



(12) 发明专利

(10) 授权公告号 CN 103677656 B

(45) 授权公告日 2016. 07. 06

(21) 申请号 201310052357. X

(22) 申请日 2013. 02. 06

(30) 优先权数据

13/600, 644 2012. 08. 31 US

(73) 专利权人 株式会社日立制作所

地址 日本东京都

(72) 发明人 出口彰

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 鄢迅

(51) Int. Cl.

G06F 3/06(2006. 01)

审查员 鱼冰

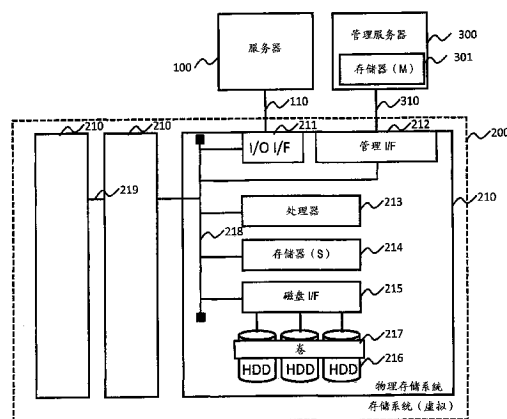
权利要求书4页 说明书14页 附图30页

(54) 发明名称

虚拟存储系统和远程复制系统的管理设备与系统

(57) 摘要

示例性的实施方式提供了管理包括远程复制系统的存储系统并且通过使复杂的操作进行自动化来改进可管理性的技术。在一个实施方式中, 计算机包括存储器和控制器。控制器可操作以: 管理要向服务器提供的虚拟卷; 管理从多个存储系统提供的多个逻辑卷; 管理对虚拟卷所要求的条件, 该条件与要发送到虚拟卷的数据的存储位置有关; 管理多个逻辑卷中的每个逻辑卷的位置信息, 逻辑卷的位置信息是基于逻辑卷的位置来限定的; 以及基于虚拟卷的条件和逻辑卷的位置信息来进行控制, 以将虚拟卷映射到多个逻辑卷中的逻辑卷。



1. 一种计算机,包括存储器和控制器,所述控制器可操作以:

管理从多个存储系统提供的多个逻辑卷;

管理要向服务器提供的多个虚拟卷,每个虚拟卷具有在所述多个存储系统上唯一的标识符;

管理对所述多个虚拟卷中的虚拟卷所要求的条件,所述条件与要发送到所述虚拟卷的数据的存储位置有关,其中所述虚拟卷能够被映射到所述多个逻辑卷;

管理所述多个逻辑卷中的每个逻辑卷的位置信息,逻辑卷的所述位置信息是基于所述逻辑卷的位置来限定的;以及

基于所述虚拟卷的所述条件和所述逻辑卷的所述位置信息来进行控制,以选择所述多个逻辑卷中要被映射到所述虚拟卷的逻辑卷,其中与要发送到所述虚拟卷的数据的存储位置有关的所述虚拟卷的所述条件被所述位置信息中的所述逻辑卷的位置满足。

2. 根据权利要求1所述的计算机,

其中,所述条件要求要发送到所述虚拟卷的数据的所述存储位置所在的站点与作为所述虚拟卷的提供对象的所述服务器所在的站点相同;并且

其中,要发送到所述虚拟卷的数据的所述存储位置是被映射到所述虚拟卷的所述逻辑卷的所述位置。

3. 根据权利要求1所述的计算机,

其中,所述虚拟卷被映射到主逻辑卷,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处。

4. 根据权利要求3所述的计算机,

其中,所述控制器可操作以基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件和所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述主逻辑卷自动地映射到所述辅逻辑卷,其中对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性。

5. 根据权利要求4所述的计算机,

其中,所述控制器可操作以自动地创建用于存储所述主逻辑卷的日志的主日志卷并创建辅日志卷,将所述日志从所述主日志卷复制到所述辅日志卷以及将所述日志从所述辅日志卷复制到所述辅逻辑卷以实现异步远程复制。

6. 根据权利要求3所述的计算机,

其中,所述主逻辑卷被映射到所述辅逻辑卷和第三逻辑卷,所述主逻辑卷、所述辅逻辑卷和所述第三逻辑卷具有远程复制关系并且处于三个不同的位置处;并且

其中,所述控制器可操作以基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件、对所述第三逻辑卷所要求的条件、所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述主逻辑卷自动地映射到所述辅逻辑卷和所述第三逻辑卷,其中,如果所述远程复制是级联的远程复制,则对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性,并且对所述第三逻辑卷所要求的所述条件包括空闲容量和相对于所述辅逻辑卷的连接性,如果所述远程复制不是级联的远程复制,则对所述辅逻辑卷所要求的所述条件包括相对于所述主逻辑卷的连接性。

7. 根据权利要求3所述的计算机,

其中,所述控制器可操作以在所述主逻辑卷出现I/O(输入/输出)故障时将I/O从所述服务器与所述主逻辑卷之间的I/O路径自动地改变为所述服务器与所述辅逻辑卷之间的另一I/O路径;并且

其中,所述辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

8.根据权利要求3所述的计算机,

其中,所述控制器可操作以在所述主逻辑卷出现I/O(输入/输出)故障时将用于运行使用来自所述服务器的数据的应用软件的虚拟机自动地迁移到目的地服务器,并且将I/O从所述服务器与所述主逻辑卷之间的I/O路径自动地改变为所述目的地服务器与所述辅逻辑卷之间的另一I/O路径;并且

其中,所述辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

9.根据权利要求3所述的计算机,

其中,所述控制器可操作以在所述主逻辑卷出现I/O(输入/输出)故障时基于所述逻辑卷的所述位置信息、对复制目的地逻辑卷所要求的条件和所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述辅逻辑卷自动地映射到所述复制目的地逻辑卷作为自动恢复远程复制对,其中对所述复制目的地逻辑卷所要求的所述条件包括空闲容量和相对于所述辅逻辑卷的连接性。

10.根据权利要求1所述的计算机,

其中,所述虚拟卷被映射到具有一个或多个主逻辑卷的主一致性组中的主逻辑卷,并且所述主一致性组被映射到具有一个或多个辅逻辑卷的辅一致性组,所述主一致性组和所述辅一致性组提供一个或多个远程复制对并且处于不同的位置处;并且

其中,所述控制器可操作以基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件、所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述主一致性组中的所述主逻辑卷自动地映射到所述辅一致性组中的辅逻辑卷,其中对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性。

11.一种包括计算机和多个存储系统的系统,所述计算机包括存储器和控制器,所述控制器可操作以:

管理从所述多个存储系统提供的多个逻辑卷;

管理要向服务器提供的多个虚拟卷,每个虚拟卷具有在所述多个存储系统上唯一的标识符;

管理对所述多个虚拟卷中的虚拟卷所要求的条件,所述条件与要发送到所述虚拟卷的数据的存储位置有关,其中所述虚拟卷能够被映射到所述多个逻辑卷;

管理所述多个逻辑卷中的每个逻辑卷的位置信息,逻辑卷的所述位置信息是基于所述逻辑卷的位置来限定的;以及

基于所述虚拟卷的所述条件和所述逻辑卷的所述位置信息来进行控制,以选择所述多个逻辑卷中要被映射到所述虚拟卷的逻辑卷,其中与要发送到所述虚拟卷的数据的存储位置有关的所述虚拟卷的所述条件被所述位置信息中的所述逻辑卷的位置满足。

12.根据权利要求11所述的系统,

其中,所述条件要求要发送到所述虚拟卷的数据的所述存储位置所在的站点与作为所述虚拟卷的提供对象的所述服务器所在的站点相同;并且

其中,要发送到所述虚拟卷的数据的所述存储位置是被映射到所述虚拟卷的所述逻辑卷的所述位置。

13.根据权利要求11所述的系统,

其中,所述虚拟卷被映射到主逻辑卷,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处;并且

其中,所述控制器可操作以基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件和所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述主逻辑卷自动地映射到所述辅逻辑卷,其中对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性。

14.根据权利要求11所述的系统,

其中,所述虚拟卷被映射到主逻辑卷,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处;

其中,所述控制器可操作以在所述主逻辑卷出现I/O(输入/输出)故障时将用于运行使用来自所述服务器的数据的应用软件的虚拟机自动地迁移到目的地服务器,并且将I/O从所述服务器与所述主逻辑卷之间的I/O路径自动地改变为所述目的地服务器与所述辅逻辑卷之间的另一I/O路径;并且

其中,提供所述辅逻辑卷的所述存储系统被配置为将所述辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

15.一种用于管理虚拟存储的设备,包括:

用于管理从多个存储系统提供的多个逻辑卷的装置;

用于管理要向服务器提供的多个虚拟卷的装置,每个虚拟卷具有在所述多个存储系统上唯一的标识符;

用于管理对所述多个虚拟卷中的虚拟卷所要求的条件的装置,所述条件与要发送到所述虚拟卷的数据的存储位置有关,其中所述虚拟卷能够被映射到所述多个逻辑卷;

用于管理所述多个逻辑卷中的每个逻辑卷的位置信息的装置,逻辑卷的所述位置信息是基于所述逻辑卷的位置来限定的;以及

用于基于所述虚拟卷的所述条件和所述逻辑卷的所述位置信息来进行控制以选择所述多个逻辑卷中要被映射到所述虚拟卷的逻辑卷的装置,其中与要发送到所述虚拟卷的数据的存储位置有关的所述虚拟卷的所述条件被所述位置信息中的所述逻辑卷的位置满足。

16.根据权利要求15所述的设备,

其中,所述条件要求要发送到所述虚拟卷的数据的所述存储位置所在的站点与作为所述虚拟卷的提供对象的所述服务器所在的站点相同;并且

其中,要发送到所述虚拟卷的数据的所述存储位置是被映射到所述虚拟卷的所述逻辑卷的所述位置。

17.根据权利要求15所述的设备,还包括:

用于将所述虚拟卷映射到主逻辑卷的装置,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处;以及

用于基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件和所述多个逻辑卷所在的站点之间的距离的站点距离信息将主逻辑卷自动地映射到所述辅逻辑卷的装置,

其中对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性。

18. 根据权利要求15所述的设备,还包括:

用于将所述虚拟卷映射到主逻辑卷的装置,所述主逻辑卷被映射到辅逻辑卷和第三逻辑卷,所述主逻辑卷、所述辅逻辑卷和所述第三逻辑卷具有远程复制关系并且处于三个不同的位置处;以及

用于基于所述逻辑卷的所述位置信息、对所述辅逻辑卷所要求的条件、对所述第三逻辑卷所要求的条件、所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述主逻辑卷自动地映射到所述辅逻辑卷和所述第三逻辑卷的装置,其中,如果所述远程复制是级联的远程复制,则对所述辅逻辑卷所要求的所述条件包括空闲容量和相对于所述主逻辑卷的连接性,并且对所述第三逻辑卷所要求的所述条件包括空闲容量和相对于所述辅逻辑卷的连接性,如果所述远程复制不是级联的远程复制,则对所述辅逻辑卷所要求的所述条件包括相对于所述主逻辑卷的连接性。

19. 根据权利要求15所述的设备,还包括:

用于将所述虚拟卷映射到主逻辑卷的装置,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处;以及

用于在所述主逻辑卷出现I/O(输入/输出)故障时将用于运行使用来自所述服务器的数据的应用软件的虚拟机自动地迁移到目的地服务器,并且将I/O从所述服务器与所述主逻辑卷之间的I/O路径自动地改变为所述目的地服务器与所述辅逻辑卷之间的另一I/O路径的装置。

20. 根据权利要求15所述的设备,还包括:

用于将所述虚拟卷映射到主逻辑卷的装置,所述主逻辑卷被映射到辅逻辑卷,所述主逻辑卷和所述辅逻辑卷是远程复制对并且处于不同的位置处;以及

用于在所述主逻辑卷出现I/O(输入/输出)故障时基于所述逻辑卷的所述位置信息、对复制目的地逻辑卷所要求的条件、所述多个逻辑卷所在的站点之间的距离的站点距离信息将所述辅逻辑卷自动地映射到所述复制目的地逻辑卷以作为自动恢复远程复制对的装置,其中对所述复制目的地逻辑卷所要求的所述条件包括空闲容量和相对于所述辅逻辑卷的连接性。

虚拟存储系统和远程复制系统的管理设备与系统

技术领域

[0001] 本发明一般涉及存储系统,具体而言,涉及远程复制系统和远程复制系统的技术的易用性。

背景技术

[0002] US2007/0079088公开了一种用于向服务器提供具有相互连接的一个或多个物理存储系统的一个虚拟化存储设备的技术。具体而言,该技术通过相互连接实现了其它物理存储系统的资源的使用。此外,在虚拟化的存储系统(一个或多个物理存储系统)中提供了针对存储区域(卷)的唯一ID。唯一的ID被称作全局卷ID。因此,即使在物理存储系统之间移动卷,向服务器提供的卷ID也不会改变。

[0003] US2008/0034005公开了虚拟存储系统。很多存储系统产品具有用于灾难恢复的远程复制功能。远程复制功能将主存储系统中存储的数据复制到辅存储系统。操作者不得不为辅物理存储系统做准备,其包括:在初始化远程复制之前创建复制目的地卷、远程复制路径配置等。

[0004] 第7,680,919号美国专利公开了与虚拟服务器有关的技术。这是用于在一个物理机器上创建一个或多个虚拟服务器的技术。此外,该专利公开了用于在物理服务器之间移动虚拟服务器的技术。

[0005] 传统的虚拟化存储系统中的卷创建是按以下顺序操作的:在物理存储系统中使用本地卷ID进行卷创建,设置全局卷ID以及向服务器提供该卷。因此,操作者必须知道物理存储系统以进行配置。用于自动地确定物理存储区域的技术不是基于指定的全局卷ID来提供的。因此,操作成本很高。此外,因为其未考虑超出数据中心(DC)的虚拟化存储系统,因此不存在选择安装在最佳数据中心中的物理存储系统(例如,安装在其中安装了服务器的相同的数据中心中的物理存储系统)的技术。如果传统的虚拟化存储系统应用于多个数据中心环境,则将会出现超出DC的处理。这使得性能下降并且消耗DC之间的网络资源。

[0006] 传统的远程复制是使用两组存储系统来配置的。因此,当仅向虚拟化的存储系统环境应用传统的远程复制时,远程复制是使用两组虚拟化的存储系统来配置的。在该情况下,对于两组虚拟化的存储系统而言,需要设置远程复制。具体而言,设置包括物理存储系统之间的有线连接、建立I/F(接口)以进行数据传送、创建复制目的地卷、识别复制目的地的卷ID等。如果实现了超出数据中心的虚拟化存储系统环境,则在该环境中,远程复制功能可以用作本地复制功能。此外,如果对于虚拟化的存储系统而言存在连接,则用于远程复制的有线连接和I/F的建立将变得不必要。此外,如果存储系统通过使用虚拟的卷技术自动地创建卷,则创建复制目的地卷将变得不必要。此外,如果使得复制源卷和复制目的地的ID是相同的,则识别复制目的地卷ID将变得不必要。

发明内容

[0007] 本发明的示例性的实施方式提供了管理包括远程复制系统的存储系统以及通过

使复杂的操作自动化来改进可操作性的技术。

[0008] 第一实施方式涉及一种用于在虚拟存储系统中创建卷的方法。具体而言,该技术指派来自物理存储系统的物理存储容量,该物理存储系统安装在其中安装有服务器的相同的数据中心中。

[0009] 第二实施方式涉及一种自动远程复制环境的配置方法。一旦针对服务器使用的卷建立了远程复制属性,存储系统就自动地获取远程复制处理所需的资源,例如,复制目的地卷(辅卷、SVOL)或者存储系统之间的带宽。此外,使得复制源卷(主卷PVOL)的卷ID和提供给服务器的SVOL的卷ID是相同的。当PVOL发生故障时,可以仅通过将I/O路径从PVOL改变为SVOL来重新启动I/O处理。因为不必如传统的技术中一样将I/O处理改变为具有另一个ID的SVOL,因此对应用没有影响。此外,可以使得一个存储配置中的服务器的故障恢复处理和多存储配置中的处理是相同的。

[0010] 第三实施方式涉及远程复制的自动恢复方法和将上述技术应用于三数据远程复制配置的方法。

[0011] 根据本发明的一个方面,计算机包括存储器和控制器。该控制器可操作以:管理要向服务器提供的虚拟卷;管理从多个存储系统提供的多个逻辑卷;管理对虚拟卷所要求的条件,该条件与要发送到虚拟卷的数据的存储位置有关;管理多个逻辑卷中的每个逻辑卷的位置信息,逻辑卷的位置信息是基于逻辑卷的位置来限定的;以及基于虚拟卷的条件和逻辑卷的位置信息来进行控制,以将虚拟卷映射到多个逻辑卷中的任一逻辑卷。

[0012] 在一些实施方式中,该条件要求要发送到虚拟卷的数据的存储位置所在的站点与作为虚拟卷的提供对象的服务器所在的站点相同;以及要发送到虚拟卷的数据的存储位置是被映射到虚拟卷的逻辑卷的位置。虚拟卷被映射到主逻辑卷,主逻辑卷被映射到辅逻辑卷,主逻辑卷和辅逻辑卷是远程复制对并且处于不同的位置处。控制器可操作以基于逻辑卷的位置信息、对辅逻辑卷所要求的条件和多个逻辑卷所在的站点之间的距离的站点距离信息将主逻辑卷自动地映射到辅逻辑卷,其中对辅逻辑卷所要求的条件包括空闲容量和相对于主逻辑卷的连接性。控制器可操作以自动地创建用于存储主逻辑卷的日志的主日志卷并创建辅日志卷,将日志从主日志卷复制到辅日志卷以及将日志从辅日志卷复制到辅逻辑卷以实现异步远程复制。

[0013] 在具体实施方式中,主逻辑卷被映射到辅逻辑卷和第三逻辑卷,主逻辑卷、辅逻辑卷和第三逻辑卷具有远程复制关系并且处于三个不同的位置处。控制器可操作以基于逻辑卷的位置信息、对辅逻辑卷所要求的条件、对第三逻辑卷所要求的条件、多个逻辑卷所处的站点之间的距离的站点距离信息将主逻辑卷自动地映射到辅逻辑卷和第三逻辑卷,其中,如果远程复制是级联的远程复制,则对辅逻辑卷所要求的条件包括空闲容量和相对于主逻辑卷的连接性,并且对第三逻辑卷所要求的条件包括空闲容量和相对于辅逻辑卷的连接性,如果远程复制不是级联的远程复制,则对辅逻辑卷所要求的条件包括相对于主逻辑卷的连接性。

