



(12) 发明专利申请

(10) 申请公布号 CN 103827803 A

(43) 申请公布日 2014. 05. 28

(21) 申请号 201180073653. 7

(51) Int. Cl.

(22) 申请日 2011. 10. 28

G06F 3/06 (2006. 01)

H04L 12/26 (2006. 01)

(85) PCT国际申请进入国家阶段日
2014. 03. 24

(86) PCT国际申请的申请数据
PCT/JP2011/006043 2011. 10. 28

(87) PCT国际申请的公布数据
W02013/061380 EN 2013. 05. 02

(71) 申请人 株式会社日立制作所
地址 日本东京都

(72) 发明人 永见明久 兼田泰典

(74) 专利代理机构 北京市金杜律师事务所
11256

代理人 王茂华 陈颖

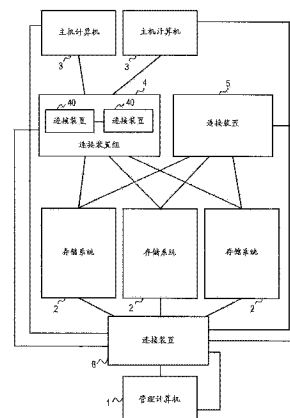
权利要求书3页 说明书23页 附图34页

(54) 发明名称

用于在存储网络中进行性能优化的计算机系统
及管理系统

(57) 摘要

在本发明的一个实施例中的一种计算机系统包括:其中执行多个主机程序的多个主机计算机;多个存储系统,包括向多个主机程序提供的多个卷;以及管理系统;以及管理系统。多个主机程序中的每个主机程序位于主机计算机的端口处。多个卷中的每个卷位于存储系统的端口处。管理系统获得关于在多个存储系统和多个主机计算机之中的数据通信网络中包括的多个路径的带宽的信息。管理系统参考该信息确定用于第一主机程序和第一卷中的至少任一项的新位置的个体候选端口的优先级别。



1. 一种计算机系统,包括:

多个主机计算机,在所述多个主机计算机中执行多个主机程序;

多个存储系统,包括向所述多个主机程序提供的多个卷;以及

管理系统,其中:

所述多个主机程序中的每个主机程序位于主机计算机的端口处;

所述多个卷中的每个卷位于存储系统的端口处;

所述管理系统获得关于包括在所述多个存储系统和所述多个主机计算机之间的数据通信网络中的多个路径的带宽的信息;并且

所述管理系统参考所述信息来确定用于第一主机程序和第一卷中的至少任一项的新位置的个体候选端口的优先级别。

2. 根据权利要求1所述的计算机系统,其中所述管理系统显示用于新位置的所述候选端口和所述优先级别。

3. 根据权利要求1所述的计算机系统,其中所述第一卷位于第一存储系统的第一端口处;

所述信息包括关于在所述第一端口与用于所述多个存储系统中的所述第一卷的新位置的所述候选端口之间的独立路径的带宽的信息;并且

所述管理系统参考关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

4. 根据权利要求3所述的计算机系统,其中:

所述信息包括关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的最小带宽的信息;并且

所述管理系统参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述最小带宽来确定用于新位置的所述个体候选端口的所述优先级别。

5. 根据权利要求3所述的计算机系统,其中:

所述管理系统还获得关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的连接装置的数目的信息;并且

所述管理系统参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的连接装置的数目来确定用于新位置的所述个体候选端口的所述优先级别。

6. 根据权利要求4所述的计算机系统,其中:

所述管理系统还获得关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的最低带宽路径的使用的信息;并且

所述管理系统参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的最低带宽路径的所述使用来确定用于新位置的所述个体候选端口的所述优先级别。

7. 根据权利要求4所述的计算机系统,其中:

所述管理系统还获得关于在用于所述第一卷的所述第一端口处的 I/O 的数量的信息;并且

所述管理系统参考关于所述 I/O 的数量的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

8. 根据权利要求1所述的计算机系统,其中:

所述第一卷是用于所述第一主机程序的卷；

所述第一卷位于第一存储系统的第一端口处；

所述信息包括关于在所述第一端口与用于所述第一主机程序的新位置的所述候选端口之间的独立路径的带宽的信息；并且

所述管理系统参考关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

9. 根据权利要求 1 所述的计算机系统, 其中：

所述第一卷是用于所述第一主机程序的卷；

所述第一主机程序位于第二端口处；

所述信息包括关于在所述第二端口与用于所述第一卷的新位置的所述候选端口之间的独立路径的带宽的信息；并且

所述管理系统参考关于在所述第二端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

10. 一种用于管理计算机系统的管理系统, 所述计算机系统包括：多个主机计算机, 在所述多个主机计算机中执行多个主机程序；以及多个存储系统, 所述多个存储系统包括向所述多个主机程序提供的多个卷, 所述管理系统包括：

处理器；以及

存储装置, 其中：

所述处理器获得关于包括在所述多个存储系统和所述多个主机计算机之间的数据通信网络中的多个路径的带宽的信息, 并且在所述存储装置中存储所述信息；

所述多个主机程序中的每个主机程序位于主机计算机的端口处；

所述多个卷中的每个卷位于存储系统的端口处；

所述多个路径中的每个路径是在所述多个存储系统的端口与所述多个主机计算机或者所述多个存储系统的另一端口之间的路径；并且

所述处理器参考所述信息来确定用于第一主机程序和第一卷中的至少任一项的新位置的个体候选端口的优先级别。

11. 根据权利要求 10 所述的管理系统, 还包括：映像输出装置, 用于输出包括用于新位置的所述候选端口和所述优先级别的映像。

12. 根据权利要求 10 所述的管理系统, 其中：

所述第一卷位于第一存储系统的第一端口处；

所述信息包括关于在所述第一端口与用于所述多个存储系统中的所述第一卷的新位置的所述候选端口之间的独立路径的带宽的信息；并且

所述处理器参考关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

13. 根据权利要求 12 所述的管理系统, 其中：

所述信息包括关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的最小带宽的信息；并且

所述处理器参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述最小带宽来确定用于新位置的所述个体候选端口的所述优先级别。

14. 根据权利要求 12 所述的管理系统,其中:

所述处理器还获得关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的连接装置的数目的信息,并且在所述存储装置中存储所述信息;并且

所述处理器参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的连接装置的所述数目来确定用于新位置的所述个体候选端口的所述优先级别。

15. 根据权利要求 13 所述的管理系统,其中:

所述处理器还获得关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的最低带宽路径的使用的信息,并且在所述存储装置中存储所述信息;并且

所述处理器参考在所述第一端口与用于新位置的所述候选端口之间的所述独立路径中的最低带宽路径的所述使用来确定用于新位置的所述个体候选端口的所述优先级别。

16. 根据权利要求 13 所述的管理系统,其中:

所述处理器还获得关于在用于所述第一卷的所述第一端口处的 I/O 的数量的信息,并且在所述存储装置中存储所述信息;并且

所述处理器参考关于所述 I/O 的数量的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

17. 根据权利要求 10 所述的管理系统,其中:

所述第一卷是用于所述第一主机程序的卷;

所述第一卷位于第一存储系统的第一端口处;

所述信息包括关于在所述第一端口与用于所述第一主机程序的新位置的所述候选端口之间的独立路径的带宽的信息;并且

所述处理器参考关于在所述第一端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

18. 根据权利要求 10 所述的管理系统,其中:

所述第一卷是用于所述第一主机程序的卷;

所述第一主机程序位于第二端口处;

所述信息包括关于在所述第二端口与用于所述第一卷的新位置的所述候选端口之间的独立路径的带宽的信息;并且

所述处理器参考关于在所述第二端口与用于新位置的所述候选端口之间的所述独立路径的所述带宽的所述信息来确定用于新位置的所述个体候选端口的所述优先级别。

用于在存储网络中进行性能优化的计算机系统及管理系统

技术领域

[0001] 本发明涉及一种计算机系统及用于其的管理系统,并且具体地涉及在计算机系统对卷或者主机程序定位。

背景技术

[0002] 计算机处理的信息量的近来剧增促使数据中心的规模扩张,这些数据中心托管用于处理信息的主机计算机和存储系统。在数据中心中安装的服务器、存储系统和网络已经获得技术进步以在个体装置的基础上减少管理成本。

