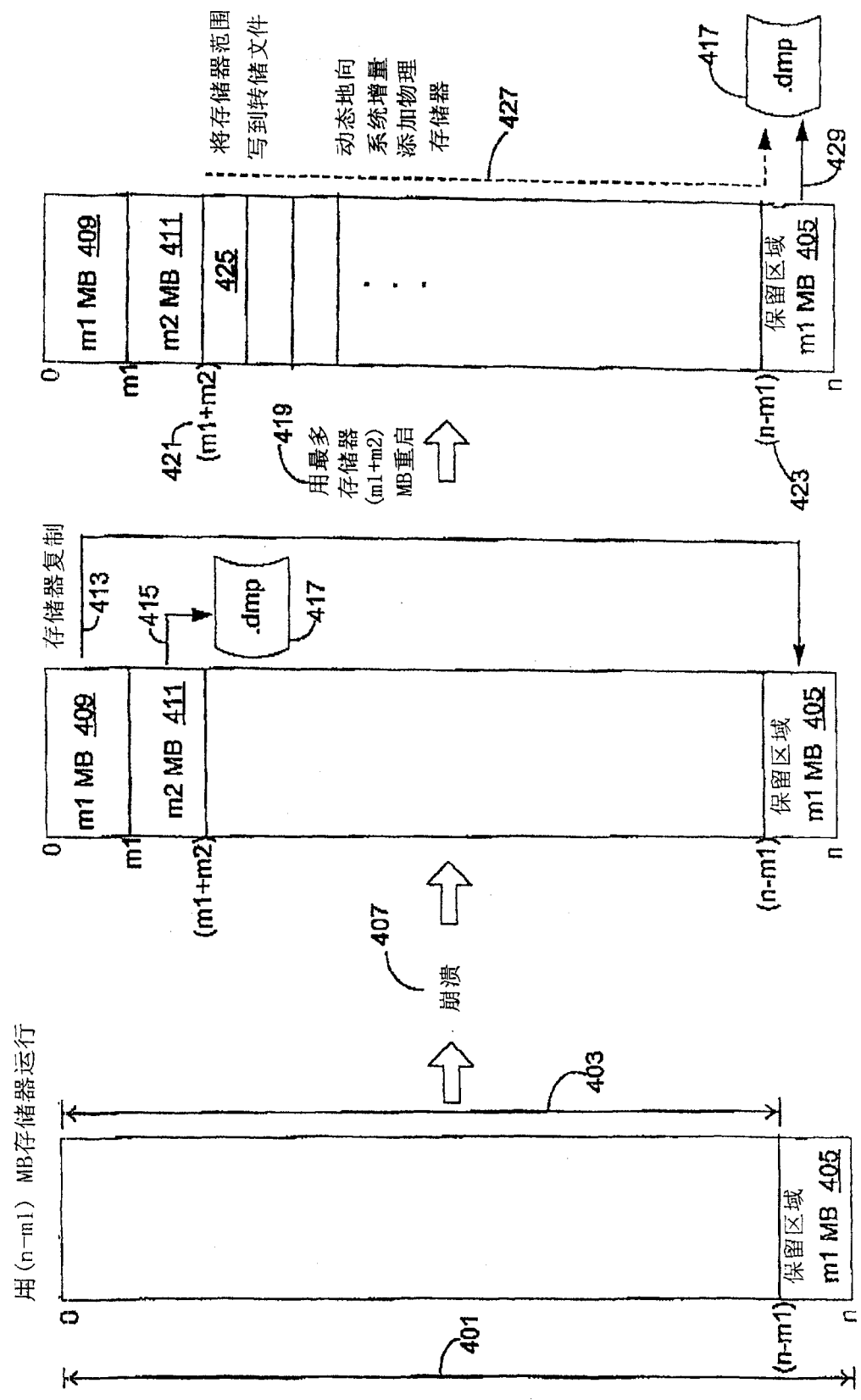


图 4

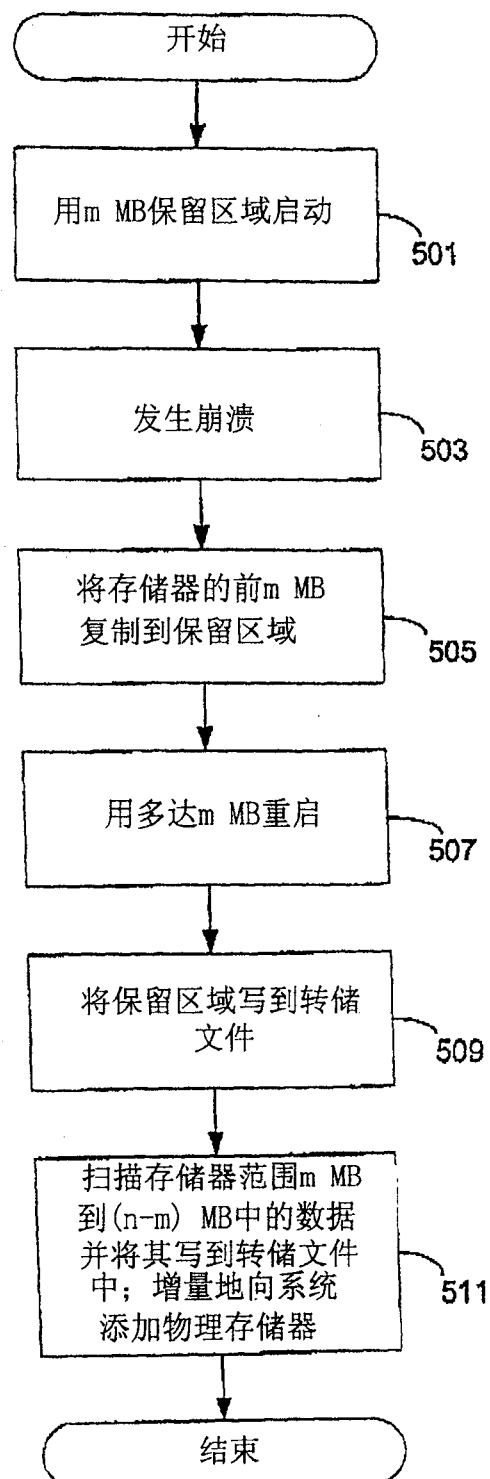