[0014] 在一些实施方式中,控制器可操作以在主逻辑卷出现I/O(输入/输出)故障时将I/O从服务器与主逻辑卷之间的I/O路径自动地改变为服务器与辅逻辑卷之间的另一I/O路径;并且辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

[0015] 在具体实施方式中,控制器可操作以在主逻辑卷出现I/O(输入/输出)故障时将用

于运行使用来自服务器的数据的应用软件的虚拟机自动地迁移到目的地服务器,并且将I/O从服务器与主逻辑卷之间的I/O路径自动地改变为目的地服务器与辅逻辑卷之间的另一I/O路径;并且辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

[0016] 在一些实施方式中,控制器可操作以在主逻辑卷出现I/O(输入/输出)故障时基于逻辑卷的位置信息、对自动恢复辅逻辑卷所要求的条件和多个逻辑卷所处的站点之间的距离的站点距离信息将辅逻辑卷自动地映射到自动恢复辅逻辑卷作为自动恢复远程复制对,其中对自动恢复辅逻辑卷所要求的条件包括空闲容量和相对于辅逻辑卷的连接性。

[0017] 在具体实施方式中,虚拟卷被映射到具有一个或多个主逻辑卷的主一致性组中的主逻辑卷,并且主一致性组被映射到具有一个或多个辅逻辑卷的辅一致性组,主一致性组和辅一致性组提供一个或多个远程复制对并且处于不同的位置处。控制器可操作以基于逻辑卷的位置信息、对辅逻辑卷所要求的条件、多个逻辑卷所在的站点之间的距离的站点距离信息将主一致性组中的主逻辑卷自动地映射到辅一致性组中的辅逻辑卷,其中对辅逻辑卷所要求的条件包括空闲容量和相对于主逻辑卷的连接性。

[0018] 根据本发明的另一个方面,系统包括计算机和多个存储系统,计算机包括存储器和控制器。控制器可操作以:管理要向服务器提供的虚拟卷;管理从多个存储系统提供的多个逻辑卷;管理对虚拟卷所要求的条件,条件与要发送到虚拟卷的数据的存储位置有关;管理多个逻辑卷中的每个逻辑卷的位置信息,逻辑卷的位置信息是基于逻辑卷的位置来限定的;以及基于虚拟卷的条件和逻辑卷的位置信息来进行控制,以将虚拟卷映射到多个逻辑卷中的任一逻辑卷。

[0019] 在一些实施方式中,虚拟卷被映射到主逻辑卷,主逻辑卷被映射到辅逻辑卷,主逻辑卷和辅逻辑卷是远程复制对并且处于不同的位置处。控制器可操作以在主逻辑卷出现I/O(输入/输出)故障时将用于运行使用来自服务器的数据的应用软件的虚拟机自动地迁移到目的地服务器,并且将I/O从服务器与主逻辑卷之间的I/O路径自动地改变为目的地服务器与辅逻辑卷之间的另一I/O路径。提供辅逻辑卷的存储系统被配置为将辅逻辑卷的状态从I/O不可接收改变为I/O可接收。

[0020] 本发明的另一个方面涉及存储用于控制数据处理器以管理数据存储设备的多个指令的计算机可读存储介质。多个指令包括:使得数据处理器管理要向服务器提供的虚拟卷的指令;使得数据处理器管理从多个存储系统提供的多个逻辑卷的指令;使得数据处理器管理对虚拟卷所要求的条件的指令,条件与要发送到虚拟卷的数据的存储位置有关;使得数据处理器管理多个逻辑卷中的每个逻辑卷的位置信息的指令,逻辑卷的位置信息是基于逻辑卷的位置来限定的;以及使得数据处理器基于虚拟卷的条件和逻辑卷的位置信息来进行控制以将虚拟卷映射到多个逻辑卷中的任一逻辑卷的指令。

[0021] 对于本领域普通技术人员来说,根据具体实施方式的以下详细描述,本发明的这些和其它特征和优点将变得显而易见。

附图说明

[0022] 图1示出了在其中可以应用本发明的方法和装置的系统的硬件配置的示例。

[0023] 图2是示出了根据第一实施方式的管理服务器中的存储器(M)的示例的详细框图。

[0024] 图3是在管理服务器中管理的服务器站点管理表的示例。

- [0025] 图4是在管理服务器中管理的存储站点管理表的示例。
- [0026] 图5是示出了根据第一实施方式的物理存储系统中的存储器(S)的示例的详细框图。
- [0027] 图6是在物理存储系统中管理的全局ID管理表的示例。
- [0028] 图7是示出了卷创建处理流程的流程图示例。
- [0029] 图8示出了用于同步远程复制处理的远程复制配置的示例。
- [0030] 图9示出了用于异步远程复制处理的远程复制配置的示例。
- [0031] 图10是示出了根据第二实施方式的物理存储系统中的存储器(S)的示例的详细框图。
- [0032] 图11是示出了根据第二实施方式的管理服务器中的存储器(M)的详细框图。
- [0033] 图12是物理存储系统的存储器(S)中的主存储对表的示例。
- [0034] 图13是物理存储系统的存储器(S)中的辅存储对表的示例。
- [0035] 图14是在物理存储系统中管理的流入表的示例。
- [0036] 图15是在管理服务器中管理的站点距离表的示例。
- [0037] 图16是用于数据复制的设置屏幕的示例。
- [0038] 图17是示出了对创建处理流程的流程图示例。
- [0039] 图18是示出了从图17中的对创建程序调用的辅选择程序的处理流程的流程图示例。
- [0040] 图19是示出了用于异步远程复制的对创建处理流程的流程图的一个示例。
- [0041] 图20是示出了用于计算JVOL所需的容量的JVOL容量程序的处理流程的流程图示例。
- [0042] 图21是示出了辅选择程序的处理流程的流程图的一个示例。
- [0043] 图22是示出了在I/O故障之后的处理的概念图。
- [0044] 图23是示出了在PVOL出现故障之后向SVOL发起的写入处理 的流程图示例。
- [0045] 图24是示出了在PVOL出现故障以后服务器的变化和涉及SVOL的处理的后续变化的概念图。
- [0046] 图25是示出了在PVOL出现故障以后的虚拟机器的迁移处理和用于确定目的地的处理的流程图示例。
- [0047] 图26是示出了远程复制对的一致性组的概念图。
- [0048] 图27是示出了考虑了一致性组的辅选择程序的处理流程的流程图的一个示例。
- [0049] 图28是示出了远程复制配置的自动恢复方法的概念图。
- [0050] 图29是示出了配置恢复程序的处理流程的流程图示例。
- [0051] 图30是示出了三数据中心远程复制的级联配置的概念图。
- [0052] 图31是示出了三数据中心远程复制的多目标配置的概念图。
- [0053] 图32示出了三数据中心远程复制的设置屏幕的一个示例。
- [0054] 图33是示出了用于考虑三数据中心远程复制来确定复制目的地的三数据中心目的地的选择程序的处理流程的流程图示例。

具体实施方式

[0055] 在本发明的以下详细描述中参照附图,这些附图构成了本公开内容的一部分,并且在附图中通过举例说明而非限制性的方式示出了可以通过其实现本发明的示例性的实施方式。在附图中,相似的数字描述贯穿多个视图的基本上相似的组件。此外,应当注意的是,虽然详细描述提供了如下面所描述的并且在附图中示出的各个示例性的实施方式,但是本发明不限于本文所描述和示出的实施方式,而是可以扩展到本领域技术人员将知道或者即将知道的其它实施方式。在说明书中提及“一个实施方式”、“该实施方式”或者“这些实施方式”意味着结合实施方式所描述的特定的特征、结构或特性包括在本发明的至少一个实施方式中,并且这些短语在说明书中的各个位置的出现不一定均是指代相同的实施方式。此外,在下面的详细描述中,给出了大量具体细节以提供对本发明的彻底理解。然而,对于本领域普通技术人员而言显而易见的是,为了实现本发明,不是所有这些具体细节都是必需的。在其它环境中,公知的结构、材料、电路、过程和接口已经被详细描述,和/或可以以框图的形式被示出,以不必模糊本发明。

[0056] 此外,根据计算机中的操作的算法和符号表示给出了下面的详细描述的一些部分。这些算法描述和符号表示是数据处理领域中的技术人员用于最有效地向本领域其它技术人员传达其创新的实质的方式。算法是导致期望的最终状态或结果的一系列定义的步骤。在本发明中,所执行的步骤需要物理操纵有形的量以实现有形的结果。通常,虽然不是必需的,但这些量具有能够被存储、传送、组合、比较和以其它方式操纵的电信号或磁信号或者指令的形式。已经证明主要由于共同使用的原因,有时将这些信号称作比特、值、要素、符号、字符、术语、数字、指令等是便利的。然而,应当记住的是,所有这些和类似的术语将与适当的物理量相关联,并且仅仅是应用于这些量的便利的标签。除非另外专门声明,通过以下讨论显而易见的是,应当理解,贯穿说明书使用诸如“处理”、“计算”、“运算”、“确定”、“显示”等术语的讨论可以包括操纵计算机系统的寄存器和存储器中的表示为物理(电子)量的数据并将这些数据转换为类似地表示为计算机系统的存储器或寄存器或者其它信息存储、传输或显示设备中的其它数据的计算机系统或其它信息处理设备的动作和过程。

[0057] 本发明还涉及一种用于执行本文的操作的装置。该装置可以被专门构造用于所需的目的,或者其可以包括由一个或多个计算机程序选择性地激活或重新配置的一个或多个通用计算机。此类计算机程序可以存储在计算机可读存储介质中,所述计算机可读存储介质包括非瞬态介质,例如但不限于:光盘、磁盘、只读存储器、随机存取存储器、固态设备和驱动器或者适合于存储电子信息的任何其它类型的介质。本文呈现的算法和显示器未固有地涉及任何特定的计算机或其它装置。各个通用系统可以与根据本文的教导的程序和模块一起使用,或者可以证明构造更专用的装置来执行期望的方法步骤是便利的。此外,本发明不是参照任何特定的编程语言来描述的。将清楚的是,各种编程语言可以用于执行本文所描述的本发明的教导。编程语言的指令可以由诸如中央处理单元(CPU)、处理器或控制器的一个或多个处理设备来执行。

[0058] 下面将更详细描述的本发明的示例性实施方式提供了用于管理包括远程复制系统的存储系统以及通过对复杂的操作进行自动化来改进可管理性的装置、方法和计算机程序。

[0059] 在下面的描述中,将描述处理并且在一些情况下将“程序”作为主题来操作。在程序由处理器执行的情况下,执行预定的处理。因此,处理的主题也可以是处理器。当程序作

为主题被操作时所公开的处理还可以是由执行程序的处理程序或者具有处理器(例如,控制设备、控制器和存储系统)的装置执行的。此外,当处理器执行程序时被执行的处理的一部分或全部也可以由作为处理器的替代物的硬件电路或者由除了处理器以外的硬件电路来执行。

[0060] 第一实施方式

[0061] 图1示出了在其中可以应用本发明的方法和装置的系统的硬件配置的示例。该系统包括服务器100、管理服务器300和形成虚拟存储系统200的一个或多个物理存储系统210(PDKC)。

[0062] 服务器具有CPU、存储器、I/O(输入/输出)I/F等。服务器通过执行诸如数据库管理系统的操作系统和软件来提供服务。数据库管理系统处理的数据被存储在存储系统200中。服务器100经由网络110被耦合到存储系统200。

[0063] 被配置为管理存储系统200的管理服务器300经由网络310耦合到存储系统200。管理服务器300被提供有CPU、存储器(M)301、输入/输出部分和管理I/F。存储器(M)301存储用于管理的程序。CPU执行管理程序以执行管理处理。输入/输出部分例如是由鼠标、键盘、显示器等配置的。输入/输出部分从执行管理的操作者接收各种指令输入,并且在显示器上显示各种信息。管理端口作为与存储系统200通信的中介。管理服务器300可以连接到服务器100并且还可以管理服务器100。

[0064] 在物理存储系统210中,I/O I/F211经由网络110耦合到服务器100,并且作为与服务器100通信的中介。管理I/F212经由网络310耦合到管理服务器300,并且作为与管理服务器300通信的中介。处理器213通过执行已经存储在存储器(S)214的程序单元2143中的各种程序来执行各种处理。此外,处理器213通过使用已经存储在存储器(M)214的控制信息单元2141中的各种信息来执行各种处理。例如,存储器单元(S)214是由例如至少一个存储器设备配置的,并且具有被配置为存储控制信息的控制信息单元2141、被配置为存储程序的程序单元2143和作为被配置为缓存数据的缓存存储器的示例的高速缓存单元2145。通常,与卷217的容量相比,高速缓存部分2145的容量更小。磁盘I/F215经由总线耦合到作为物理存储设备的示例的至少一个HDD216。例如,被配置为管理数据的卷217是由HDD216的至少一个存储区域配置的。物理存储设备不限于HDD并且还可以是SSD(固态驱动器)或DVD。可以在奇偶校验组的单元中收集至少一个HDD216,并且还可以使用诸如RAID(独立磁盘冗余阵列)的高可靠性技术。物理存储系统210可以具有上述资源中的一个或多个。这些资源经由内部总线218相互连接。

[0065] 在虚拟存储系统200中,物理存储系统210经由网络219相互连接。可以为了冗余来复用网络219。将相互连接的两个或更多个物理存储系统210作为一个存储系统(虚拟存储系统)200提供给服务器100。为了实现对物理存储系统210的虚拟化,需要唯一的卷ID分配方法和两个或更多个物理存储系统210之间的资源共享技术。这种技术是已知的。

[0066] 图2是示出了根据第一实施方式的管理服务器300中的存储器(M)301的示例的详细框图。控制信息单元302包含服务器站点管理表303和存储站点管理表304。这些表存储稍后解释的由处理来使用的信息。在图3和图4中示出了这些表的细节。使用这些表来检测相同的数据中心中的服务器和存储。程序单元305包含用于向存储系统200指导卷创建的卷创建程序(管理)306。

[0067] 图3是在管理服务器300中管理的服务器站点管理表303的示例。服务器ID是用于在计算机系统中唯一地标识服务器的标识。站点ID是用于在计算机系统中唯一地标识站点(数据中心)的标识。

[0068] 图4是在管理服务器300中管理的存储站点管理表304的示例。PDKC ID是用于在计算机系统中唯一地标识物理存储设备的标识。站点ID是用于在计算机系统中唯一地标识站点(数据中心)的标识。

[0069] 图5是示出了根据第一实施方式的物理存储系统210中的存储器(S)214的示例的详细框图。控制信息单元2141包含全局ID管理表2142。该表存储稍后解释的由处理使用的信息。图6中示出了该表的细节。程序单元2143包含用于创建新卷的卷创建程序(存储设备)2144。提供高速缓存单元2145以用于高速缓存。

[0070] 图6是在物理存储系统210中管理的全局ID管理表2142的示例。全局卷ID是用于在虚拟化的存储系统200中唯一地标识卷的卷标识。PDKC ID是用于在计算机系统中唯一地标识物理存储系统210的标识。本地卷ID是用于在物理存储系统210中唯一地标识卷的卷标识。卷大小是由全局卷ID指定的卷的大小。

[0071] 图7是示出了卷创建处理流程的流程图的示例。该处理是由卷创建程序(管理)306和卷创建程序(存储设备)2144实现的。在步骤S100中,卷创建程序(管理)306经由管理服务器300的输入/输出部分从操作者接收卷创建请求。此时,卷创建程序将服务器ID、全局卷ID和卷大小作为参数进行接收。在步骤S101中,卷创建程序(管理)306通过使用服务器站点管理表303和存储站点管理表304来检测相同站点中的PDKC。在步骤S102中,卷创建程序(管理)306寻找具有指定大小的空闲空间的一个或多个存储系统。在步骤S103中,卷创建程序(管理)306向PDKC指导卷创建。如果在步骤S102中找到两个或更多个PDKC,则可以选择具有最大空闲空间的PDKC并且可以选择具有最低负载的PDKC。可替换地,该程序可以认为一旦找到了满足条件的第一PDKC则步骤S102就完成,并且前进至步骤S103。

[0072] 接收指导的卷创建程序(存储设备)2144创建卷,并且将该卷映射到连接到由服务器ID指定的服务器的I/O I/F(步骤S200)。该创建的卷可以是虚拟卷。当从服务器100写入数据时,虚拟卷技术分配物理存储区域。在很多存储产品中实现该技术。在卷创建之后,卷创建程序(存储设备)2144向调用方发送完成消息(步骤S201)。在步骤S104,卷创建程序(管理)306从PDKC接收完成消息并且终止处理。

[0073] 当服务器仅通过前面提到的处理来指定全局卷ID时,确定最佳的物理存储区域。此外,当虚拟化的存储系统在不同的数据中心中包括PDKC时,不选择引起I/O延迟下降的PDKC。因此,降低了用于进行配置的设计成本和操作成本。

[0074] 第二实施方式

[0075] 第二实施方式涉及一种从构成虚拟化的存储系统的物理存储系统中自动地选择用于远程复制的资源的方法。该方法还执行用于远程复制的自动配置(例如,远程复制路径配置)并且创建传送数据存储区域。通过使用图8和图9描述了远程复制功能。将远程复制分为两类:同步远程复制和异步远程复制。

[0076] 图8示出了用于同步远程复制处理的远程复制配置的示例。主存储系统600是远程复制的源存储系统,并且包括存储由服务器使用的数据的主卷(PVOL)601。辅存储系统700是远程复制的目的地存储系统,并且包括存储PVOL601的复制数据的辅卷(SVOL)701。