[0003] 对于服务器,随着装置的单位性能增强,一种在高度地集成的刀片型装置中实施虚拟化机制的架构正在开始被运用于实现在多个服务器之间的负荷平衡。向存储系统应用如下设计,该设计在具有高性能、高可靠性和高可用性的大规模存储系统中存储合并的大量数据而不是在一些小存储装置中存储大量数据。关于网络装备,增加每装置的连接数目提高性能以承载在系统之中的增加的流量。

[0004] 在这样的境况中,数据中心开始面临用于采购个体装置并且构造和操作大规模系统的管理成本这一问题。为了减少成本,正在开始提出集成系统,这些集成系统包括具有服务器、存储装置和网络的构建的系统。

[0005] 这样的集成系统包括具有确定的规格的服务器、存储系统和网络的构建的部件;因而,如果假设在它的指定的性能的范围范围内使用它,则可以减少构造和操作成本。

[0006] 对于提高系统性能的问题是存储系统管理。由于难以设计性能和容量以及难以构建存储区域网络(SAN),大规模数据中心运用在存储系统中专门化的管理员;因而,简化是特别希望的,并且技术在开发中。

[0007] 用于实现大规模存储系统的技术之一是存储区域网络。存储区域网络将一个或者多个存储系统连接到多个主机计算机,从而多个主机存储装置可以共享巨大容量存储系统。这样的计算机系统当前是普遍的。

[0008] 这样的计算机系统的优点是显著的可扩展性;可以在以后场合容易添加、算出或者替换存储资源和计算机资源。对于连接到 SAN 的存储系统,普遍使用包括 RAID(独立盘冗余阵列)结构的存储装置的存储系统。

[0009] 在 SAN 中包括的光纤通道交换机提供分区段功能。区段是光纤通道网络的逻辑分离单位。这一光纤通道分区段的方案向区段配置一组网络接口。

[0010] 在区段中注册的网络接口不接受向和从未在区段中注册的网络接口的数据传送。换言之,数据传送仅在区段中注册的网络接口之中可用。光纤信道交换机管理关于连接到 SAN 结构的装备的信息以响应于新装备登入或者登出来更新状态或者对关于连接状态的查询做出响应。

[0011] 专利文献 1 公开一种包括多个虚拟卷和池的存储系统。每个池具有分级结构,该分级结构具有性能不同的多层。存储系统管理指示在虚拟卷与实际区域之间的关系的信息和指示对实际区域的访问频率的信息。管理系统具有指示虚拟卷的所需性能和实际性能的

管理信息。管理系统参考管理信息确定在不适当条件中的虚拟卷并且执行涉及确定的虚拟卷的池的迁移以改变其条件以令人满意。

[0012] 引用列表

[0013] 专利文献

[0014] PTL1 :W02011/092738A

发明内容

[0015] 技术问题

[0016] 如果存储系统需要的系统性能扩大到超过单个存储系统的性能,则重要的是使用多个存储系统作为资源以用于扩张系统性能和减少管理成本。

[0017] 然而在多个存储系统与 SAN 互连用于数据传送并且可以经由 SAN 传送卷这样的配置(单存储池配置)中, SAN 可能未在主机计算机和多个存储系统之中统一。

[0018] 在这样的系统中,从存储系统向在管理系统指示的可传送卷范围内的另一存储系统的卷迁移可能引起用于主机计算机的卷的 I/O (输入 / 输出) 性能的显著改变。具体而言,如果在目的地的存储系统与主机计算机之间的路径的 I/O 带宽窄,则用于主机计算机的卷的 I/O 性能下降。

[0019] 在构成单存储池的存储系统之一中部署虚拟机的另一情况下, I/O 性能可以根据其中部署虚拟机的存储系统而不同,因为 SAN 未在主机计算机与存储系统之中统一。具体而言,如果在存储系统与主机计算机之间的路径的 I/O 带宽窄,则用于部署的虚拟机的卷的 I/O 性能将为低。

[0020] 对问题的解决方案

[0021] 本发明的一个方面是一种计算机系统,该计算机系统包括:其中执行多个主机程序的多个主机计算机;多个存储系统,包括向多个主机程序提供的多个卷;以及管理系统。多个主机程序中的每个主机程序位于主机计算机的端口处。多个卷中的每个卷位于存储系统的端口处。管理系统获得关于在多个存储系统和多个主机计算机之中的数据通信网络中包括的多个路径的带宽的信息。管理系统参考该信息确定用于第一主机程序和第一卷中的至少任一项的新位置的个体候选端口的优先级别。