图 5

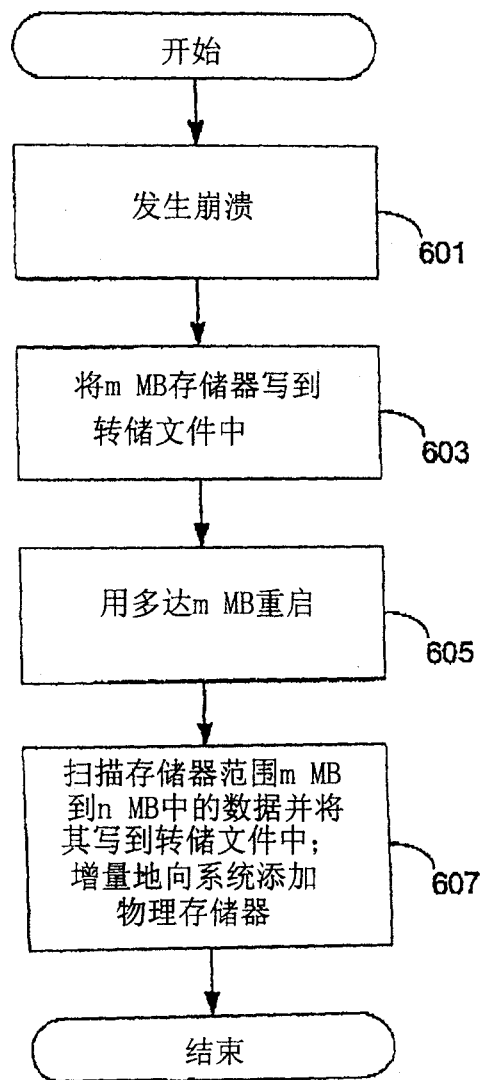


图 6