PVOL601与SVOL701之间的关系被称作远程复制对。主存储系统600和辅存储系统700分别具有用于数据传送的一个或多个I/O I/F602、702。对于同步远程复制处理而言,主存储系统600从服务器100接收写入请求。然后,主存储系统600向PVOL601写入 写入数据,并且将写入数据传送到辅存储系统700。主存储系统600不向服务器发送完成消息,直到接收到来自辅存储系统700的完成消息为止。接收到复制请求的辅存储系统700将经传送的数据写入SVOL701,并且向主存储系统600发送完成消息。最后,主存储系统600向服务器100发送完成消息。在这种同步远程复制中,在PVOL601与SVOL701之间实现了数据的高一致性。因此,即使主存储系统600出现故障,也可以使由此产生的数据丢失最小化。然而,随着主存储系统与辅存储系统之间的距离增加,存在着延长写入响应时间的缺点。

[0077] 图9示出了用于异步远程复制处理的远程复制配置的示例。主存储系统600从服务器100接收写入请求。然后,主存储系统600在PVOL601和JVOL608中写入数据作为日志(JNL)。JNL是传送数据,并且JVOL608是用于暂时地存储JNL的存储区域。JVOL不必是卷。JNL可以存储在高速缓存单元2145中。此后,主存储系统600向服务器发送完成消息。在辅存储系统700中向JVOL703异步传送JNL。最后,向SVOL701复制JVOL703中存储的JNL。因此,与同步远程复制相比,写入响应时间更短。然而,还存在以下缺点:与同步远程复制相比,如果主存储系统600出现故障,则存在更容易数据丢失和更大量的数据丢失的缺点。在现有的远程复制系统中,操作必须配置辅存储系统、SVOL、JVOL、用于远程复制的I/O I/F、服务器(辅)、远程复制对等。这些操作非常复杂。

[0078] 图10是示出了根据第二实施方式的物理存储系统210中的存储器(S)214的示例的详细框图。与图5中所示的第一实施方式相比,添加了主存储对表2146、辅存储对表2147、p-对创建程序2149、s-对创建程序2150、写入程序(存储设备)2151和流入表2148。主存储对表2146和辅存储对表2147是用于管理PVOL601与SVOL701之间的关系的表。在图12和图13中示出了这些表的细节。p-对创建程序2149和s-对创建程序2150是用于在主存储系统和辅存储系统中创建远程复制对的程序。在图17中示出了这些程序的处理流程。写入程序(存储设备)2151是用于向目标卷存储写入数据的程序。在该实施方式中,在图22和图24中描述了在PVOL故障以后接收SVOL701写入的两种情况。流入表2148是用于管理写入到PVOL601中的写入数据的量的表。提供该信息以考虑在主存储系统与辅存储系统之间所需的带宽。

[0079] 图11是示出了根据第二实施方式的管理服务器300中的存储器(M)301的详细框图。与图2中所示的第一实施方式相比,添加了站点距离表307、对创建程序308、辅选择程序309、JVOL容量程序311和迁移站点程序312。站点距离表307是用于管理数据中心之间的距离的表。在图15中示出了表的细节。对创建程序308和辅选择程序309是用于选择复制目的地并且向存储系统指导远程复制对的创建的程序。在图17至图19中示出了这些程序的处理流程。JVOL容量程序311是用于估计所需的JVOL容量的程序。在图20中示出了程序的处理流程。迁移站点程序312是用于确定迁移的目的地以进行服务器处理的程序。在图24和图25中示出了程序的概念和处理。

[0080] 图12是物理存储系统210中的存储器(S)214中的主存储对表2146的示例。图13是物理存储系统210的存储器(S)214中的辅存储对表2147的示例。当物理存储系统210被配置为远端复制的主存储系统时,使用主存储对表2146。当物理存储系统210被配置为远端复制的辅存储系统时,使用辅存储对表2147。主卷ID是作为复制源卷的卷的ID。辅PDKC ID是具

有SVOL701的辅物理PDKC700的ID。辅卷ID是作为PVOL601的复制目的地卷的本地卷ID。主PDKC ID是具有PVOL601的主物理PDKC600的ID。

[0081] 图14是在物理存储系统210中管理的流入表2148的示例。该表管理去往每一个卷的写入数据的流入。本地卷ID存储每一个卷的ID,并且流入存储针对每个卷的每秒写入数据量。

[0082] 图15是在管理服务器300中管理的站点距离表307的示例。站点ID的意义与图3中的站点ID的意义相同。距离存储由站点ID指定的数据中心之间的距离。

[0083] 通过使用图16至图21来描述自动远程复制配置处理。

[0084] 图16是用于数据复制的设置屏幕800的示例。HA(高可靠性)意味着本地复制。RC(同步)意味着同步类型的远程复制。RC(异步)意味着异步类型的远程复制。设置屏幕800具有复选框。通过复选该框来应用诸如远程复制的功能。在所示的示例中,向具有卷ID1的卷应用异步远程复制。

[0085] 图17是示出了对创建处理流程的流程图示例。在步骤S300中,对创建程序308调用辅选择程序309以选择辅PDKC。在图18中示出了辅选择程序309的处理流程。在步骤S301中,已经从辅选择程序309接收到辅信息的对创建程序308向所选择的存储系统指导卷创建。所选择的存储系统中的卷创建程序(存储设备)2144的处理与图7的步骤S200和S201相同。在步骤S302中,对创建程序308向主PDKC指导对创建,并且等待完成消息。此时,将操作者指定的全局卷ID和在步骤S301中选择的本地卷ID以及在步骤S300中选择的辅PDKC700的ID作为参数进行发送。

[0086] 接收对创建请求的主PDKC600执行p-对创建程序2149。p-对创建程序2149确认是否可以创建对(步骤S400)。例如,其包括确定主存储对表2146是否具有空闲条目,主PDKC600是否连接到辅PDKC700等等。在确认以后,p-对创建程序2149向在步骤S300中选择的辅PDKC指导对创建。将全局卷ID、PVOL和SVOL的本地卷ID和主PDKC的ID作为参数进行发送。

[0087] 接收对创建请求的辅PDKC700执行s-对创建程序2150。s-对创建程序2150确认该对是否可以被创建并且更新辅存储对表2147(步骤S500和S501)。在步骤S502中,s-对创建程序2150更新全局ID管理表2142并且向主PDKC600发送完成消息。所接收的全局卷ID被存储在全局卷字段中,并且所接收的本地卷ID被存储在本地卷ID字段中。其自己的ID被存储在PDKC ID字段中。由本地ID指定的卷的大小被存储在卷大小字段中。

[0088] 在步骤S401中,p-对创建程序2149在从辅PDKC700中接收到完成消息以后,更新主PDKC600中的主存储对表2146。在步骤S402中,p-对创建程序2149向调用方(对创建程序308)发送完成消息。在步骤S303中,对创建程序308终止该处理。

[0089] 图18是示出了从图17中的对创建程序308调用的辅选择程序309的处理流程的流程图示例。该程序用于选择最佳的复制目的地存储设备。在步骤S600中,辅选择程序309确认该请求是否是远程复制请求。如果该请求不是远程复制请求,则程序前进至步骤S601并且选择其自己的PDKC作为复制目的地PDKC(步骤S601),并且然后在步骤S607中向调用方报告PDKC ID。这是因为该请求针对本地复制。如果请求是远程复制请求,则辅选择程序309标识直接连接到具有PVOL的主PDKC的存储系统(步骤S602)。然后,辅选择程序309确认该请求是否是同步远程复制请求(步骤S603)。如果该请求是同步远程复制请求,则程序前进至

步骤S604并且寻找与指定的主PDKC相距约100km的存储系统(步骤S604)。100km的意义是可以大体上提供可允许的I/O响应时间以进行同步远程复制的距离。可替换地,可以应用可以提供可允许的I/O响应时间的一些其它距离。如果该请求不是同步远程复制请求,则辅选择程序309寻找与指定的主PDKC相距约1000km或者更远的存储系统(步骤S605)。1000km的意义是可以避免主数据中心灾难的后果的距离。可替代地,可以应用可以避免灾难的后果的某一其它距离。在步骤S604或者S605之后,在步骤S606中,辅选择程序309选择满足条件(例如,针对SVOL的空闲容量、连接性等等)的PDKC。最后,辅选择程序309向调用方报告决定的PDKC ID(步骤S607)。

[0090] 通过使用图19和图20来描述异步远程复制的情况。

[0091] 图19是示出了用于异步远程复制的对创建处理流程的流程图的一个示例。在步骤S700中,对创建程序308调用JVOL容量程序311以计算JVOL608、703所需的容量。对创建程序308调用辅选择程序309并且指导卷创建(步骤S701和S702)。步骤S701和S702与图17中的步骤S300和S301相同。卷创建程序(存储设备)2144的处理与图7的步骤S200和S201相同。然后,对创建程序308向主PDKC和辅PDKC指导JVOL创建(步骤S703)。将在步骤S701中计算出的JVOL大小作为参数进行发送。接收JVOL创建请求的每一个PDKC创建具有指定的大小的卷,并且该创建的卷可以是虚拟的卷(步骤S800)。PDKC中的每一个向调用方发送完成消息(步骤S801)。最后,在步骤S704中,对创建程序308向主PDKC指导对创建。步骤S704之后的后续步骤与图17中的步骤S302之后的后续步骤相同。

[0092] 图20是示出了用于计算JVOL608、703所需的容量的JVOL容量程序311的处理流程的示例。下面将描述JVOL608、703的容量。在例如主PDKC600与辅PDKC700之间发生路径故障的情况下,对于JVOL608而言,JNL开始维持。在此后解决路径故障的情况下,JVOL608中存储的JNL被发送到JVOL703。因此,在JVPL608具有足以存储已经生成的JNL的容量的情况下,可以在无需暂停和重新同步远程复制对的情况下继续远程复制。根据用户希望抵抗路径故障的所需时间来设计JVOL608的容量。因此,可以通过使用所需的时间和从服务器100写入的数据的量来计算JVOL容量。JVOL容量程序311从主PDKC600获得针对PVOL601的每秒写入数据量(步骤S900)。然后,当发生路径故障时,JVOL容量程序311获得所需的持续时间(步骤S901)。通过使用输入/输出部分向管理服务器300输入所需的持续时间。最后,JVOL容量程序311计算JVOL容量(步骤S902)。通过写入数据的量乘以持续时间来计算JVOL容量。

[0093] 图21是示出了辅选择程序309的处理流程的流程图的一个示例。该示例考虑主PDKC与辅PDKC之间的带宽。仅解释与图18不同的步骤。在步骤S602之后执行额外的步骤S1000和S1001,并且在步骤S605或者步骤S604之后执行步骤S1002而不是步骤S606。在步骤S1000中,辅选择程序309从主PDKC600获得针对PVOL的每秒写入数据量(MB/s)。在步骤S1001中,辅选择程序309计算主PDKC与辅PDKC之间所需的带宽。在辅PDKC选择步骤(S1002)中,辅选择程序309考虑计算出的带宽。

[0094] 图22是示出了在I/O故障之后的处理的概念图。服务器100使用其卷ID为0的卷。服务器100具有交替路径软件。在图22的示例中,交替路径软件具有去往卷0的两个路径。虚拟化的存储系统200被提供给服务器100。在存储系统层中,向物理存储系统600发起来自服务器100的I/O,并且向(物理存储系统700的)卷820复制在(物理存储系统600的)卷810中存储

的数据。当主PDKC出现故障时,交替路径软件改变I/O路径并且再次发起I/O。在存储系统层中,向物理存储系统700发起来自服务器100的I/O。在物理存储系统700的卷820中存储写入数据。因为向服务器100提供的卷ID是相同的,因此服务器100没有意识到物理存储系统的改变。通常,为了一致性,远程复制的SVOL拒绝服务器I/O。在PVOL故障以后,SVOL的状态必须变为可以接收和处理服务器I/O的状态。在图23中示出了该处理。

[0095] 图23是示出了在PVOL出现故障以后向SVOL发起的写入处理的流程图的示例。该写入程序(存储设备)2151接收去往SVOL的I/O(步骤S1100)并且确认PVOL故障是否已经发生(步骤S1101)。如果PVOL故障未发生,则写入程序(存储设备)2151向服务器100报告错误(步骤S1102)。如果PVOL故障已经发生,则写入程序(存储设备)2151将SVOL状态改变为I/O可接收的状态(步骤S1103)。然后,写入程序(存储设备)执行I/O并且终止处理(步骤S1104和S1105)。

[0096] 在图22中示出了服务器未经改变的情况。在本发明的配置的情况下,因为安装在不同的数据中心中的物理存储系统210可以构成虚拟化的存储系统200,因此I/O目的地物理存储系统可以改变为与其中安装了服务器的数据中心不同的数据中心。因此,需要超出数据中心的I/O并且性能下降。为了避免这种情况,还可以改变服务器。

[0097] 图24是示出了在PVOL出现故障以后的服务器变化和涉及SVOL的处理中的后续变化的概念图。仅解释与图22的区别。服务器400作为迁移的源被安装,并且服务器500作为迁移的目的地被安装。服务器400连接到服务器500。在物理服务器400和500中的每一个中创建虚拟机410。通过虚拟机来为应用提供服务。可以从一个物理服务器向另一个物理服务器迁移虚拟机。可以通过传统的技术来执行迁移。在卷810故障以后或者在PDKC600和I/O目的地改变为PDKC700中的卷820以后,虚拟机410从服务器400迁移到处于与PDKC700相同的数据中心中的服务器500。因此,可以避免超出数据中心的I/O。

[0098] 图25是在PVOL出现故障以后的虚拟机410的迁移处理以及用于确定目的地的处理的流程图的示例。在步骤S1200中,写入程序(服务器)检测I/O故障。与预定的值相比更长的响应时间也可以是检测的条件。写入程序(服务器)向管理服务器300查询迁移的目的地。接收查询的管理服务器300调用迁移站点程序312。迁移站点程序312向具有PVOL的PDKC查询具有SVOL的PDKC的信息(步骤S1300)。PDKC可以使用主存储对表2146来对其进行确认。然后,迁移站点程序312向服务器报告PDKC的所获得的信息(步骤S1301)。最后,写入程序(服务器)向所报告的PDKC移动虚拟机410。

[0099] 通常,远程复制功能具有由一个或多个远程复制对构成的一致性组609。图26是示出了远程复制对的一致性组609的概念图。由一个应用使用的一个或多个远程复制对应当属于相同的一致性组609。操作者可以针对一致性组609执行远程复制操作。因此,操作者不需要操作每一个远程复制对。此外,使用远程复制的异步类型,在恢复处理中向SVOL701确认去往属于相同的一致性组609的PVOL601的写入顺序。在远程复制对创建处理时,应当从相同的数据中心中选择属于相同的一致性组609的PVOL的复制目的地。否则,故障以后的I/O将超出数据中心。

[0100] 图27是示出了考虑一致性组609的辅选择程序309的处理流程的流程图的一个示例。仅解释与图21的区别。在步骤S1001之后执行额外的步骤S1400、S1401和S1402。首先,当使用复选框应用远程复制时,由操作者经由设置屏幕800输入一致性组ID。在步骤S1400

中,辅选择程序309从主PDKC获得指定的一致性组信息。在步骤S1401中,辅选择程序309确定指导的远程复制对是否是一致性组的第一对。如果是,则辅选择程序309执行步骤S603和S607并且确定辅PDKC。否则,如果指导的远程复制对不是一致性组的第一对,则辅选择程序309前进至步骤S1402,并且选择具有已经属于指定的一致性组的SVOL的辅PDKC(步骤S1402)。然后,辅选择程序309前进至步骤S607,并且向调用方报告PDKC ID。

[0101] 通过使用上面提到的技术,用户的操作将是指定全局卷ID以及应用远程复制功能。因此,准备SVOL、配置远程复制路径等变得不必要。

[0102] 图28是示出了远程复制配置的自动恢复方法的概念图。仅解释与图24的区别。通过使用图24所示的技术,在移动虚拟机410并且将I/O目的地改变到PDKC700中的卷820以后,计算机系统处于不应用远程复制的状态中。因此,可靠性较低。自动恢复方法基于在初始设置时输入的内容自动地确定复制目的地,并且重新启动远程复制处理。添加对于配置恢复所需的第三物理存储设备900,如图28所示。第三物理存储设备900中的卷830是PDKC700中的卷820的复制目的地,并且也具有相同的全局卷ID0。

[0103] 通过图29中所示的管理服务器300中的配置恢复程序来实现上面的处理。可以在图25中的迁移站点程序312的步骤S1301之后调用配置恢复程序。首先,配置恢复程序接收PVOL或主PDKC600故障的通知(步骤S1500)。该配置恢复程序获得针对目标卷820的初始设置的内容(步骤S1501)。在图28的示例中,目标卷是具有全局卷ID0的卷。例如,内容是“RC(同步)”、“RC(异步)”等。然后,配置恢复程序使用获得的内容调用对创建程序308(步骤S1502)并且终止处理(步骤S1503)。

[0104] 在服务器400发生故障而不是PDKC600发生故障的情况下,远程复制的复制目的地可以是PDKC600。在该情况下,卷820成为复制源卷,并且卷810成为复制目的地卷。在一般的远程复制功能中,只将不同的数据从卷820复制到卷810。

[0105] 第三实施方式

[0106] 第三实施方式涉及远程复制配置的自动恢复方法和将上面的技术应用于三数据中心远程复制配置的方法。远程复制功能支持同步类型的远程复制和异步类型的远程复制的组合。在图30和图31中示出了两个示例。

[0107] 图30是示出了三数据中心远程复制的级联配置的概念图。级联的远程复制将PVOL中存储的数据同步地复制到SVOL,并且将SVOL中存储的数据异步地复制到TVOL(第三卷)。

[0108] 图31是示出了三数据中心远程复制的多目标配置的概念图。多目标远程复制将PVOL中存储的数据同步地复制到SVOL,并且将PVOL中存储的数据异步地复制到TVOL。

[0109] 级联和多目标的远程复制被简单地称作三数据中心远程复制(3DC远程复制)。这些功能既实现了作为同步类型的优点的无数据损失,又实现了作为异步类型的优点的更短的响应时间。

[0110] 图32是用于三数据中心远程复制的设置屏幕800的另一个示例。添加了用于级联配置和多目标配置的复选框。

[0111] 图33是示出了用于考虑到三数据中心远程复制来确定复制目的地的3DC目的地选择程序的处理流程的流程图示例。3DC目的地选择程序是由于到图32的设置屏幕800的输入而从对创建程序308调用的。