[0022] 本发明的有利效果

[0023] 本发明的方面实现对卷和主机程序的的适当位置的选择。实施例的以下描述将阐明待解决的其它问题、配置和有利效果。

附图说明

[0024] [图 1] 图 1 是在第一实施例中的整个控制系统的配置图。

[0025] [图 2] 图 2 是举例说明在第一实施例中的管理计算机的硬件配置的图。

[0026] [图 3] 图 3 是举例说明在第一实施例中的存储系统的硬件配置的图。

[0027] [图 4] 图 4 是举例说明在第一实施例中的主机计算机的硬件配置的图。

[0028] [图 5] 图 5 是举例说明在第一实施例中的连接装置的硬件配置的图。

[0029] [图 6] 图 6 是举例说明在第一实施例中的存储系统中的非易失性存储装置和存储的信息的卷配置的图。

[0030] [图 7] 图 7 是举例说明在第一实施例中的管理计算机中的非易失性存储装置中存储的信息的图。

[0031] [图 8] 图 8 是图示在第一实施例中的管理计算机中的存储系统管理表的配置示例的图。

[0032] [图 9] 图 9 是图示在第一实施例中的管理计算机中的连接装置管理表的配置示例的图。

[0033] [图 10] 图 10 是图示在第一实施例中的管理计算机中的传送范围管理表的配置示例的图。

[0034] [图 11] 图 11 是图示在第一实施例中的管理计算机中的存储拓扑管理阈值表的配置示例的图。

[0035] [图 12] 图 12 是图示在第一实施例中的管理计算机中的卷管理表的配置示例的图。

[0036] [图 13] 图 13 是图示在第一实施例中的管理计算机中的存储系统使用管理表的配置示例的图。

[0037] [图 14] 图 14 是图示在第一实施例中的管理计算机中的精简配备使用管理表的配置示例的图。

[0038] [图 15] 图 15 是图示在第一实施例中的管理计算机中的目的地推荐级别表的配置示例的图。

[0039] [图 16] 图 16 是举例说明在第一实施例中的存储系统中的管理信息存储部中存储的信息的图。

[0040] [图 17] 图 17 是图示在第一实施例中的存储系统中的存储拓扑管理表的配置示例的图。

[0041] [图 18] 图 18 是举例说明在第一实施例中的主机计算机中的非易失性存储装置中存储的信息的图。

[0042] [图 19] 图 19 是图示在第一实施例中的主机计算机中的连接的存储系统的卷管理表的配置示例的图。

[0043] [图 20] 图 20 是举例说明在第一实施例中的连接装置中的非易失性存储装置中存储的信息的图。

[0044] [图 21] 图 21 是图示在第一实施例中的连接装置中的连接装置管理表的配置示例的图。

[0045] [图 22] 图 22 是图示在第一实施例中的卷迁移的示例的序列图。

[0046] [图 23] 图 23 是图示在第一实施例中的计算候选目的地端口的推荐级别的方法的示例的流程图。

[0047] [图 24] 图 24 是图示在第一实施例中的更新存储拓扑管理表的示例的序列图。

[0048] [图 25] 图 25 是举例说明在第二实施例中的存储系统中的管理信息存储部中存储的信息的图。

[0049] [图 26] 图 26 是图示在第二实施例中的存储系统中的卷访问条件管理表的配置示例的图。

[0050] [图 27A] 图 27A 是示出在第二实施例中的计算候选目的地端口的推荐级别的方法

的示例的流程图的一个部分的图。

[0051] [图 27B] 图 27B 是示出在第二实施例中的计算候选目的地端口的推荐级别的方法的示例的流程图的另一部分的图。

[0052] [图 28] 图 28 是举例说明在第三实施例中的管理计算机中的非易失性存储装置中存储的信息的图。

[0053] [图 29] 图 29 是图示在第三实施例中的管理计算机中的主机管理表的配置示例的图。

[0054] [图 30] 图 30 是图示在第三实施例中的管理计算机中的主机拓扑管理阈值表的配置示例的图。

[0055] [图 31] 图 31 是举例说明在第三实施例中的存储系统中的管理信息存储部中存储的信息的图。

[0056] [图 32] 图 32 是举例说明在第三示例中的存储系统中的主机拓扑管理表的配置示例的图。

[0057] [图 33] 图 33 是图示在第三实施例中的计算候选虚拟机位置端口的推荐级别的方法的示例的流程图。

[0058] [图 34] 图 34 是图示在第三实施例中的计算候选虚拟机位置端口的推荐级别的方法的示例的流程图。

具体实施方式

[0059] 下文将参照附图描述本发明的实施例。为了说明清楚，以下描述和附图如适合的那样包含省略和简化。实施例中的任一实施例的配置的部分可以替换为不同实施例的配置的部分。实施例中的任一实施例的配置可以被添加到不同实施例的配置。实施例举例说明本发明的一些方面，但是本发明未仅限于实施例。

[0060] 第一实施例

[0061] 在第一实施例中的信息存储系统包括多个存储系统并且可以在存储系统之中传送卷(数据)。在优选配置中，信息存储系统可以从一个存储系统向另一存储系统传送卷而不停止来自主机的 I/O(实况卷迁移)。这样的允许从一个存储系统向另一存储系统传送卷的信息存储系统也被称为单存储池。

[0062] 主机计算机可以经由网络访问在信息存储系统中的多个存储系统中的卷。操作系统(OS)在主机计算机上运行，该 OS 是主机程序或者虚拟机(包括客户 OS)。

[0063] 第一实施例参考关于在主机计算机与存储系统之中的数据网络中的存储系统之间的路径的配置和带宽性质的信息来确定用于卷的候选目的地端口的优先级别。候选目的地端口是在当前存储系统和不同存储系统中的端口。这一确定造成从主机计算机对卷的访问在卷迁移之前与之后之间的更少性能改变。

[0064] 为了计算候选目的地端口的优先级别，该实施例使用关于在源存储系统端口与目的地存储系统端口之间的路径的配置和带宽性质的信息。这使得能够在没有关于在主机计算机与存储系统之间的路径的信息的情况下确定用于卷的适当目的地。

[0065] 第一实施例参考在路径中的中转连接装置(交换机)数目作为关于路径的配置信息的示例，而在路径中的最小带宽作为路径的带宽性质的示例。在示例中，第一实施例还参

考使用路径以确定优先级别。参考这样的信息造成恰当确定候选目的地存储系统的优先级别。下文将参照附图说明第一实施例的更具体配置。

[0066] 图 1 是用于图示在第一实施例中的整个计算机系统的配置示例的图。如在这一幅图中所示,计算机系统包括管理计算机 1、多个(在图 1 的示例中为三个)存储系统 2、一个或者多个(在图 1 的示例中为两个)主机计算机 3、包括多个连接装置 40 的连接装置组 4 以及连接装置 5 和 6。在这一计算机系统连接装置是交换机,这些交换机分析在它们的输入端口接收的数据以选择用于发送数据的输出端口并且从选择的输出端口发送接收的数据。例如连接装置是光纤信道交换机。

[0067] 管理计算机 1、存储系统 2、主机计算机 3、在连接装置组 4 中的连接装置 40 以及连接装置 5 和 6 中的每项具有一个或者多个端口;装置(包括管理计算机 1、存储系统 2 和主机计算机 3)的每个端口连接到不同装置的端口(相反端口)。端口通常用线缆连接到它的相反端口。

[0068] 连接装置组 4 连接主机计算机 3 和存储系统 2 以相互可通信并且在它们之中传送数据(主机数据)。在连接装置组 4 中的每个连接装置 40 具有多个端口;连接装置 40 的每个端口连接到主机计算机 3、存储系统 2 或者在连接装置组 4 中的不同连接装置 40 的端口。

[0069] 连接装置组 4 向存储系统 2 传送来自主机计算机 3 的数据 I/O 请求(对于输入和输出数据的请求)和写入数据。此外,它向主机计算机 3 传送来自存储系统 2 的读取数据。连接装置组 4 还包括连接存储系统 2 的路径。然而在这一架构中,经由连接装置 5 执行从存储系统 2 向另一存储系统的卷(数据)传送。

[0070] 连接装置 5 向另一存储系统传送来自存储系统 2 的数据。连接装置 5 具有连接到存储系统 2 的端口的多个端口。连接装置 5 在存储系统 2 之中传送数据 I/O 请求和请求的数据。在这一实施例中的存储系统 2 从存储系统 2 向另一存储系统传送卷(数据)。经由连接装置 5 执行卷迁移。

[0071] 连接装置 6 将管理计算机 1 连接到其它装置、即存储系统 2、主机计算机 3、在连接装置组 4 中的连接装置 40 和连接装置 5,以可通信并且在管理计算机 1 与其它装置之间传送数据。连接装置 6 具有多个端口;每个端口连接到管理计算机 1 或者其它装置中的任一装置的端口。连接装置 6 向存储系统 2、主机计算机 3、在连接装置组 4 中的连接装置 40 和连接装置 5 传送来自管理计算机 1 的对于管理通信的请求并且向管理计算机 1 传送它们的响应。

[0072] 在这一计算机系统中,在主机计算机 3 和存储系统 2 之中用连接装置组 4 配置 SAN(存储区域网络)。在存储系统 2 之中用连接装置 5 配置另一 SAN。存储系统 2 可以在没有连接装置 5 的情况下直接相互连接。在这一示例中,可以使用连接装置 6 经由 IP 网络从管理计算机 1 连接整个计算机系统。

[0073] 连接装置组 4 组成的网络和连接装置 5 组成的网络可以是在物理上相同的装置提供的在逻辑上独立的网络。连接装置组 4 组成的网络和连接装置 5 组成的网络可以用 iSCSI(因特网小型计算机系统接口)配置的 IP 网络。备选地,它们可以是连接装置 6 提供的 IP 网络的相同装置提供的在逻辑上独立的网络。

[0074] 另外,可以用 FCoE(在以太网之上的光纤信道)配置连接装置组 4 组成的以及连接装置 5 和 6 组成的网络;它们可以是相同装置提供的在逻辑上独立的网络。

[0075] 图 2 图示管理计算机 1 的硬件配置的示例。管理计算机 1 是包括高速缓冲存储器 11、非易失性存储装置 12、处理器 13、程序存储器 14、管理通信 I/F16、内部总线 17、映像输出装置 18、输入装置 19 的计算机。非易失性存储装置 12 可以是经由网络连接的外部存储装置。

[0076] 管理计算机 1 向程序存储器 14 加载在非易失性存储装置 12 中存储的软件、比如操作系统(OS)和应用程序;处理器 13 从程序存储器 14 取回它们并且执行它们以提供预定功能。处理器 13 可以包括多个芯片或者多个封装。这同样适用于其它装置(包括存储系统 2)。

[0077] 管理计算机 1 在高速缓冲存储器 11 中暂时存储频繁使用的数据以增强处理速度。管理计算机 1 包括输入装置 19、比如键盘和鼠标以及映像输出装置 18、比如监视器。输入装置 19 接受来自管理员(用户)的输入,并且映像输出装置 18 输出处理器 13 指定的信息。管理计算机 1 还包括用于连接到连接装置 6 的管理通信 I/F16。

[0078] 在图 1 的配置示例中,用管理计算机 1 配置管理系统,但是可以用多个计算机配置管理系统:一个计算机可以是用于显示的计算机,而其它多个计算机可以执行与在管理计算机中的任务等效的任务以求快速管理和更高可靠性。管理系统可以包括存储系统或者连接装置的功能的一部分。

[0079] 图 3 图示存储系统 2 的硬件配置的示例。存储系统 2 包括高速缓冲存储器 21、非易失性存储装置 22、处理器 23、程序存储器 24、数据 I/O 通信 I/F25、管理通信 I/F26、内部总线 27、映像输出装置 28 和输入装置 29。