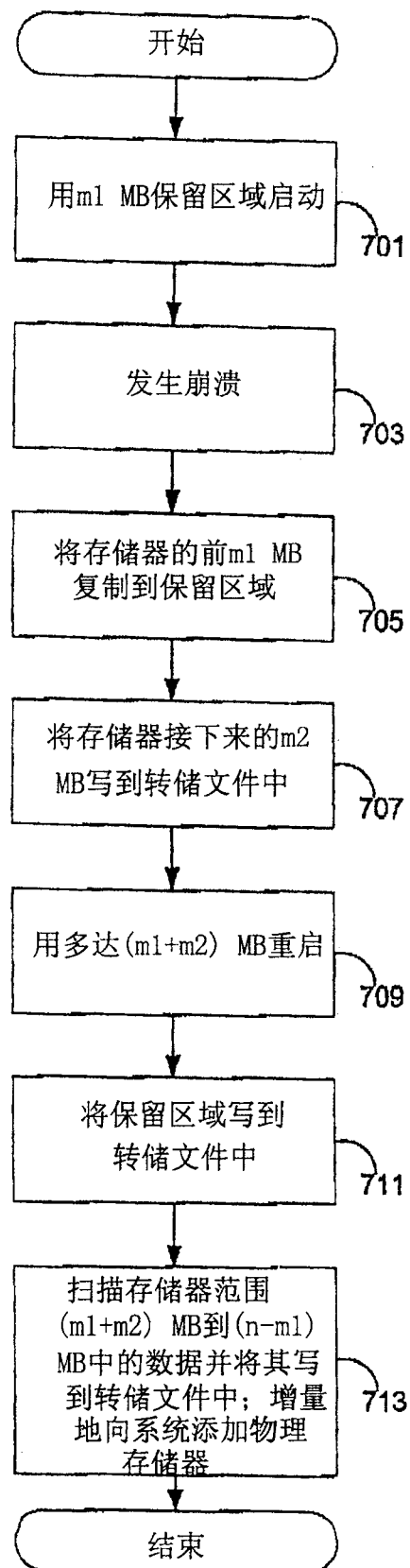


图 7

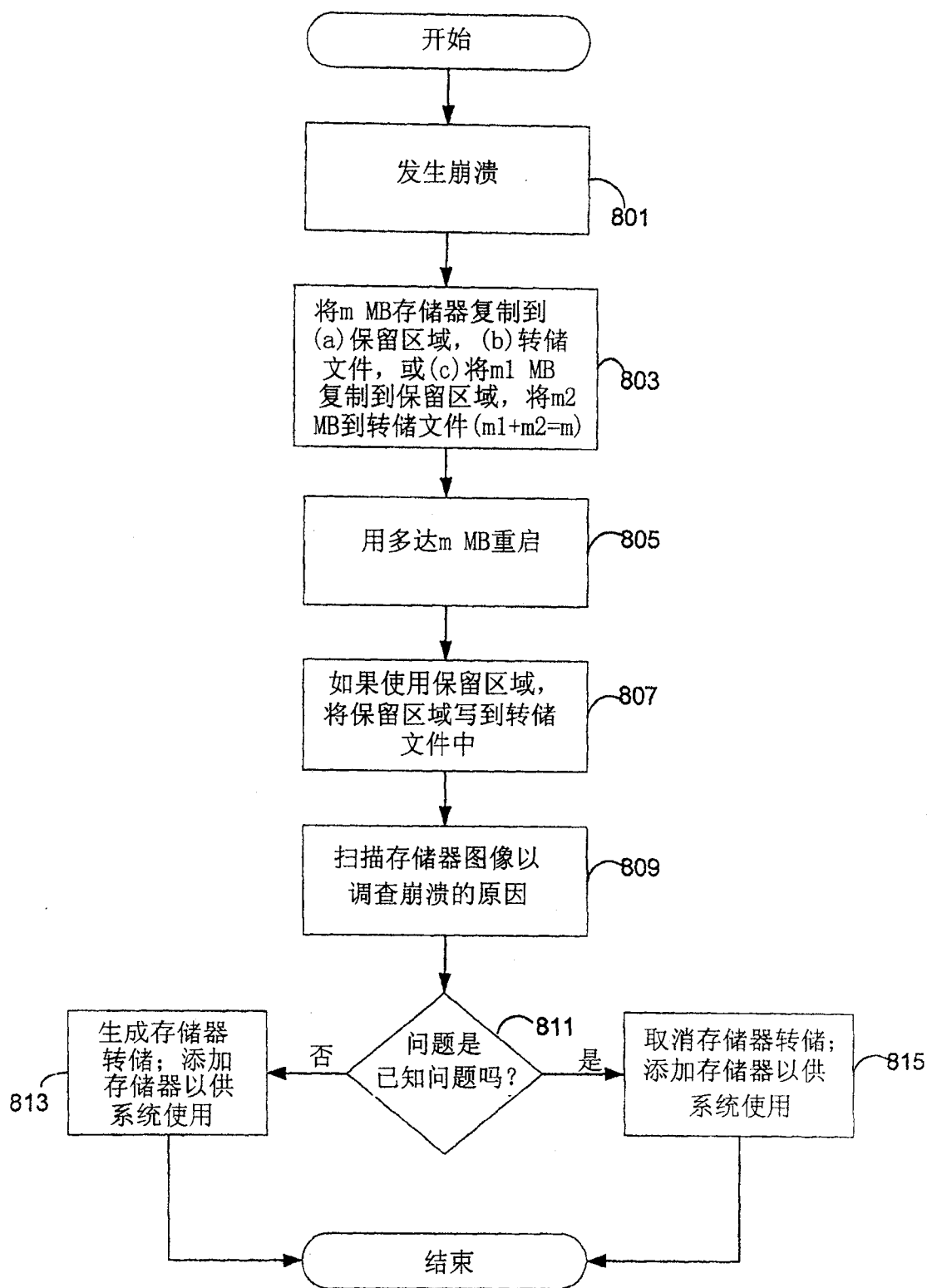


图 8