[0112] 在步骤S1600中,3DC目的地选择程序识别直接连接到主PDKC的存储系统。该步骤

与图18中的步骤S602是相同的。然后,3DC目的地选择程序确定同步类型的远程复制的复制目的地。3DC目的地选择程序查找与指定的主PDKC仅相距约100km的存储系统(步骤S1601),并且选择满足作为同步远程复制的条件的辅PDKC(步骤S1602)。3DC目的地选择程序确定指定的类型是否是级联的(步骤S1603)。如果指定的类型不是级联的,则3DC目的地选择程序从在步骤S1600中获得的信息中查找与主PDKC和辅PDKC相距约1000km或者更远的存储系统,并且然后前进至步骤S1607。如果指定的类型是级联的,则3DC目的地选择程序识别直接连接到所选择的辅PDKC的存储系统(步骤S1605),并且从前一步骤S1605中获得的信息中查找与主PDKC和辅PDKC相距约1000km或者更远的存储系统(步骤S1606)。在(步骤S1604或步骤S1606之后的)步骤S1607中,3DC目的地选择程序选择满足诸如异步远程复制的条件的PDKC。最后,3DC目的地选择程序向调用方报告辅PDKC ID和第三PDKC ID(步骤S1608)。

[0113] 在级联的情况下,不能检测到最佳的PDKC。因此,3DC目的地选择程序返回步骤S1602,并且在选择下一个最佳的辅PDKC之后再次执行步骤S1605至步骤S1607。

[0114] 图33中的步骤可以作为辅选择程序309的一部分被执行(图21或图27)。例如,这些步骤可以在图27中的步骤S1001之后被执行。在该情况下,针对三数据中心配置考虑PDKC之间的带宽等。

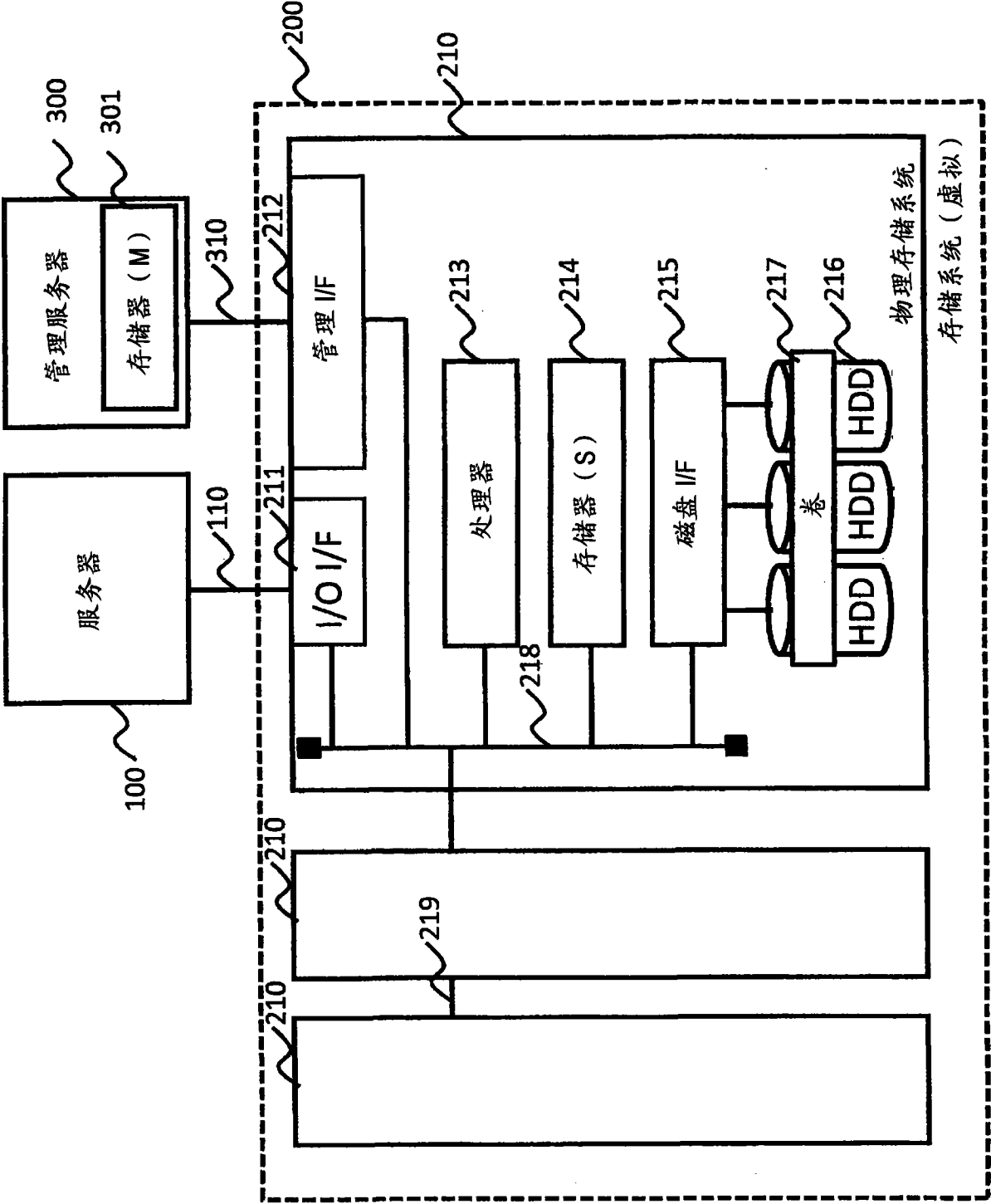
[0115] 当然,图1、8、9、24、26、28、30和31中所示的系统配置仅仅是可以在其中实现本发明的信息系统的示例,并且本发明并不限于特定的硬件配置。实现本发明的计算机和存储系统还可以具有已知的I/O设备(例如,CD和DVD驱动器、软盘驱动器、硬盘驱动器等),该已知的I/O设备可以存储和读取用于执行上面所描述的发明的模块、程序和数据结构。可以在这些计算机可读介质上编码这些模块、程序和数据结构。例如,本发明的数据结构可以独立于其上存储有本发明中使用的程序的一个或多个计算机可读介质而存储在计算机可读介质上。可以通过数字数据通信的任何形式或介质(例如,通信网络)来对系统的组件进行互连。通信网络的示例包括局域网、诸如因特网的广域网、无线网络、存储区域网络等。

[0116] 在说明书中,给出了大量细节以用于解释的目的,从而提供对本发明的彻底理解。然而,对于本领域技术人员而言显而易见的是,为了实现本发明,不是所有这些具体细节都是必需的。还应当注意的是,可以将本发明描述为通常描绘为流程图、流程示意图、结构示意图或框图的过程。虽然流程图可以将操作描述为连续过程,但是可以并行地或并发地执行这些操作中的很多操作。此外,可以重新排列操作的顺序。

[0117] 如本领域中公知的,上文所描述的操作可以由硬件、软件或者软件和硬件的某一组合来执行。本发明的实施方式的各个方面可以使用电路和逻辑设备(硬件)来执行,而其它方面可以使用存储在机器可读介质(软件)上的指令来执行,所述指令如果由处理器执行,则将使得处理器执行用于实现本发明的实施方式的方法。此外,本发明的一些实施方式可以单独地用硬件来执行,而其它实施方式可以单独地用软件来执行。此外,所描述的各个功能可以在单个单元中执行,或者可以以任意数量的方式分布在大量组件中。当由软件执行时,这些方法可以由诸如通用计算机的处理器基于存储在计算机可读介质上的指令来执行。如果期望的话,指令可以以压缩和/或加密的格式存储在介质上。

[0118] 根据上述内容将清楚的是,本发明提供了用于管理包括远程复制系统的存储系统并且通过使复杂的操作自动化来改进可管理性的方法、装置和存储在计算机可读介质上的

程序。此外,虽然已经在本说明书中示出和描述了具体的实施方式,但是本领域普通技术人员将清楚的是,可以用被计算为实现相同的目的的任何安排来替代所公开的具体实施方式。本公开内容旨在涵盖本发明的任何或所有调节或变化,并且应当理解的是,下面的权利要求中使用的术语不应当理解为将本发明限制于说明书中公开的具体实施方式。更确切地说,本发明的范围应当完全由下面的权利要求进行确定,其应当根据所建立的权利要求解释的教导以及这些权利要求有权享有的等同形式的整个范围来进行理解。



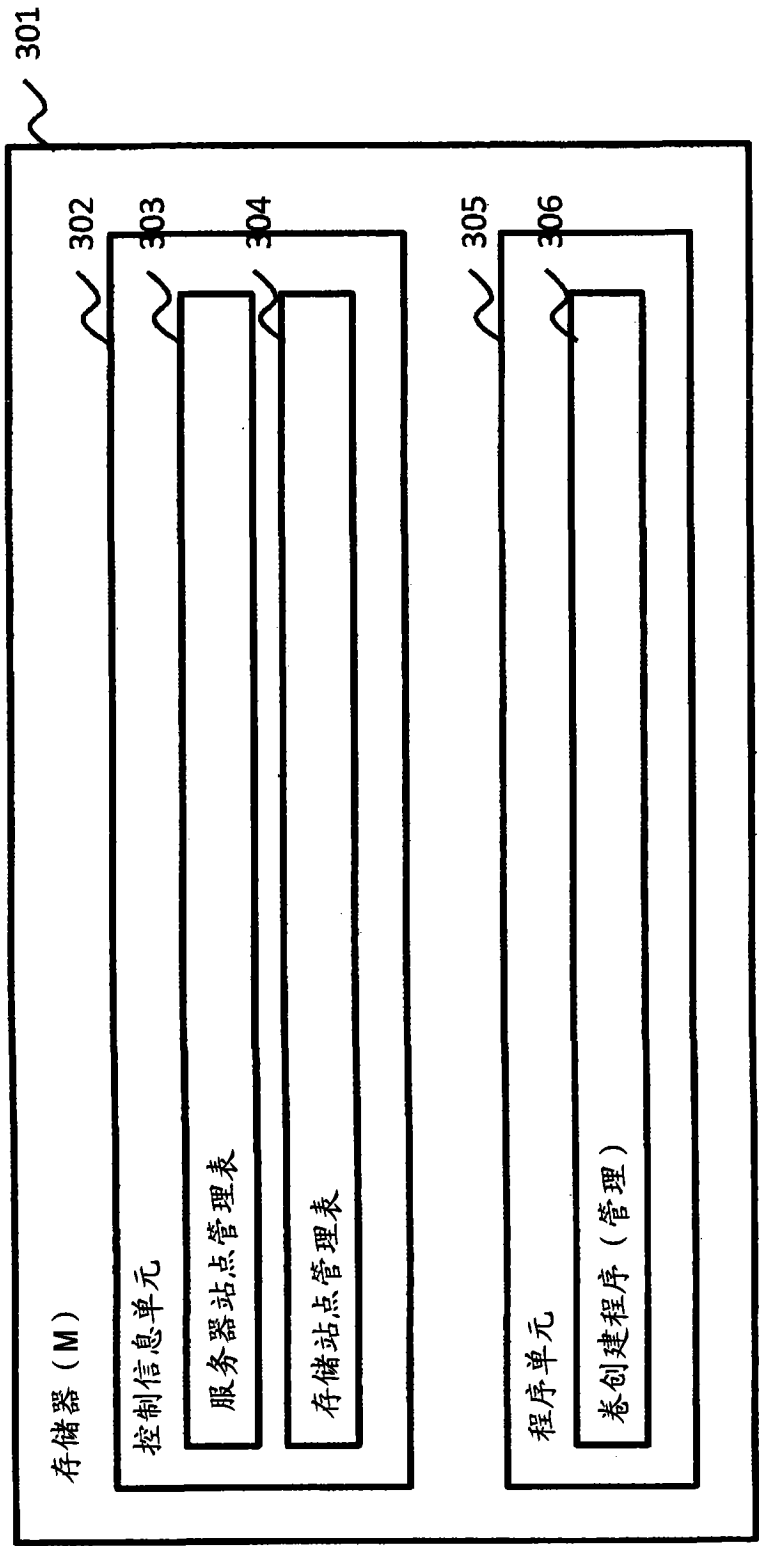


图2

303

服务器 ID	站点 ID
0	0
1	0
2	0
3	0
...	...

图3

304

PDKC ID	站点 ID
0	0
1	0
2	1
3	1
...	...

图4

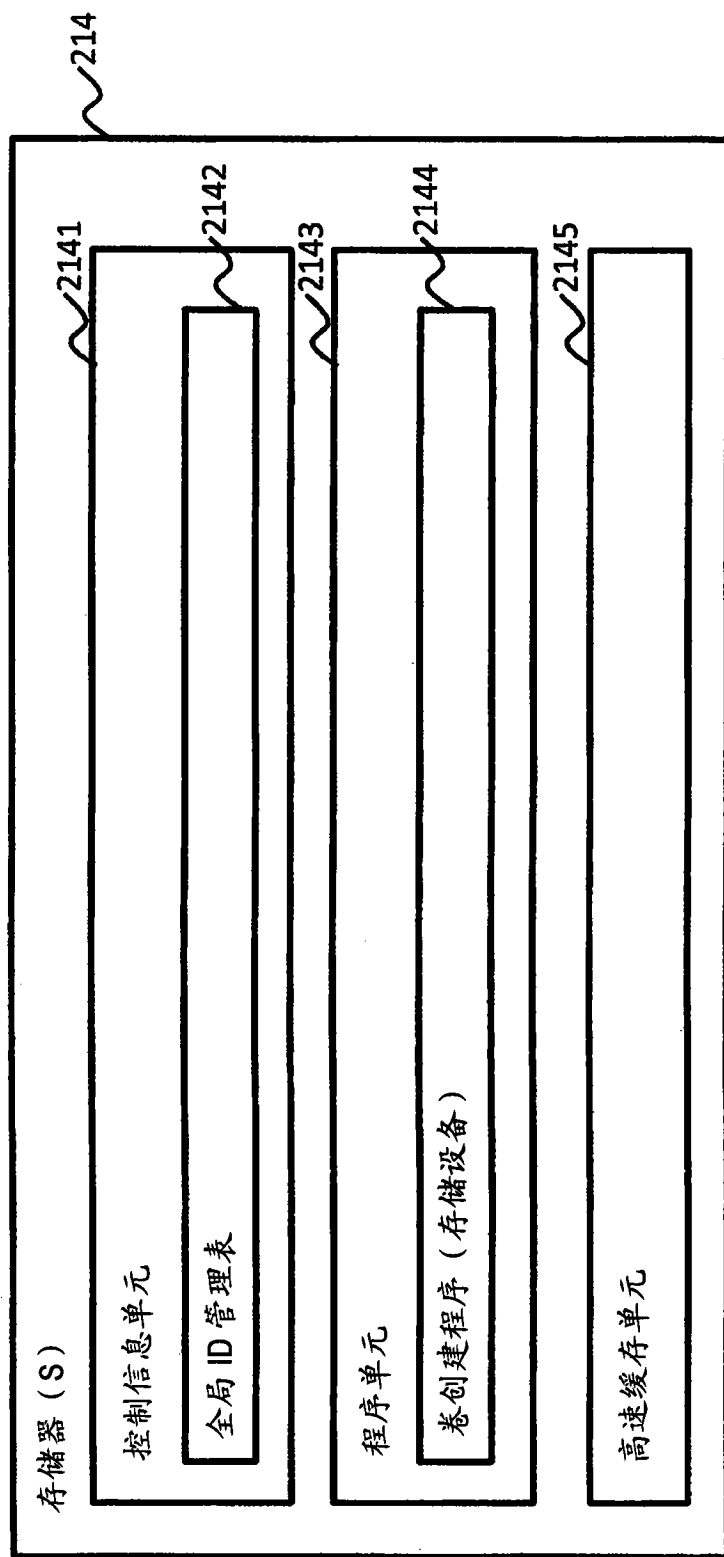


图5

2142

全局卷 ID	PDKC ID	本地卷 ID	卷大小
0	0	0	500GB
1	0	1	800GB
2	2	0	300GB
3	3	5	500GB
...	...	...	...