[0080] 存储系统 2 向程序存储器 24 加载在非易失性存储装置 22 中存储的软件、比如 OS 和控制程序;处理器 23 从程序存储器 24 取回它们并且执行它们以提供预定功能。处理器 23 充当存储系统 2 的控制器。存储系统 2 在高速缓冲存储器 21 中存储频繁使用的数据。

[0081] 非易失性存储装置 22 包括多个存储驱动。存储驱动包括具有非易失性磁盘的硬盘驱动(HDD)和具有在其中嵌入的非易失性半导体存储器(比如闪存)的固态驱动(SSD)。存储驱动存储从主机计算机 3 发送的数据。多个存储驱动通过 RAID 操作实现数据冗余性从而防止在某个存储驱动出现故障时丢失数据。稍后将描述在非易失性存储装置 22 中的卷配置。

[0082] 存储系统 2 包括输入装置 29、比如键盘和鼠标以及映像输出装置 28、比如监视器、输入装置 29 接受来自管理员(用户)的输入,并且映像输出装置 28 输出处理器 23 指定的信息。存储系统 2 包括各自用于连接到连接装置组 4 和连接装置 5 的数据 I/O 通信 I/F25 以及用于连接到连接装置 6 的管理通信 I/F26。管理员可以在经由连接装置 6 从管理计算机 1 访问存储系统 2 之时管理和控制存储系统 2。

[0083] 为了简化,图 1 示出三个存储系统 2,但是存储系统数目依赖于系统设计。向数据 I/O 通信 I/F25 指派计算机系统唯一的端口 ID,并且可以使用 NPIV(N_端口 ID 虚拟化)或者任何其它虚拟化方案来指派多个虚拟端口 ID。

[0084] 例如 NPIV 虚拟化 N_端口的 ID 并且向一个物理端口(N_端口)指派多个虚拟端口。NPIV 可以向一个物理端口指派多个主机计算机 3 或者虚拟机的虚拟端口。为了使用虚拟端口,虚拟端口 ID 与在计算机系统内的物理端口 ID 关联。在以下说明中,计算机、存储系统或者连接装置的术语端口 ID 除非具体指示,否则意味着物理端口的 ID。

[0085] 图 4 图示主机计算机 3 的硬件配置的示例。主机计算机 3 是包括高速缓冲存储器 31、非易失性存储装置 32、处理器 33、程序存储器 34、数据 I/O 通信 I/F35、管理通信 I/F36、内部总线 37、映像输出装置 38 和输入装置 39 的计算机。

[0086] 主机计算机 3 向程序存储器 34 加载在非易失性存储装置 32 中存储的软件、比如 OS 和应用程序；处理器 33 从程序存储器 34 取回它们并且执行它们以提供预定功能。处理器 3 在高速缓冲存储器 31 中存储频繁使用的数据。

[0087] 主机计算机 3 包括输入装置 39、比如键盘和鼠标以及映像输出装置 38、比如监视器。输入装置 39 接受来自管理员(用户)的输入，并且映像输出装置 38 输出处理器 33 指令的信息。主机计算机 3 包括用于连接到连接装置组 4 的一个或者多个数据 I/O 通信 I/F35 和用于连接到连接装置 6 的管理通信 I/F36。

[0088] 为了简化，图 1 示出两个主机计算机 3，但是系统可以包括任何数目的主机计算机 3。向数据 I/O 通信 I/F35 指派计算机系统唯一的端口 ID，并且可以使用 NPIV 或者任何其它虚拟化方案来指派多个虚拟端口 ID。例如 NPIV 可以向主机计算机 3 的物理端口指派多个虚拟机的虚拟端口。

[0089] 图 5 图示在连接装置组 4 中的连接装置 40 的硬件配置的示例。连接装置包括高速缓冲存储器 41、非易失性存储装置 42、处理器 43、成像存储器 44、数据 I/O 通信 I/F45、管理通信 I/F46 和内部总线 47。

[0090] 连接装置 40 向程序存储器 44 加载在非易失性存储装置 42 中存储的软件、比如 OS 和控制程序；处理器 43 从程序存储器 44 取回它们并且执行它们以提供预定功能。连接装置 40 在高速缓冲存储器 41 中存储频繁使用的数据。连接装置 5 和 6 可以具有与连接装置 40 相同的配置。

[0091] 图 6 图示在存储系统 2 中的非易失性存储装置 22 的配置。非易失性存储装置 22 包括存储系统管理信息存储部 221 和用非易失性存储介质装配的不同类型的介质 222A 至 222C。在这一实施例中，包括一个或者多个存储驱动及其接口提供的存储区域的部件为此被称为介质。通常，介质是包括 RAID 组及其接口提供的存储区域的部件。

[0092] 在图 6 的示例中，举例说明利用半导体存储器的闪存的 SSD 介质 222A、包括 SAS 接口的串行附着 SCSI (SAS) 介质 222B 和包括 SATA 接口的串行高级技术附着 (SATA) 介质 222C。非易失性存储装置 22 通常包括多个 SSD 介质 222A、多个 SAS 介质 222B 和多个 SATA 介质 222C。

[0093] 存储系统 2 提供由介质提供的实际存储区域形成的并且从主机计算机 3 或者精简配备池 223 可读取和可写入的一个或者多个卷(实际卷) 2221。每个卷 2221 可以由单个类型的存储驱动的存储区域组成或者可以从用多个类型的存储驱动配置的 RAID(独立盘冗余阵列) 被提供。下文将描述其中每个卷 2221 由单个类型的存储驱动的存储区域形成的示例。

[0094] 在精简配备池 223 中，将管理员(用户)从介质 222A 至 222C 的卷 2221 之中选择的卷 2221 分离成多个页面 2231。以页面为单位管理精简配备池 223。根据将在精简配备卷 224 中存储的数据量，向精简配备卷 224 分配页面。

[0095] 精简配备卷 224 是向主机计算机 3 提供的虚拟卷，并且虚拟化容量。每当主机计算机 3 向精简配备卷 224 写入数据以引起需要数据存储区域时，存储系统 2 向精简配备卷

224 分配页面 2231。

[0096] 精简配备可以使主机计算机 3 识别的精简配备卷 224 的容量大于向精简配备卷 224 分配的实际容量(所有页面的总容量)并且可以使用于实现向主机计算机 3 分配的容量的实际容量小于分配的容量。

[0097] 在图 6 中,精简配备卷 223 包括多个实际卷(池卷) 2221。精简配备卷 223 通常包括在存储系统 2 中的性能能力不同的多个介质并且根据访问性能具有分级结构,该分级结构具有多层。每层由一个或者多个实际卷 2221 组成。

[0098] 存储系统 2 分析从主机计算机 3 到精简配备卷 224 的 I/O 并且自动向由高性能高成本介质的资源组成的高级别层分配 I/O 负荷高的页面而向由更低性能更低成本的资源组成的更低级别分配其它页面。

[0099] 虽然在图 6 中未示出,但是存储系统 2 可以提供常规属性的卷,这些卷的容量未被虚拟化,并且主机识别的容量等于实际容量。

[0100] 图 7 图示在管理计算机 1 中的非易失性存储装置 12 中存储的信息。存储系统管理程序 121 检查在计算机系统 1 中的存储系统 2 的状态并且通过经由从管理通信 I/F16 连接连接装置 6 的通信来设置关于它们的配置信息。连接装置管理程序 122 检查在连接装置组 4 中的连接装置 40 和连接装置 5 的状态并且通过经由从管理通信 I/F16 连接连接装置 6 的通信来设置关于它们的配置信息。

[0101] 主机计算机管理程序 123 检查在计算机系统 1 中的主机计算机 3 的状态,并且通过经由从管理通信 I/F16 连接连接装置 6、与在主机计算机 3 中的代理 322 的通信来设置关于它们的配置信息。操作管理程序 124 参考在管理计算机 1 中的非易失性存储装置 12 中的表设计卷位置并且执行其它作业。操作管理程序 124 还与各种程序配合地管理在计算机系统 1 中的操作并且向管理员提供用户界面。操作系统 125 是用于执行各种程序和用于管理各种表的平台。

[0102] 程序由处理器 13 执行以使用存储装置和通信接口来执行预定作业。因而在这一实施例中,以程序为主题的说明可以是以处理器为主题的说明。程序执行的作业是程序在其上运行的计算机或者计算机系统执行的作业。

[0103] 处理器 13 根据程序工作,以作为用于实施预定功能的操作部来工作。例如处理器 13 根据存储系统管理程序 121 工作,以充当存储系统管理器,并且根据连接装置管理程序 122 工作,以充当连接装置管理器。这同样适用于其它程序。

[0104] 另外,处理器 13 也作为用于实施程序执行的个体作业的操作部工作。包括处理器 13 和程序的管理计算机 1 是包括这些操作部的装置。关于在程序与处理器之间的关系和在这些与包括它们的装置之间的关系说明适用于其它装置和系统、即每个主机计算机 3、在连接装置组 4 中的每个连接装置、连接装置 5 或者 6 和每个存储系统 2。

[0105] 非易失性存储装置 12 还存储存储系统管理表 100、连接装置管理表 110、传送范围管理表 130、存储拓扑管理阈值表 140、卷管理表 150、存储系统使用管理表 160、精简配备使用管理表 170 和目的地推荐级别表 180。

[0106] 在这一实施例中,管理计算机 1 使用的信息不依赖于数据结构,而是可以在任何数据结构中被表达。例如从表、列表、数据库和队列适当选择的数据结构可以存储信息。这同样适用于其它装置(包括存储系统)。

[0107] 图 8 图示在管理计算机 1 中的存储系统管理表 100 的配置示例。存储系统管理表 100 具有用于存储用于管理在计算机系统 2 中的存储系统 ID 列 1001 和管理端口 IP 地址列 1002。管理端口 IP 地址是用于管理计算机 1 与存储系统 2 通信的 IP 地址。

[0108] 例如这些值由管理员输入。在这一实施例中,使用 ID、名称和编号作为用于标识对象的信息,但是这些名义标示可以相互替换。在计算机系统中添加或者删除存储系统 2 时添加或者删除存储系统管理表 100 的条目。

[0109] 图 9 是图示在管理计算机 1 中的连接装置管理表 110 的配置示例。连接装置管理表 110 具有连接装置 ID 列 1101 和管理端口 IP 地址列 1102。管理员通过操作管理程序 124 输入在每列中的值。连接装置 ID 是计算机系统唯一的标识符。管理端口 IP 地址是用于管理计算机 1 与连接装置通信的 IP 地址。连接装置 1 参考这一个表用于与连接装置组 4 或者连接装置 5 或者 6 的管理通信。

[0110] 图 10 图示在管理计算机 1 中的传送范围管理表 130 的配置示例。传送范围管理表 130 定义卷可以被传送到范围、具体为物理端口,这些物理端口可以是卷的目的地。在计算机系统中用存储系统 ID 和端口 ID 唯一标识物理端口。传送范围管理表 130 是用于存储关于单存储池和存储系统 2 以及它们的在个体单存储池中包括的端口的信息的表。

[0111] 在这一示例中,可以在卷属于的单存储池内传送卷。主机计算机 3 可以访问在单存储池中的任何端口并且在单存储池中执行实况卷迁移(在不停止系统的情况下传送卷)。可以根据设计而不同地定义卷传送范围。

[0112] 如图 10 中所示,传送范围管理表 130 包括单存储池 ID 列 1301、存储系统 ID 列 1302 和端口 ID 列 1303。例如在图 10 的传送范围管理表 130 中,单存储池 SPoo11 由存储系统 ST1、ST2 和 ST3 构成。可以向在单存储池 SPoo11 中的不同端口传送单存储池 SPoo11 的端口中的任一端口已经被指派到的卷。

[0113] 可以从保持卷的存储系统 2 向在相同单存储池内的不同存储系统 2 传送卷;否则,可以在保持卷的存储系统 2 内传送它。在相同存储系统 2 内的这样的传送是指派的端口的改变。

[0114] 图 11 示出管理计算机 1 将使用的存储拓扑管理阈值表 140 的示例。存储拓扑是在连接存储系统 2 和主机计算机 3 的网络(连接装置组 4 组成的网络)中的连接存储系统 2 的网络的拓扑。具体而言,它包括关于在存储系统 2 之间的路径的配置和带宽的信息。这一示例运用在路径中的中转连接装置数目作为关于路径的配置信息而在路径中的最小带宽作为关于路径的带宽信息。

[0115] 存储拓扑管理阈值表 140 包括用于最小带宽的阈值列 1401、用于中转连接装置数目的阈值列 1402、用于最低带宽路径的平均带宽使用速率的阈值列 1403、用于最低带宽路径的峰值带宽使用速率的阈值列 1404 和用于延时的阈值列 1405。操作管理程序 124 在计算卷的目的地端口的推荐级别时参考存储拓扑管理阈值表 140。稍后将描述计算推荐级别的方法的细节。在存储拓扑管理阈值表 140 中的值由管理员预设。

[0116] 用于最小带宽的阈值是用于在两个存储系统 2 的端口之间的路径(包括一个或者多个连接装置 40)中的最小带宽的阈值。用于中转连接装置数目的阈值是用于在两个存储系统 2 的端口之间的路径中的中转连接装置数目的阈值。最低带宽路径是如下路径,该路

径的带宽在两个存储系统 2 的端口之间的路径中包括的路径中最低。

[0117] 在两个存储系统端口之间的路径包括连接装置的端口和在存储系统端口与连接装置端口之间的路径。它可以包括在两个连接装置端口之间的路径。例如在存储系统端口之间的路径贯穿多个连接装置 40, 并且在某个连接装置对的端口之间的路径的带宽最低的情况下, 在连接装置之间的路径是最低带宽路径。根据在用于终止路径的端口段的带宽最小值确定路径的最小带宽。

[0118] 平均带宽使用率是路径的带宽使用率在预定时间段、例如一小时中的平均值。具体而言, 它是通过将在预定时间段中使用的带宽(例如以 bps 为单位表达)的平均值除以用于路径的指定的带宽而获得的值。峰值带宽使用率是在预定时间段中的最高带宽使用率。例如在以 bps 为单位表达用于路径的指定的带宽的情况下, 它是通过将在预定时间段中的最高每秒比特值除以用于路径的指定的带宽而获得的值。

[0119] 图 12 示出在管理计算机 1 中的卷管理表 150 的配置。卷管理表 150 用指派的唯一编号管理在计算机系统中存在的卷。卷管理表 150 包括卷编号列 1501; 向每列指派卷编号。与卷编号关联地管理关于其它项的信息。

[0120] 容量列 1502 存储个体卷的卷容量, 使用率列 1503 存储卷的使用数量。卷类型列 1504 指示每个卷是在正常卷、精简配备卷和池卷(向精简配备池提供实际容量的卷)中的哪个卷。

[0121] 卷管理表 150 还具有指示保持卷的存储系统的存储系统 ID 列 1505、指示用于唯一标识(仅)在特定存储系统中的卷的标识信息的卷 ID 列 1506 和指示向卷指派的端口的端口 ID 列 1507。端口 ID 是向每个存储系统 2 指派的值。

[0122] 例如管理员可以在创建卷时使用操作管理程序 124 来创建卷管理表 150 的条目。存储系统管理程序 121 从存储系统 2 获得将在卷管理表 150 中存储的信息并且存储它。

[0123] 图 13 是图示在管理计算机 1 中的存储系统使用管理表 160 的配置示例的图。表 160 管理存储系统 2 的容量使用。存储系统使用管理表 160 具有存储系统 ID 列 1601。

[0124] 存储系统使用管理表 160 还具有总容量列 1602、容量阈值列 1603 和存储系统 ID 标识的存储系统 2 的使用数量列 1604。从管理员或者存储系统 2 获得总容量的值, 并且容量阈值由管理员确定。为了获得使用量, 在管理计算机 1 中的存储系统管理程序 121 经由连接装置 6 向每个存储系统 2 周期性地查询并且在表 160 中存储获得的值。