图6

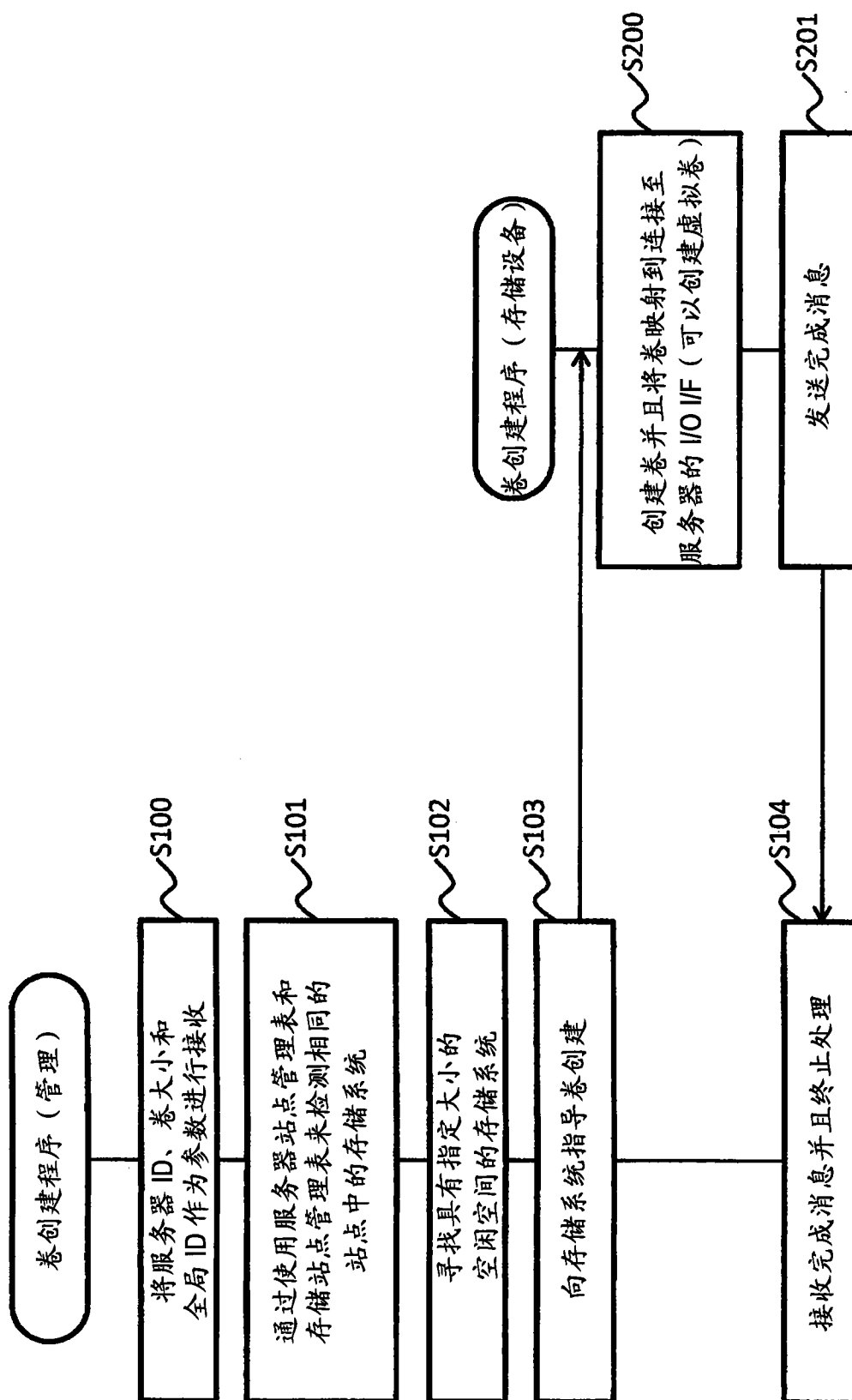


图7

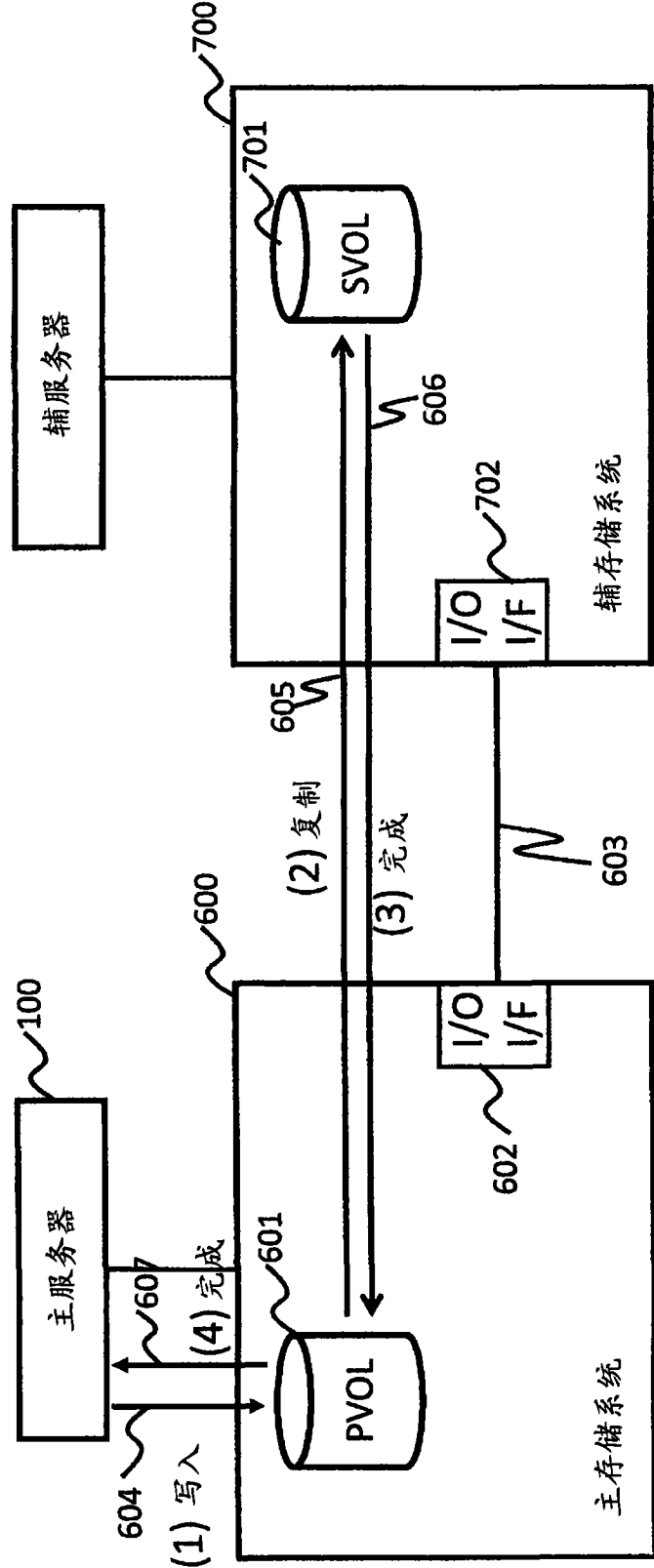


图8

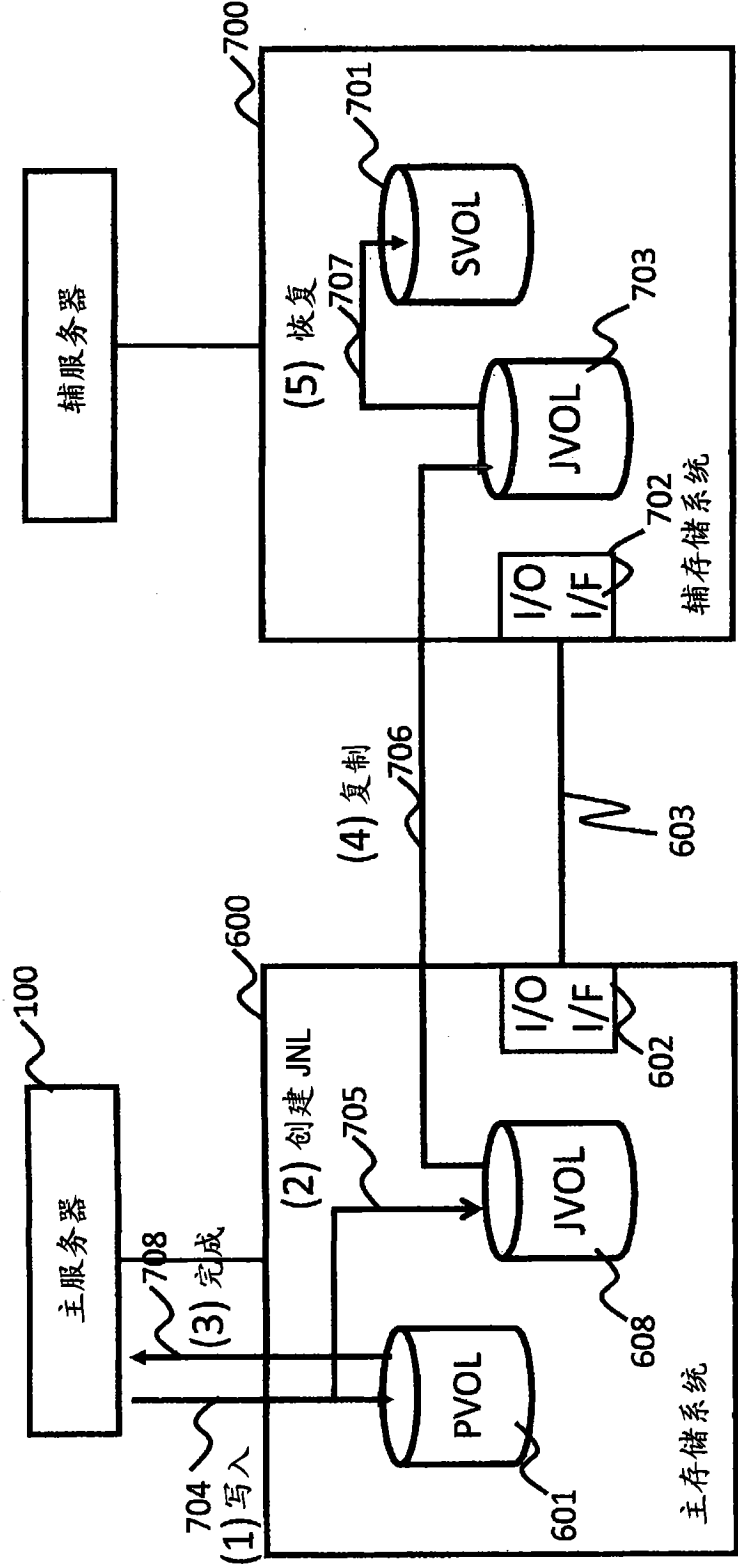


图9

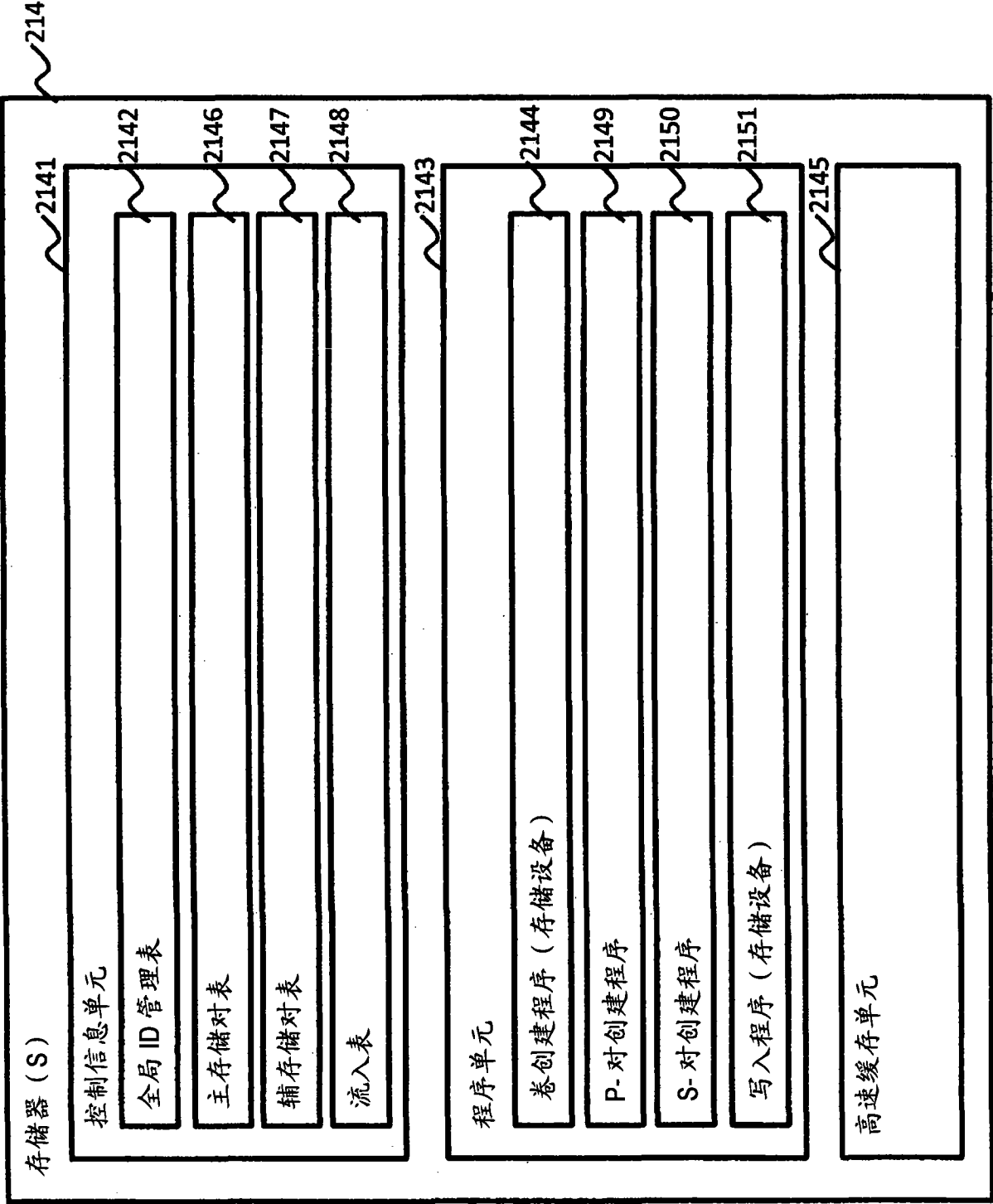


图10

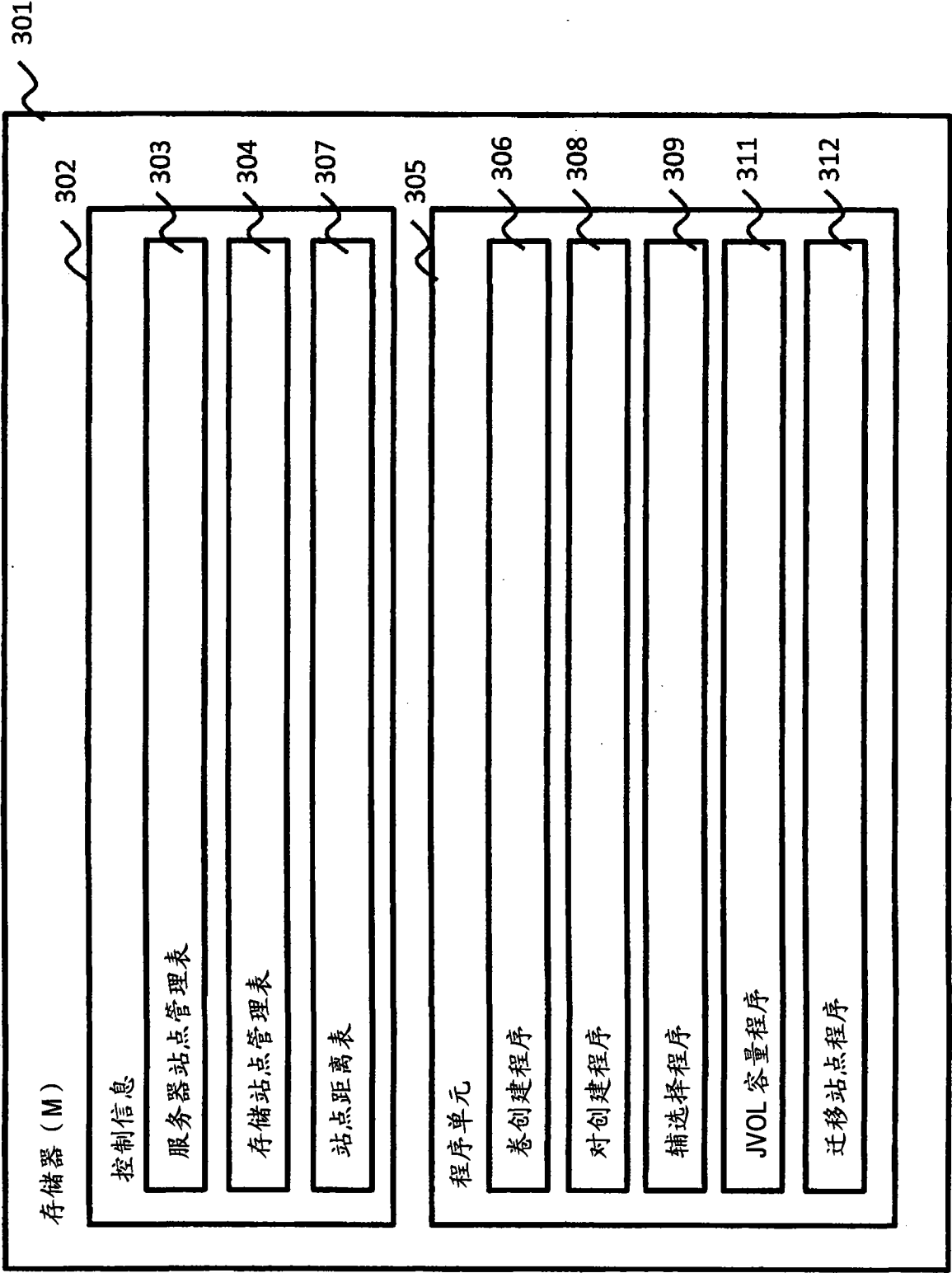


图11

2146

主卷 ID	辅 PDKC ID	辅卷 ID
0	2	0
1	2	10
2	3	3
3	3	6
...	...	...

图12

2147

辅卷 ID	主 PDKC ID	主卷 ID
0	0	0
1	0	9
2	0	12
3	0	5
...	...	...

图13

2148

本地卷 ID	流入
0	5MB/s
1	3MB/s
2	10MB/s
...	...

图14

307

站点 ID	站点 ID	距离
0	1	100km
0	2	1000km
0	3	300km
1	2	900km
1	3	200km
...	...	...