[0125] 图 14 是图示在管理计算机 1 中的精简配备池使用管理表 170 的示例。精简配备池使用管理表 170 管理在计算机系统中的精简配备池 223 的使用。图 14 举例说明关于仅一个精简配备池 TPoo11 的信息。精简配备池使用管理表 170 具有在计算机系统中公共确定的池 ID 列 1701、一个或者多个存储系统提供的总容量列 1702 和那些存储系统的使用量列 1703。

[0126] 另外, 精简配备池使用管理表 170 在列 1704 中存储构成池的个体存储系统 2 的 ID, 并且具有用于个体存储系统的容量列 1705、使用量的列 1706 和 IOPS (每秒输入输出) 列 1708。

[0127] 另外, 精简配备池使用管理表 170 具有将在存储系统 2 中的使用量和 IOPS 超过容量阈值和性能阈值时的协调中使用的容量阈值列 1707 和性能阈值列 1709。例如在管理员产生用于创建精简配备池 223 的指令时添加而在管理员产生用于删除精简配备池 223 的指

令时删除精简配备池使用管理表 170 的条目。

[0128] 管理员可以使用操作管理程序 124 来设置容量阈值和性能阈值。为了获得用于使用量列 1706 和 IOPS 列 1708 的值,在管理计算机 1 中的存储系统管理程序 121 经由连接装置 6 向存储系统 2 周期性地发出查询并且向那些字段存储从存储系统 2 获得的值。

[0129] 图 15 示出在管理计算机 1 中的目的地推荐级别表 180 的示例。目的地推荐级别表 180 是指示其它端口的推荐级别的表,这些端口将是用于将要连接到端口的卷的目的地端口。图 15 的目的地推荐级别表 180 是如下表的示例,该表指示用于在存储系统 ST1 中的端口 PT1 的目的地端口的推荐级别。推荐级别是用于指示目的地端口的优先级级别的值。

[0130] 在管理计算机 1 中的操作管理程序 124 可以根据从存储系统 2 获得的路径信息计算每个端口的目的地推荐级别。通常,操作管理程序 124 在确定卷的目的地时创建用于在存储系统 2 中的每个端口的目的地推荐级别表 180。稍后将描述推荐级别的计算。

[0131] 在传送卷时,向管理员呈现在目的地推荐级别表 180 中的推荐级别。管理员参考呈现的推荐级别以确定卷的目的地。备选地,管理计算机 1 可以参考计算的推荐级别(优先级)自动确定卷的目的地。

[0132] 图 15 的目的地推荐级别表 180 具有单存储池 ID 列 1801、存储系统 ID 列 1802、目的地端口 ID 列 1803 和目的地推荐级别列 1804。在这一示例中,在目的地推荐级别列 1804 中的星号数目越多,推荐级别就越高。

[0133] 在图 15 的示例中,可以向在单存储池 SPoo11 中的除了在存储系统 ST1 中的端口 PT1 之外的其它端口传送连接到特定端口 PT1 的卷。如图 15 中所示,目的地端口包括在相同存储系统 ST1 中的其它端口。从在一个存储系统 2 中的不同端口到主机计算机 3 的路径贯穿相同连接装置或者不同连接装置。如果在两个路径中的连接装置 40 的端口不同,则它们的带宽性质可以不同。

[0134] 图 16 图示在每个存储系统 2 中的存储系统管理信息存储部 221 中存储的信息。数据 I/O 程序 2211 经由从数据 I/O 通信 I/F25 连接的连接装置组 4 与在计算机系统的主机计算机 3 通信、发送和接收用于卷 2221 和精简配备卷 224 的数据并且向它们写入。

[0135] 池管理程序 2212 管理在精简配备池 223 中的页面 2231、作为池卷的卷 2221 和页面 2231 向精简配备卷 224 的分配。卷管理程序 2213 管理卷 2221 和精简配备卷 224。

[0136] 操作管理程序 2214 与各种程序配合地管理存储系统 2 的操作。操作系统 2215 是用于执行各种程序和用于管理各种表的平台。存储拓扑管理程序 2216 管理关于网络的拓扑信息,该网络连接着程序 2216 在其上运行的存储系统 2 和其它存储系统 2 (在存储系统端口之间的路径)。

[0137] 这一网络是连接主机计算机 3 和存储系统 2 的网络并且是用连接装置组 4 配置的数据网络。存储系统拓扑管理信息包括关于在存储系统 2 的端口之间的路径的配置信息;具体而言,它包括在路径中的连接装置配置和带宽。

[0138] 存储系统管理信息存储部 221 还保持存储拓扑管理表 200 和管理通信 I/F 地址管理表 210。管理通信 I/F 地址管理表 210 保持管理通信 I/F26 的 IP 地址。管理员在向计算机系统安装存储系统 2 时确定这一值。

[0139] 存储拓扑管理表 200 包括关于在特定存储系统 2 和其它存储系统 2 之中的网络的拓扑信息。这一网络包括连接在相同存储系统 2 中的不同端口的路径。如以上描述的那样,

存储拓扑管理表 200 存储关于在特定存储系统 2 的每个端口与其它端口(包括相同存储系统 2 的其它端口)之间的路径的配置信息。

[0140] 图 17 图示在存储系统 2 中的存储拓扑管理表 200 的示例。每个存储系统 2 具有存储拓扑管理表 200。图 17 举例说明在具有 ST1 的存储 ID 的存储系统中的存储拓扑管理表 200。

[0141] 存储拓扑管理表 200 具有端口 ID 列 2001、相反存储系统 ID 列 2002、相反端口 ID 列 2003、最小带宽列 2004、中转连接装置数目列 2005、最低带宽路径的平均带宽使用率列 2006、最低带宽路径的峰值带宽使用率列 2007 和延时列 2008。

[0142] 端口 ID 列 2001 存储在拥有表 200 的存储系统 2 中的端口的 ID。相反存储系统 ID 列 2002 存储在端口 ID 列 2001 中的端口 ID 标识的端口连接到的相反存储系统 2 的 ID。在这一示例中,可以如以上描述的那样在单存储池内传送卷;因而在表 200 中注册与存储系统 ST1 一起构成单存储池的存储系统 ST2 和 ST3。

[0143] 相反端口 ID 列 2003 存储在用端口 ID 标识的端口连接到的相反存储系统 2 中的端口的 ID。最小带宽列 2004 存储在每个条目的存储系统端口之间的路径中包括的路径的带宽的最小值(在存储系统端口之间的路径中包括的每个路径是在存储系统之间的路径的一部分)。中转连接装置数目列 2005 存储在个体条目的存储系统端口之间的路径中的连接装置数目。

[0144] 最低带宽路径的平均带宽使用率和最低带宽路径的峰值带宽使用率列与在关于图 11 的存储拓扑管理阈值表 140 的说明中描述的最低带宽路径的平均带宽使用率和最低带宽路径的峰值带宽使用率列相同。延时列 2008 存储在个体条目中的存储系统端口之间的延时的值。在延时列 2008 中,“>1ms”意味着少于 1ms 的值。

[0145] 在图 17 中所示存储拓扑管理表 200 的示例中,在路径中的连接装置数目代表关于每个路径(条目)的配置信息,并且路径的最小带宽的值代表关于路径的带宽信息。在最低带宽路径上的平均带宽使用率和峰值带宽使用率代表关于带宽的使用的信息,并且包括延时作为指示使用的信息。

[0146] 例如在向计算机系统添加存储系统 2 时,在存储系统 2 中的存储拓扑管理表 200 由存储拓扑管理程序 2216 创建。例如管理员分别向端口 ID 列 2001、相反存储系统 ID 列 2002 和相反端口 ID 列 2003 设置端口 ID、在单存储池中包括的相反存储系统的 ID 和它们的端口的 ID。

[0147] 存储拓扑管理程序 2216 从在连接装置组 4 中的连接装置 40 获得将在最小带宽列 2004、中转连接装置数目列 2005、最低带宽路径的平均带宽使用率 2006 和最低带宽路径的峰值带宽使用率列 2007 中存储的信息。稍后将描述在连接装置组 4 中的连接装置 40 保持的关于网络的配置信息的细节。