图15

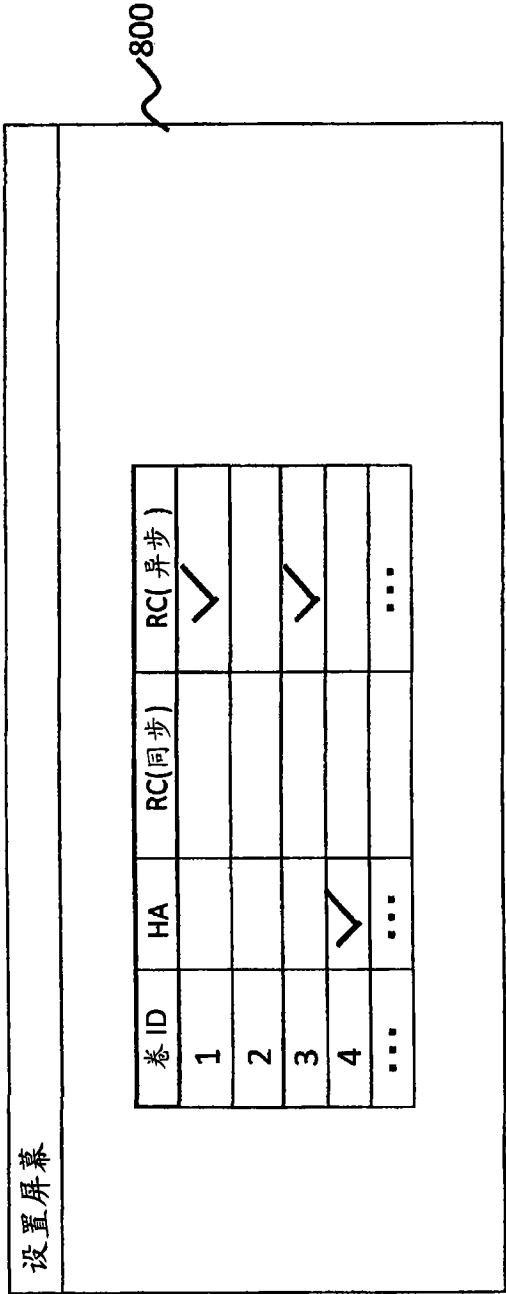


图16

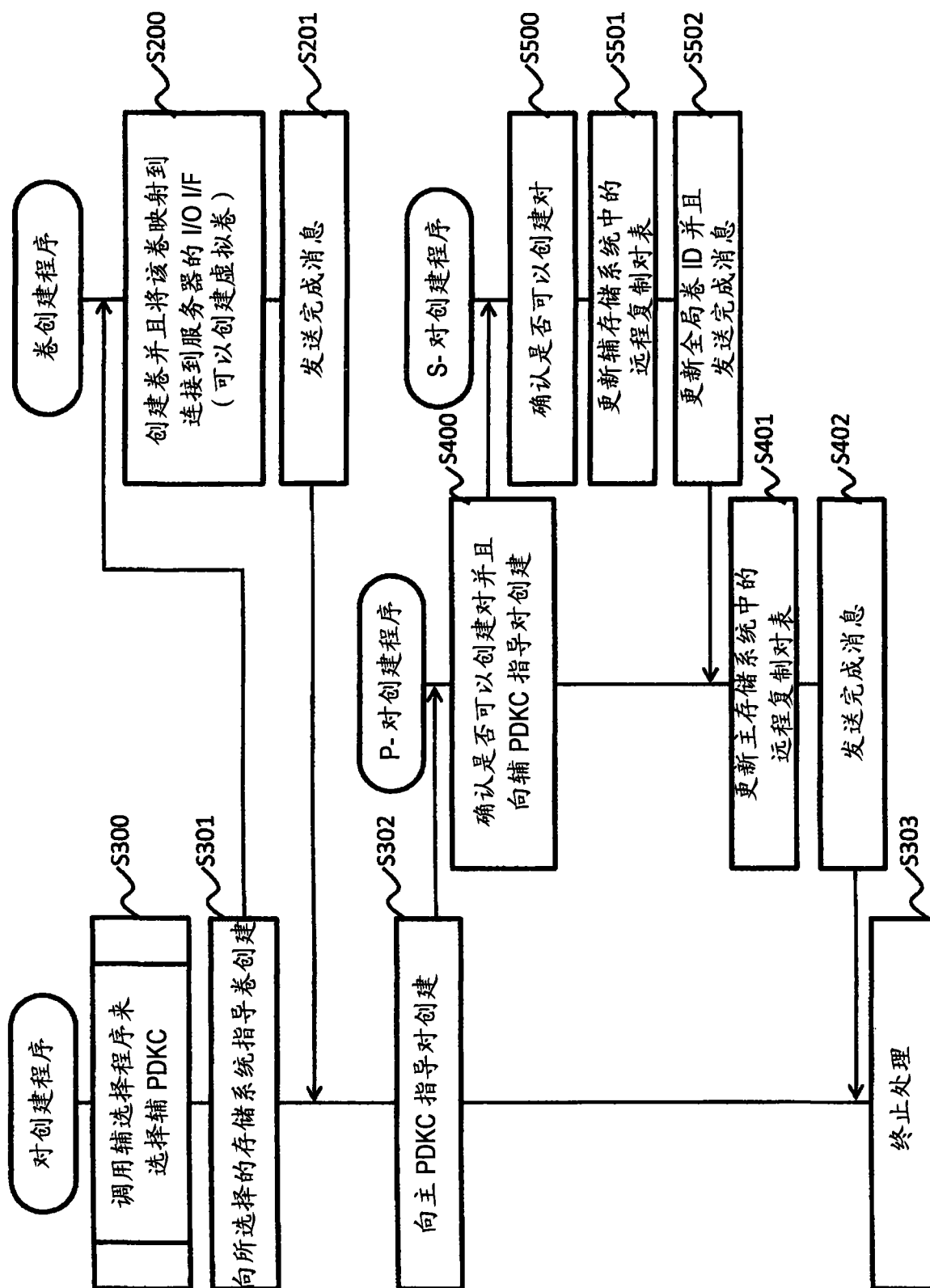


图17

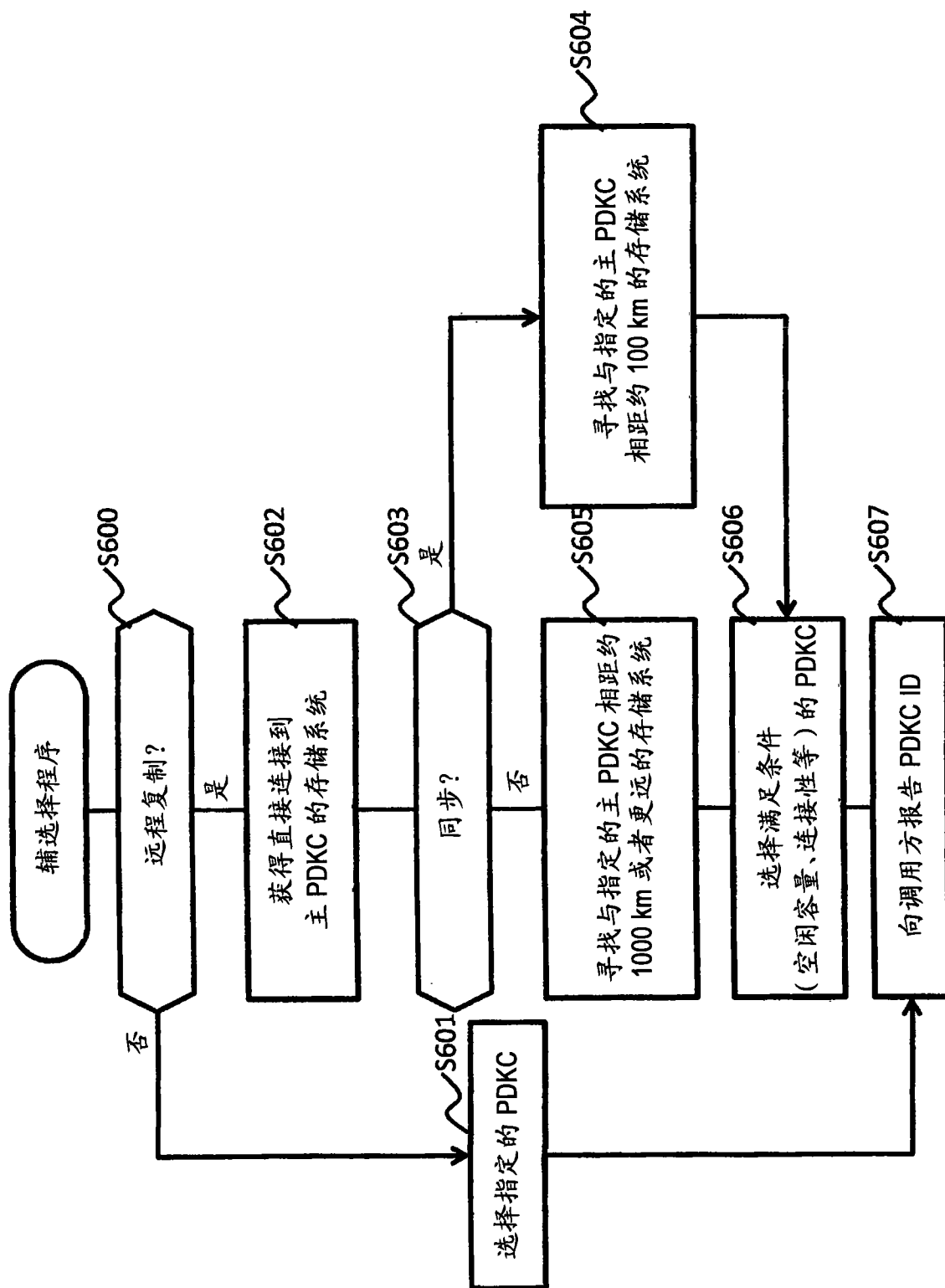


图18

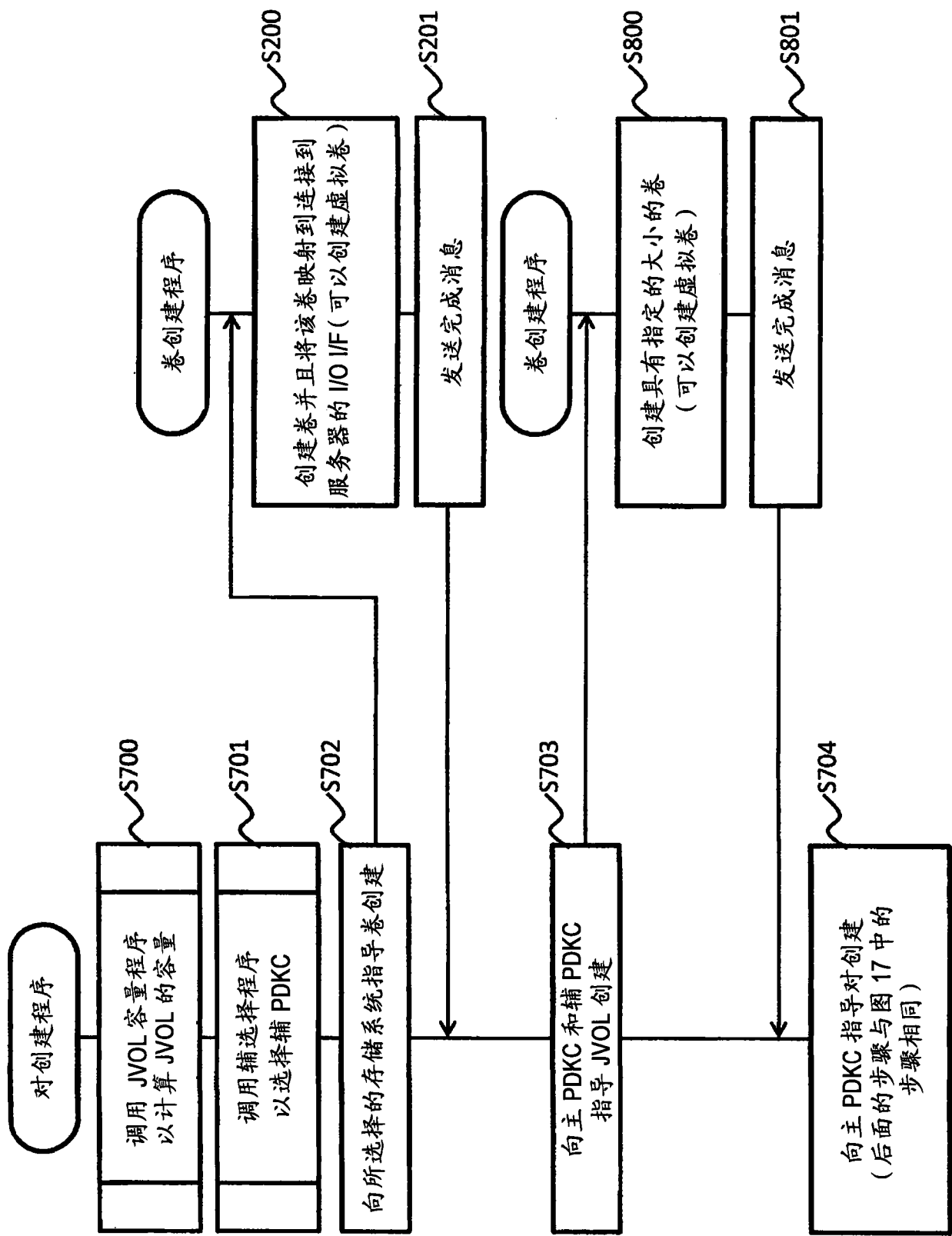


图19

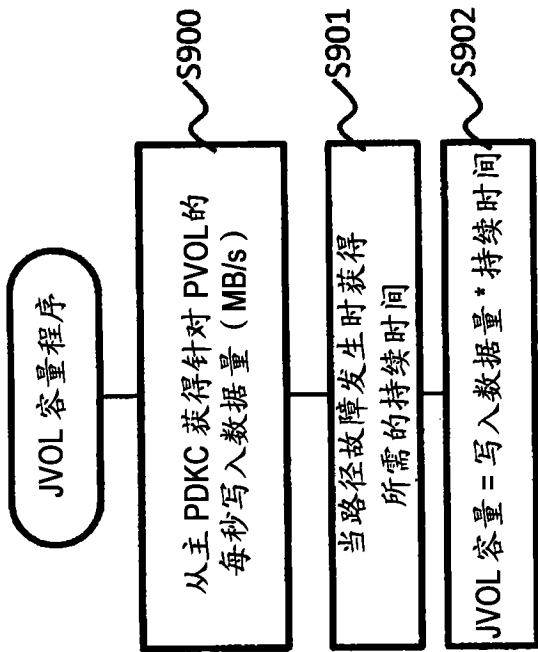


图20

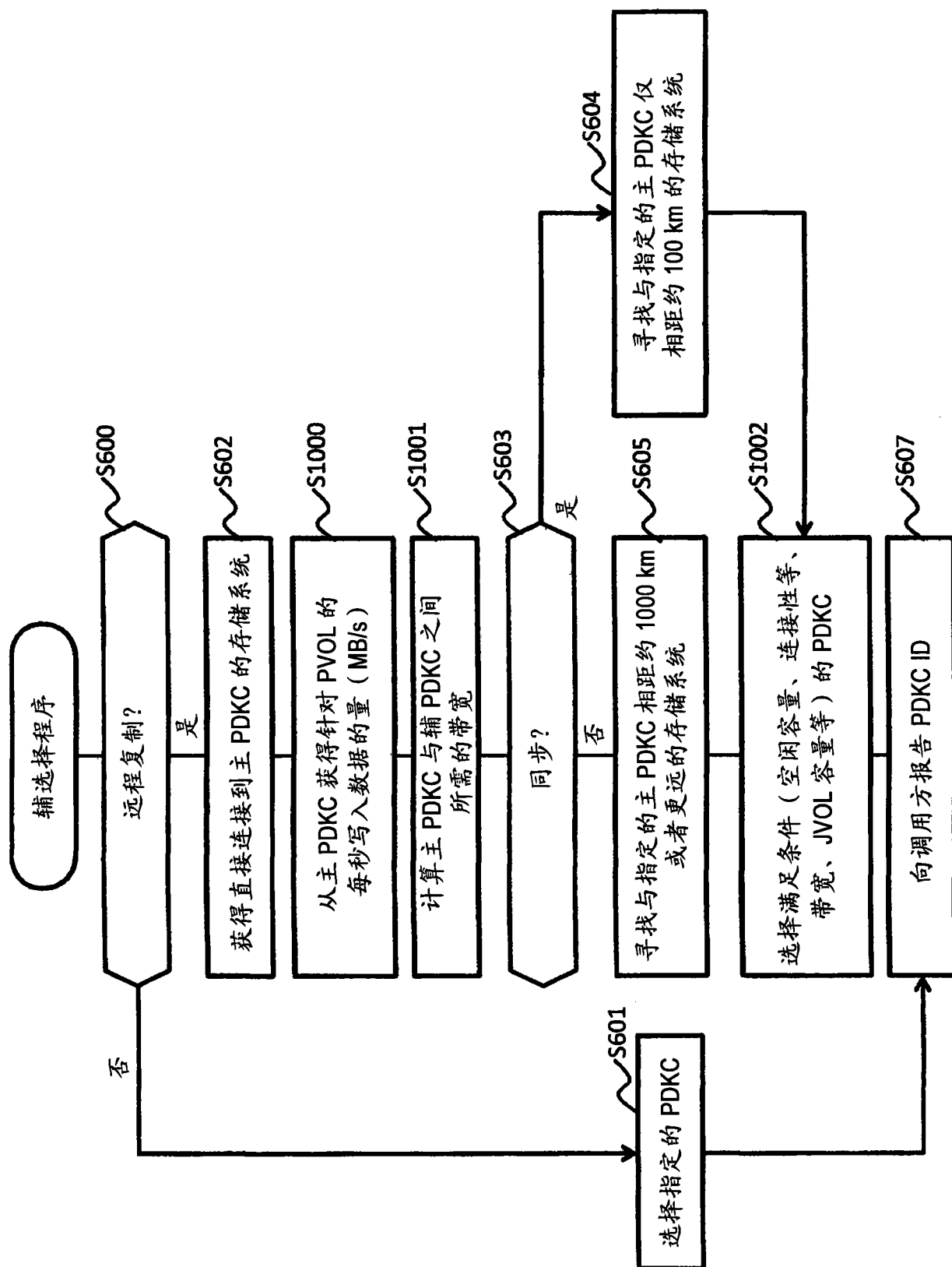


图21