[0148] 存储拓扑管理程序 2216 用预定定时更新使用率的值。在添加新存储系统 2 时或者在存储系统的端口连接到从网络装置 / 从网络装置断开存储系统的端口时,在其它存储系统 2 中的存储拓扑管理程序 2216 向 / 从它们的存储拓扑管理表 200 添加 / 删除关于存储系统 2 的信息。

[0149] 存储拓扑管理程序 2216 可以在网络中例如通过 ping 命令(比如 fcping 命令)获得将在延时列表 2208 中存储的值。存储拓扑管理程序 2216 用预定定时更新延时列 2008

的值。

[0150] 图 18 图示在每个主机计算机 3 中的非易失性存储装置 32 中存储的信息。应用 321 是为了主机计算机 3 提供功能或者服务而需要的程序。代理 322 检查主机计算机 3 的状态并且通过与在管理计算机 1 中的主机计算机管理程序 123 的通信来指定关于主机计算机 3 的配置信息。操作系统 323 是用于执行在应用 321 中的各种程序和用于管理各种表的平台。

[0151] 非易失性存储装置 32 还存储连接的存储系统的卷管理表 300 和管理通信 I/F 地址管理表 310。管理通信 I/F 地址管理表 310 存储管理通信 I/F36 的 IP 地址。管理员在计算机系统中安装主机计算机 3 时设置值。

[0152] 图 19 图示在主机计算机 3 中的连接的存储系统的卷管理表 300 的示例。主机计算机 3 具有这一个表 300, 该表存储关于向拥有这一个表的主机计算机 3 分配并且从主机计算机 3 可读取或者向主机计算机 3 可写入的卷的信息。这一个表 300 具有存储系统 ID 列 3001、卷 ID 列 3002 和端口 ID 列 3003。

[0153] 存储系统 ID 是保持这样的卷的存储系统的 ID。卷 ID 是在条目的卷的个体存储系统 2 中的 ID。端口 ID 是在访问目的地的个体存储系统 2 中的端口 ID。这些值可以由管理员确定; 否则, 代理 322 可以存储从管理计算机 1 接收的值。在主机计算机 3 中的操作系统 323 参考端口 ID 访问存储系统 2。

[0154] 图 20 图示在连接装置组 4 中的每个连接装置 40 中的非易失性存储装置 42 中存储的信息。数据 I/O 程序 421 从在从数据 I/O 通信 I/F45 连接的计算机系统 4 中的计算机和装置接收 I/O 请求并且参考在非易失性存储装置 42 中的表与它们通信。管理程序 422 提供用于设置在非易失性存储装置 42 中的表中的值的界面。操作系统 423 是用于执行各种程序和用于管理各种表的平台。

[0155] 非易失性存储装置 42 还存储连接装置管理表 400 和管理通信 I/F 地址管理表 410。管理通信 I/F 地址管理表 410 存储管理通信 I/F46 的地址。在计算机系统中安装连接装置时, 管理员向这一个表设置值。连接装置 5 和 6 具有相同程序和表。

[0156] 图 21 图示连接装置管理表 400 的配置示例。连接装置管理表 400 管理关于连接装置的每个端口的信息。具体而言, 它存储关于相反装置和每个端口连接到的相反端口以及每个带宽的带宽和使用的信息。在这一示例中, 连接装置管理表 400 具有端口 ID 列 4001、相反装置 ID 列 4002、连接的端口 ID 列 4003、物理带宽列 4004、流量平均数量列 4005 和流量峰值数量列 4006。

[0157] 端口 ID 列 4001 存储连接装置的个体端口的 ID。相反装置 ID 列 4002 存储从个体端口连接的相反的装置的 ID。从端口连接的相反装置是存储系统 2、主机计算机 3 或者不同连接装置。在图 21 的示例中, 存储系统 2 和连接装置的 ID 存储于字段中。

[0158] 连接的端口 ID 列 4003 存储在相反装置中连接的端口的 ID。物理带宽列 4004 存储向个体端口预设的带宽的值。流量平均数量是在预定时间段、比如一小时中的流量平均数量并且在这一示例中以 bps 为单位来表达。流量峰值数量是在前述预定时间段中并且在这一示例中以 bps 为单位表达的流量最大数量。

[0159] 在新装置连接到连接装置的端口时, 在连接装置中的管理程序 422 获得相反装置和相反端口的 ID 的值并且在连接端口的在连接装置管理表 400 中的条目中存储获得的 ID。

向个体端口预定物理带宽的值。数据 I/O 程序 421 测量在每个端口的流量数量并且用预定间隔更新在连接装置管理表 400 中的流量平均数量列 4005 和流量峰值数量列 4006 的值。

[0160] 如先前描述的那样,在存储系统 2 中的存储拓扑管理程序 2216 向在连接装置组 4 中的每个连接装置查询用于存储拓扑管理表 200 的信息。已经从存储系统 2 接收查询的管理程序 422 从连接装置管理表 400 获得需要的信息以向存储系统 2 发送它。

[0161] 例如在存储系统 2 新连接到连接装置时,存储拓扑管理程序 2216 请求在连接装置组 4 中的每个连接装置 40 发送指示与网络的连接条件的配置信息。响应于请求,每个连接装置 40 向查询源的存储系统 2 发送在连接装置管理表 400 中保持的所有信息、即在列 4001 至 4006 中保持的值。

[0162] 存储拓扑管理程序 2216 可以从在连接装置组 4 中的连接装置 40 获得的信息获得将在存储拓扑管理表 200 的列中存储的值。例如存储拓扑管理程序 2216 可以根据在连接装置 40 的端口与来自连接装置的连接的端口之间的关系断定在存储系统 2 的端口与其相反端口和中转端口之间的中转连接装置。

[0163] 存储拓扑管理程序 2216 可以从存储系统端口之间的路径的连接装置配置获得在路径中的中转连接装置数目的值。另外,在断定在路径中的连接装置 40 的端口时,存储拓扑管理程序 216 可以从在路径中的端口的物理带宽的值获得路径的最小带宽的值。

[0164] 存储拓扑管理程序 2216 可以确定在用于在路径中的端口的物理带宽的值中的最小值为路径的最小带宽的值。在连结多个端口以形成一个物理端口的情况下,在确定路径的最小带宽时参考由多个端口构成的合并端口的物理带宽。

[0165] 存储拓扑管理程序 2216 可以根据在连接装置管理表 400 中的流量平均数量和流量峰值数量计算在两个存储端口之间的最低宽度路径的平均带宽使用率和峰值带宽使用率。为了仅更新这些值,存储拓扑管理程序 2216 请求在两个存储端口之间的每个路径中的最低带宽路径中的去往连接装置的流量平均数量和流量峰值数量的值并且使用获得的值来更新平均带宽使用率和峰值带宽使用率。

[0166] 现在参照图 22 的流程图,将描述用于选择将传送的卷和卷的目的地用于在计算机系统资源平衡的过程的示例。在这一示例中,管理计算机 1 基于存储系统 2 的使用来选择将传送的候选卷并且还基于在两个存储系统端口之间的路径计算卷的目的地端口的推荐级别。管理员可以参考这一推荐级别判决卷的目的地。在这一示例中,用于被计算推荐级别的卷的候选目的地端口是与当前存储系统 2 不同的存储系统 2 的端口。

[0167] 在管理计算机 1 中的操作管理程序 124 搜寻用于迁移的卷。具体而言,它参考存储系统使用管理表 160 并且关于在存储系统 ID 列 1601 中的每个字段搜寻如下条目,该条目的在使用数量列 1604 中的字段的值与在总容量列 1602 中的字段的值的比率高于在容量阈值列 1603 中的字段的值(S1001)。

[0168] 如果存在其中使用数量的比率高于容量阈值的任何存储系统 2,则操作管理程序 124 在卷管理表 150 中的存储系统 ID 列 1505 中搜寻存储系统 ID 的条目。操作管理程序 124 从在匹配条目的卷中的在卷类型列 1504 中的值为正常的卷选择卷作为迁移卷(S102)。存储系统管理程序 121 选择例如使用数量最大的卷。

[0169] 接着,操作管理程序 124 从用于精简配备的卷选择用于迁移的卷。操作管理程序 124 参考精简配备池使用管理表 170 并且搜寻如下条目的存储系统 ID,该条目的在使用数