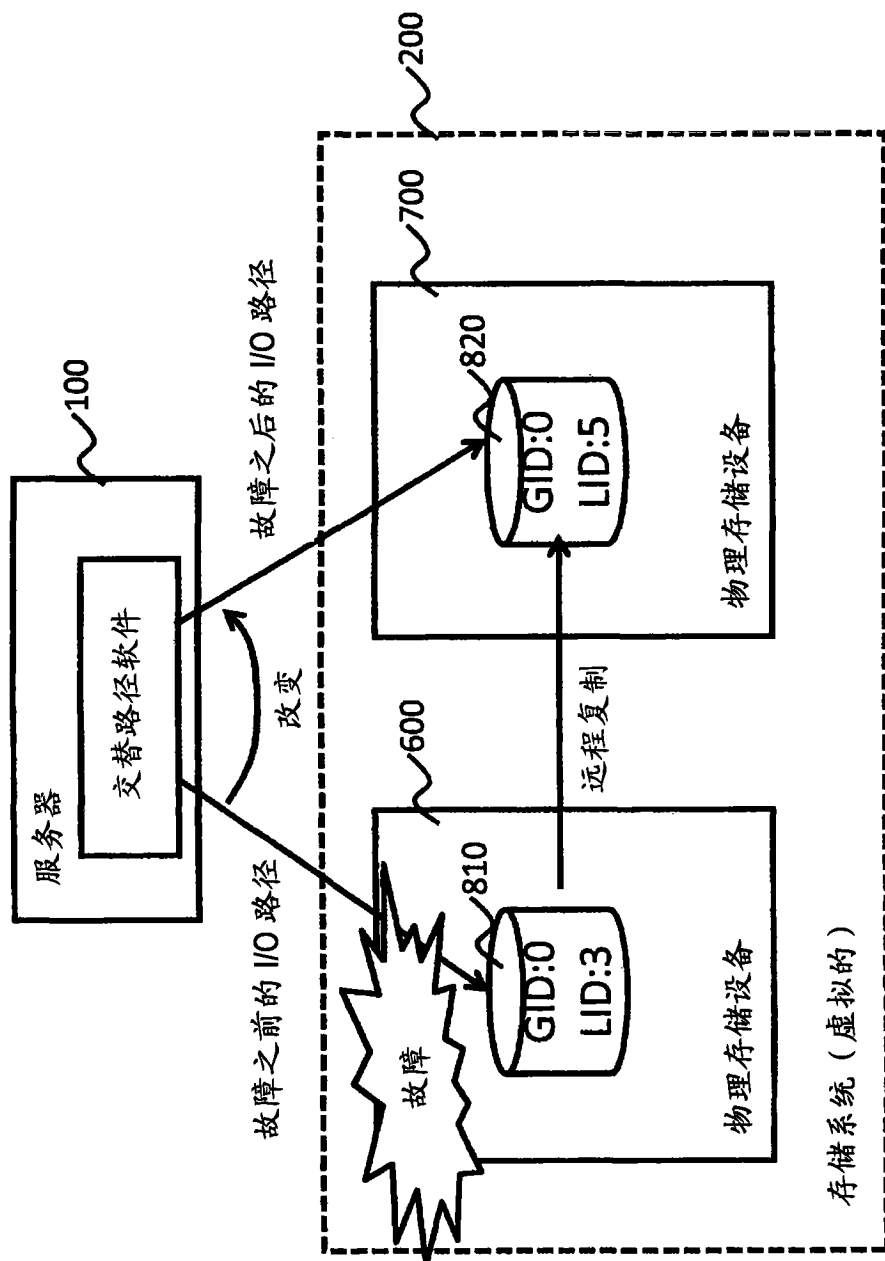


图22

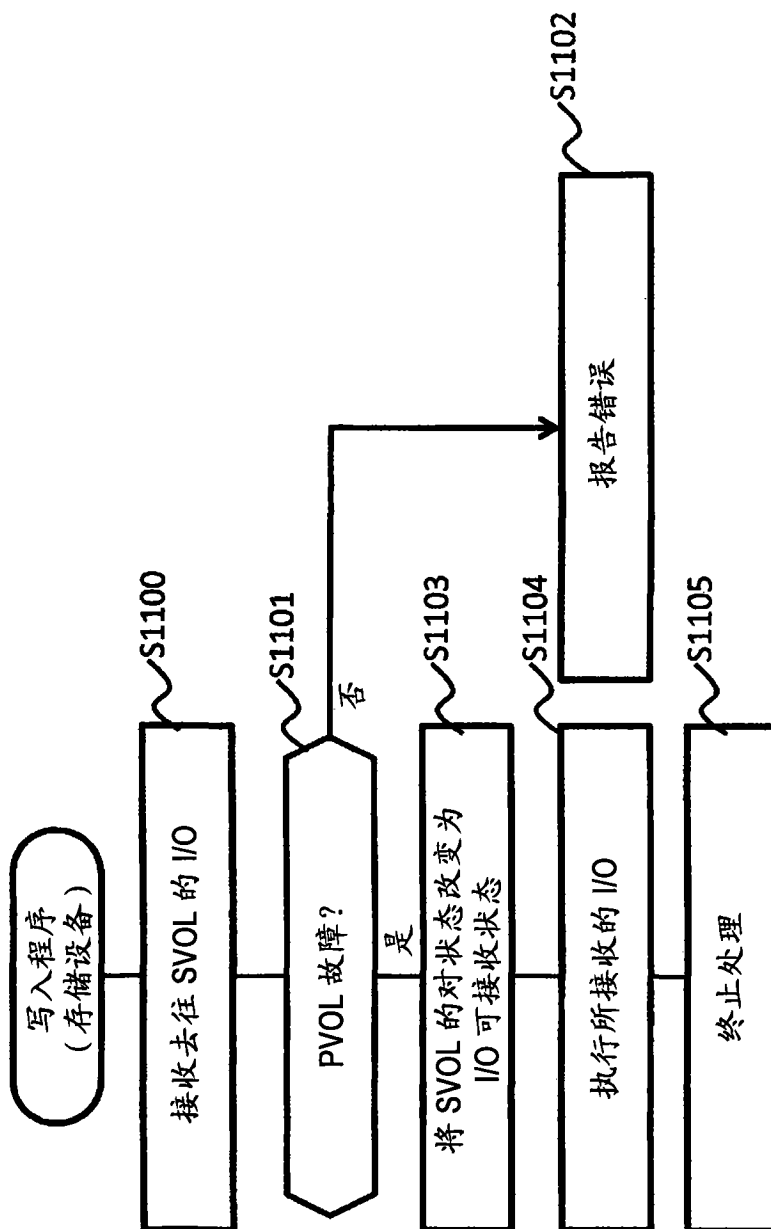


图23

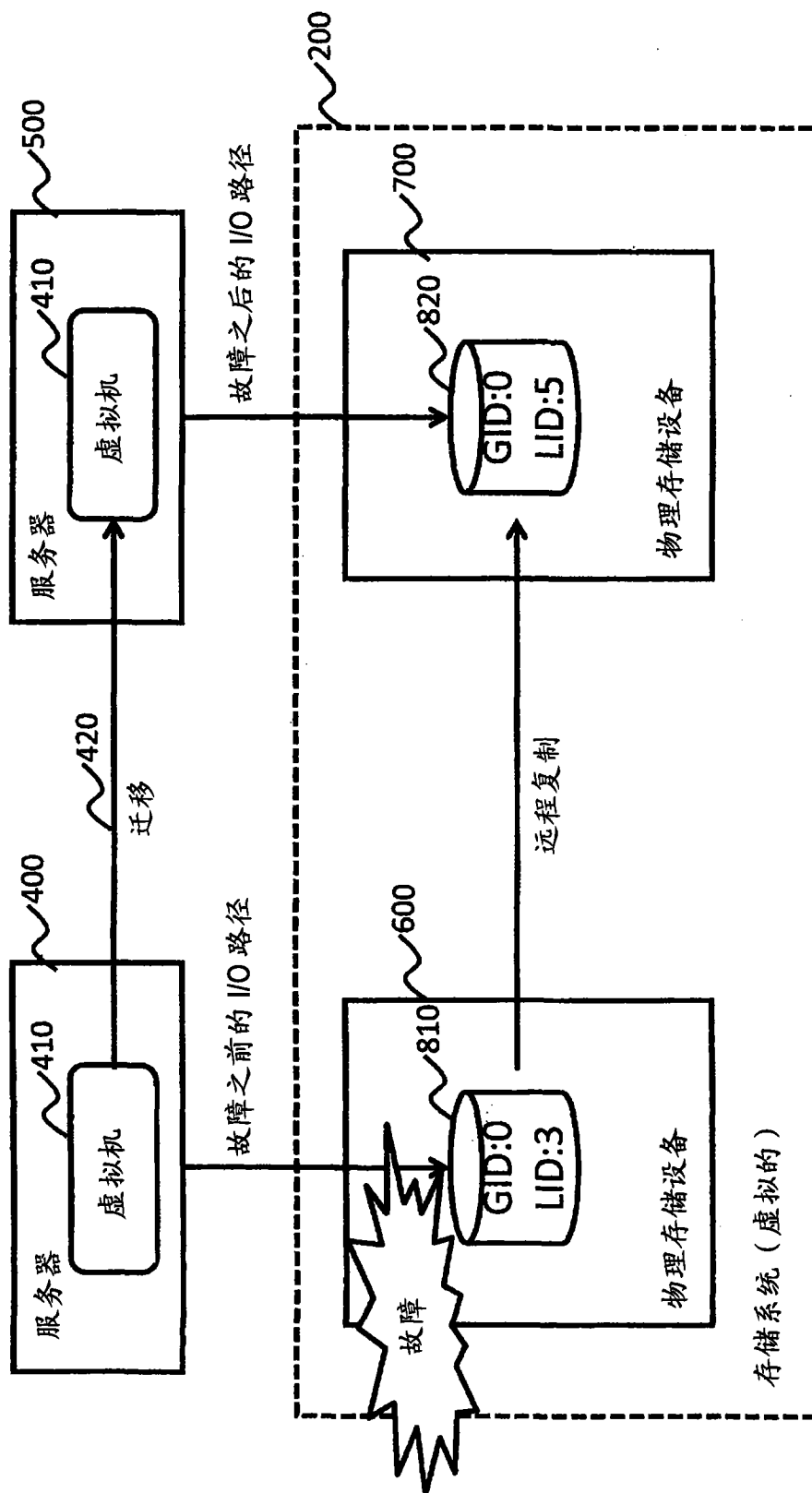


图24

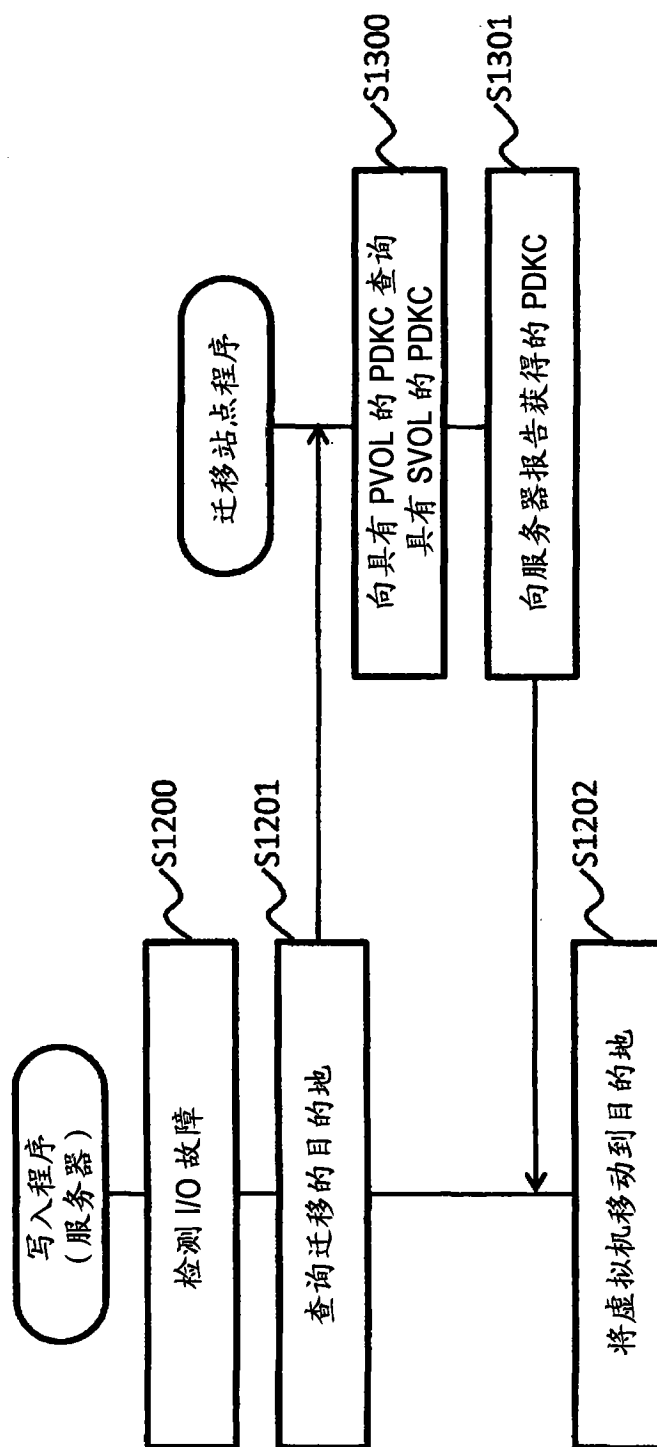


图25

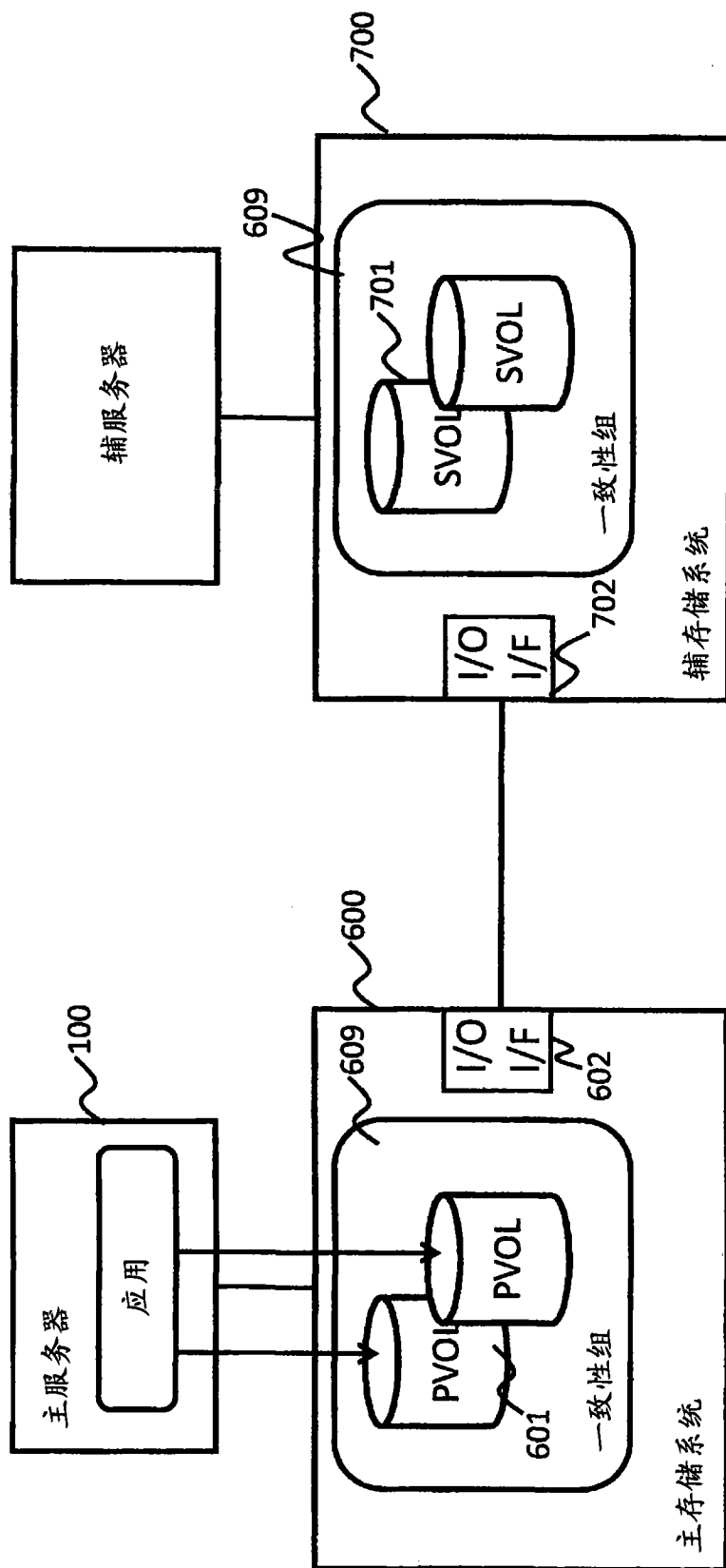


图26

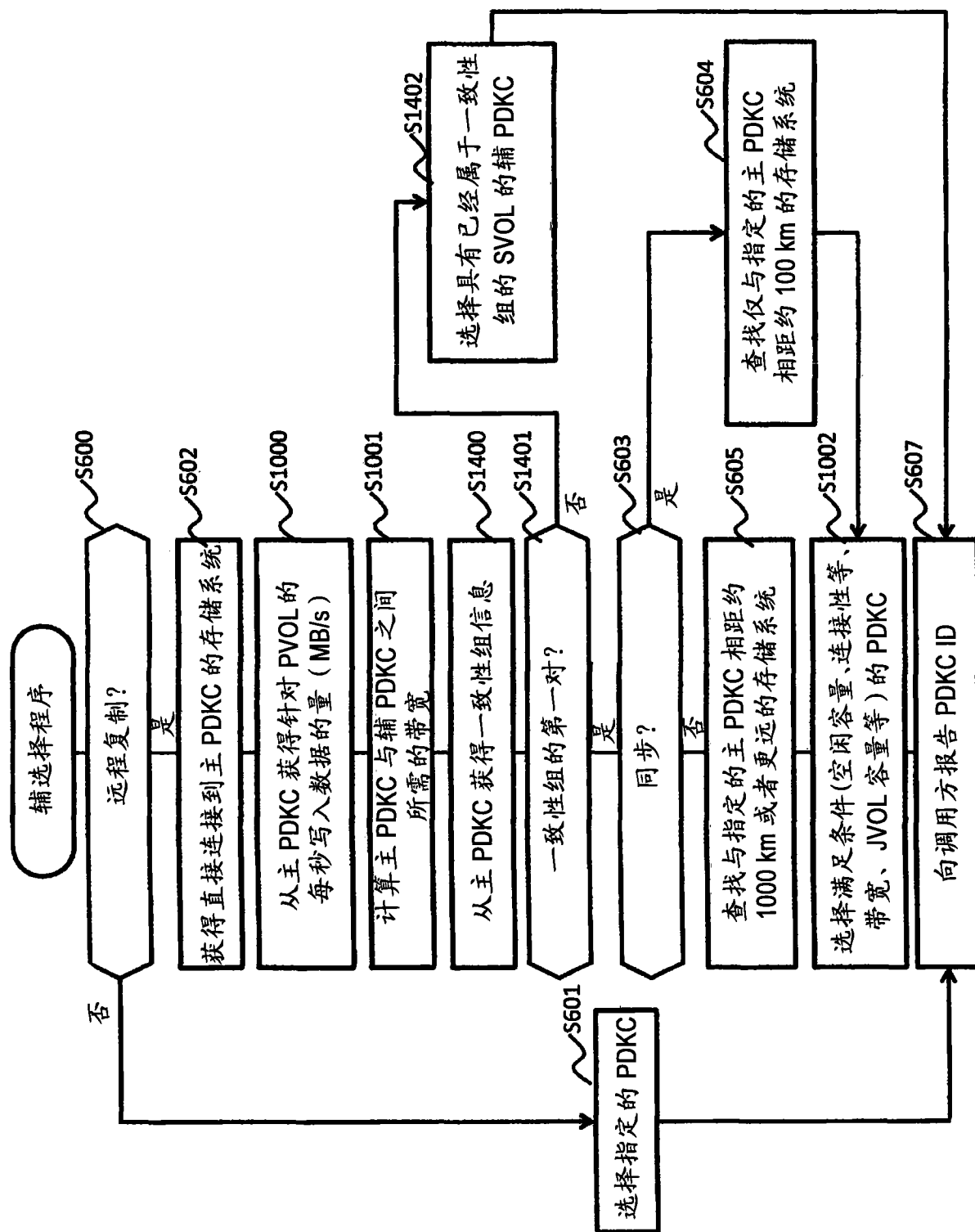


图27

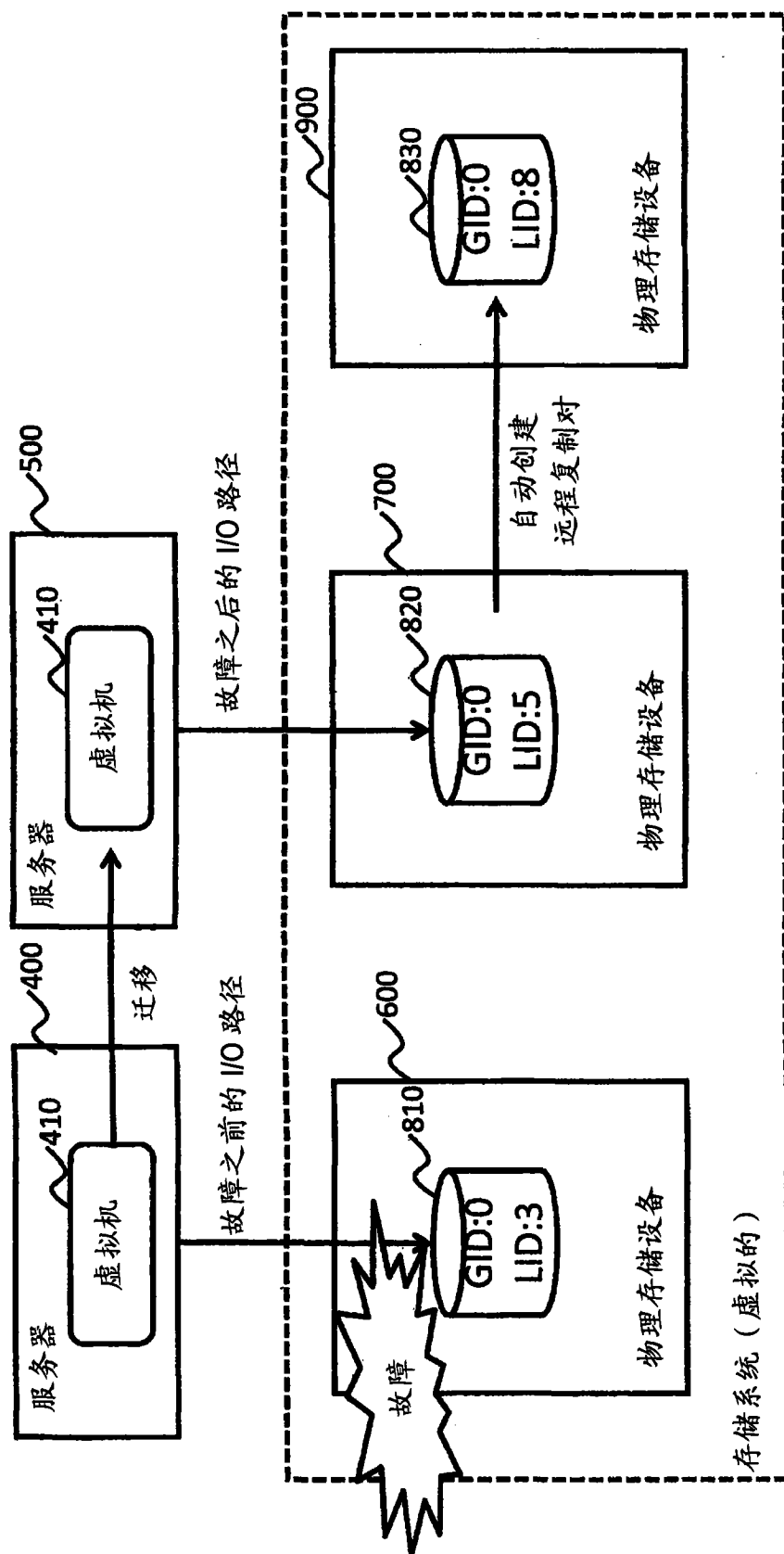


图28

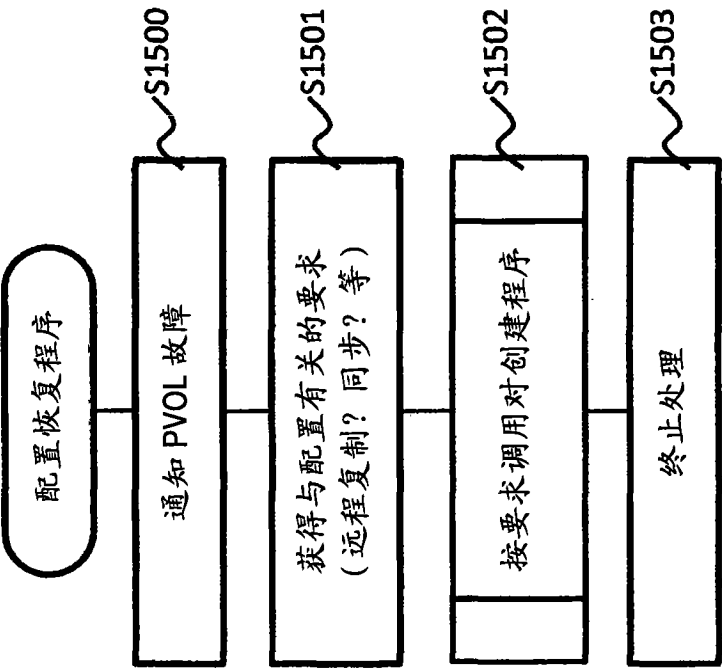


图29

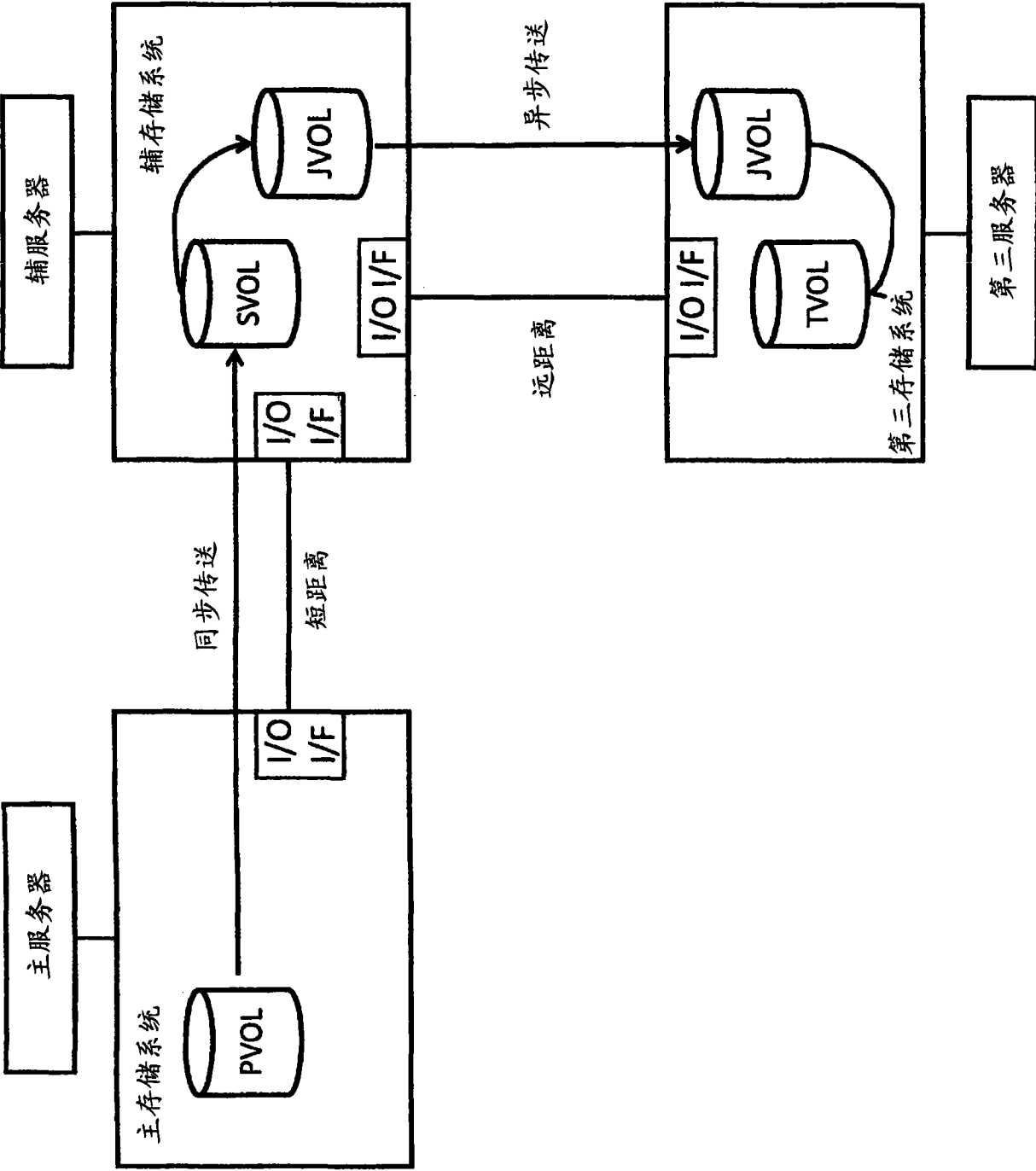


图30

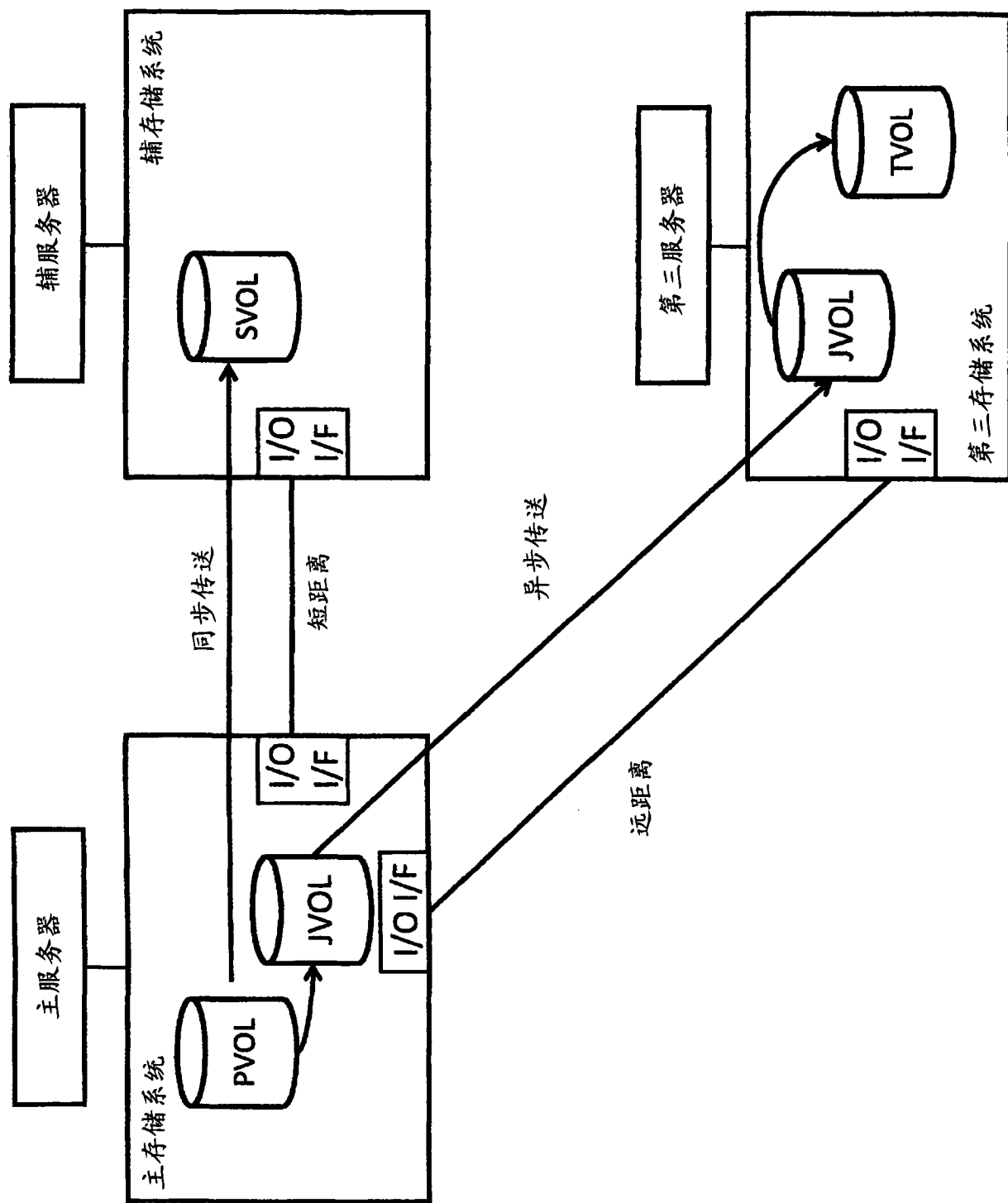


图31

800

设置屏幕						
卷ID	HA	RC(同步)	RC(异步)	级联	多目标	
1			✓			
2					✓	
3			✓			
4	✓					
...	...	...	...	...	...	...

图32

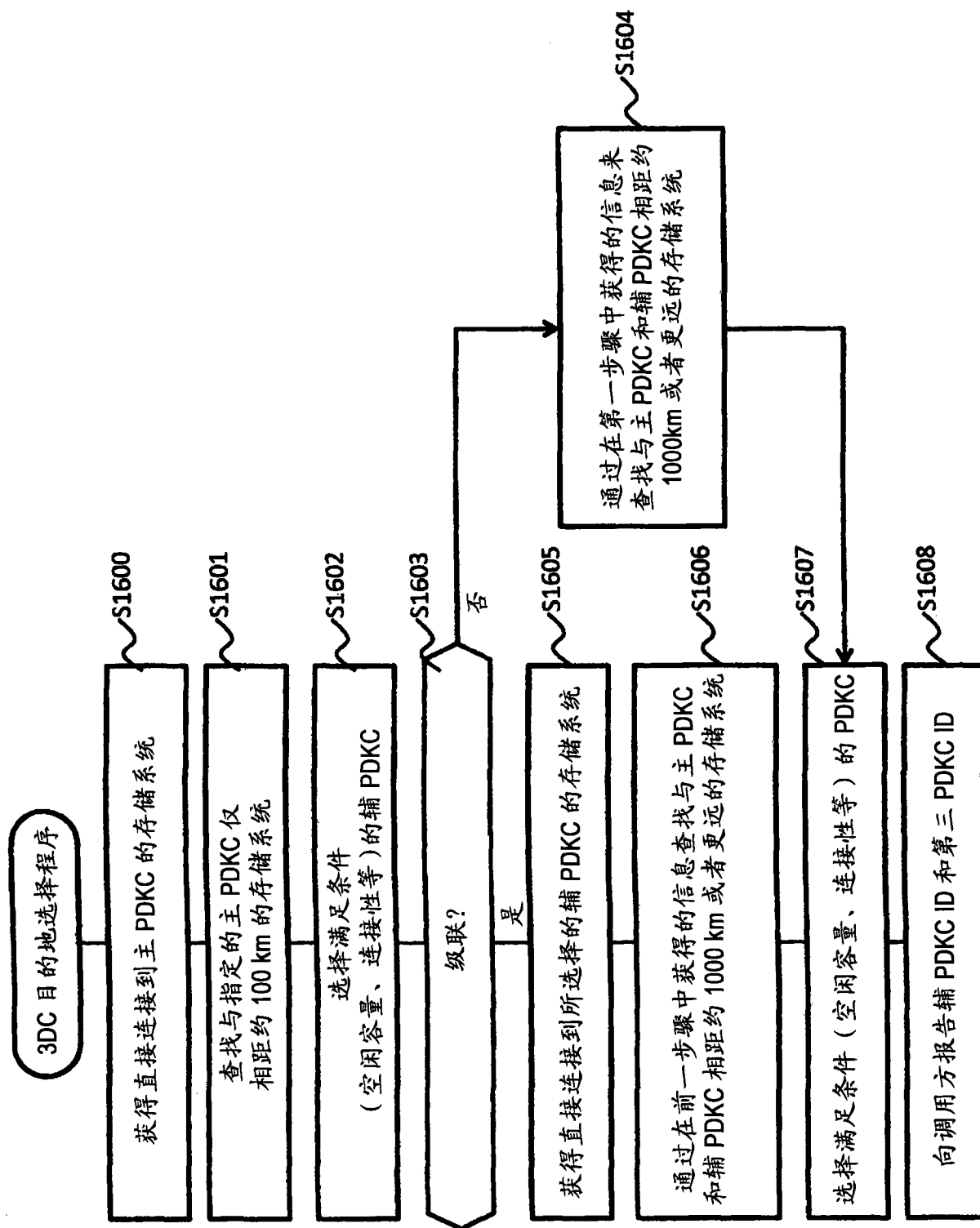


图33