量列 1706 中的字段的值与在容量列 1705 中的字段的值的比率高容量阈值列 1707 的字段的值。此外,操作管理程序 124 搜寻如下条目的存储系统 ID,该条目的 IOPS 列 1708 的字段的值高于性能阈值列 1709 的字段的值(S103)。

[0170] 接着,操作管理程序 124 参考卷管理表 150 并且在存储系统 ID 列 1505 中搜索在步骤 S103 确定使用数量或者 IOPS 比它们的相应阈值更高的存储系统的条目。在存储系统中的卷(条目)之中,它从在卷类型列 1504 中的值为精简配备的条目选择卷作为迁移卷(S104)。存储系统管理程序 121 选择例如使用数量最大的卷。应当注意在这一流程的系列处理中执行从适于精简配备的卷和其它正常卷的选择;然而可以选择执行步骤 S101 和 S102 或者步骤 S103 和 S104。

[0171] 接着,操作管理程序 124 参考选择的卷的在卷管理表 150 中的条目的端口 ID 列 1507 以断定卷的端口 ID (S105)。

[0172] 另外,操作管理程序 124 根据选择的卷的端口 ID 确定卷可以被传送到的端口(S106)。在这一示例中,操作管理程序 124 选择如下端口作为选择的卷可以被传送到的端口,这些端口在选择的卷的端口属于的单存储池中以及在具有足以存储选择的卷的剩余容量的其它存储系统 2 中的端口。

[0173] 操作管理程序 124 可以参考传送范围管理表 130 断定将传送的卷的端口属于的单存储池以及其它存储系统和在其中包括的端口。

[0174] 例如操作管理程序 124 参考存储系统使用管理表 160 确定用于正常类型的卷的候选目的地存储系统 2 (端口) 并且参考精简配备池使用管理表 170 确定用于精简配备卷的候选目的地存储系统 2 (端口)。

[0175] 操作管理程序 124 可以参考存储系统使用管理表 160 断定存储系统的容量和使用数量并且可以参考精简配备池使用管理表 170 断定在精简配备池中的存储系统的容量和使用数量。精简配备卷的目的地可以限于向相同精简配备卷提供存储区域的存储系统或者是其它精简配备池。

[0176] 接着,操作管理程序 124 计算确定的候选目的地端口的推荐级别(S107)。稍后将参照图 23 描述推荐级别的计算方法的细节。如果已经选择多个卷,则操作管理程序 124 计算每个卷的推荐级别。存储系统管理程序 121 在目的地推荐级别表 180 中存储计算的推荐级别。另外,操作管理程序 124 向映像输出装置 18 输出包括从目的地推荐级别表 180 获得的推荐级别的信息(S108)。

[0177] 例如,映像输出装置 18 显示具有与目的地推荐级别表 180 相同的内容的表。在精简配备卷的情况下,操作管理程序 124 也可以显示关于当前精简配备池和候选目的地精简配备池的信息。

[0178] 管理员参考显示的候选目的地端口和它们的推荐级别用输入装置 19 选择卷被传送到的端口(S1309)。操作管理程序 124 确定管理员选择的端口为用于卷的目的地端口并且指令源存储系统和 / 或目的地存储系统迁移卷。在图 22 的示例中,存储系统 2A 是用于卷迁移的源,并且存储系统 2B 是用于卷迁移的目的地。

[0179] 存储系统 2A 和 2B 执行将选择的卷从包括选择的卷的存储系统 2A 向选择的迁移目的地的存储系统 2B 迁移(S110)。存储系统管理程序 121 指令存储系统 2B 传送卷。在存储系统 2B 中的卷管理程序 2213 请求向存储系统 2A 传送指定的卷(中的数据)。

[0180] 在存储系统 2A 和 2B 中的卷管理程序 2213 向在存储系统 2B 中的卷传送在存储系统 2A 中的目标卷中的数据。另外,在存储系统 2A 和 2B 中的卷管理程序 2213 更新它们的关于卷的相应管理信息(未示出)。在管理计算机 1 中的操作管理程序 124 接收到指定的卷的迁移完成通知时,它更新卷管理表 150 的相关条目。

[0181] 前述流程实现考虑卷经由网络的访问中的性能在传送卷时选择更适合的目的地端口。由于基于在存储系统端口之间的路径计算候选目的地的推荐级别,所以即使在未管理关于主机计算机 3 的连接信息的系统中仍然可以获得推荐级别。

[0182] 在前述示例中,管理系统自动选择将传送的卷,但是管理员可以指定将传送的卷。被计算推荐级别的候选卷目的地端口可以包括当前存储系统 2 的端口,该存储系统包括该卷。管理计算机 1 关于指定的卷执行前述流程。管理系统可以根据推荐级别确定目的地端口而未向管理员呈现候选目的地端口的推荐级别。

[0183] 接着参照图 23 的流程图,将描述计算候选目的地端口的推荐级别的方法。如参照图 22 说明的那样,操作管理程序 124 计算被选择作为候选目的地的端口的推荐级别(S107)。存储系统管理程序 121 基于在源端口与目的地端口之间的路径的配置(连接装置配置和带宽)和使用来计算推荐级别。操作管理程序 124 对于每个候选目的地端口执行图 23 的流程。

[0184] 在图 23 的流程图中,操作管理程序 124 设置推荐级别项计数为项数(S201)。在这一示例中,项数为 5。接着,操作管理程序 124 选择将被计算推荐级别的候选目的地端口、从卷传送源的存储系统 2 获得关于在存储拓扑管理表 200 中的用于卷的源端口和选择的候选目的地端口的条目的信息并且在高速缓冲存储器 11 或者非易失性存储装置中存储它(S202)。

[0185] 操作管理程序 124 比较从存储拓扑管理表 200 获得的值与在存储拓扑管理阈值表 140 中保持的阈值并且计算推荐级别。具体而言,操作管理程序 124 比较路径的最小带宽的值与用于最小带宽的阈值(S203)。如果最小带宽的值等于或者低于阈值(在 S203 为否),则操作管理程序 124 将计数递减一(S204)。如果最小带宽的值高于阈值(在 S203 为是),则维持计数。

[0186] 接着,存储系统管理程序 121 比较在路径中的中转连接装置数目的值与用于中转连接装置数目的阈值(S205)。如果中转连接装置数目的值等于或者高于阈值(在 S205 为否),则操作管理程序 124 将计数递减一(S206)。如果中转连接装置数目的低于阈值(在 S205 为是),则维持计数。

[0187] 接着,操作管理程序 124 比较在路径中的最低带宽路径的平均使用率的值与用于平均带宽使用率的阈值(S207)。如果平均带宽使用率的值等于或者高于阈值(在 S207 为否),则存储系统管理程序 121 将计数递减一(S208)。如果平均带宽使用率的值低于阈值(在 S207 为是),则维持计数。

[0188] 接着,操作管理程序 124 比较在路径中的最低带宽路径的峰值带宽使用率的值与用于峰值带宽使用率的阈值(S209)。如果峰值带宽使用率的值等于或者高于阈值(在 S209 为否),则操作管理程序 124 将计数递减一(S210)。如果峰值带宽使用率的值低于阈值(在 S209 为是),则维持计数。

[0189] 接着,操作管理程序 124 比较在路径中的延时的值与用于延时的阈值(S211)。如

果延时的值等于或者高于阈值(在 S211 为否),则操作管理程序 124 将计数递减一(S212)。如果延时的值低于阈值(在 S211 为是),则维持计数。

[0190] 最后,操作管理程序 124 将计数的值替换成预定公式以获得推荐级别(S213)。在示例中,推荐级别的计算将推荐级别的最高值(在这一示例中为五)与计数的乘积除以项数并且将结果取整为整数。操作管理程序 124 在传送推荐级别表 180 中存储这一个值作为这一路径的推荐级别的值。

[0191] 这一示例基于在存储系统端口之间的路径的配置和使用来计算目的地端口的推荐级别。在主机存储装置之中的网络中,在网络中选择与当前卷位置(端口)接近的新位置(端口)实现选择其中从主机计算机 3 的访问性能被更少改变的卷位置。

[0192] 计算推荐级别的前述方法仅为示例,并且操作管理程序 124 可以使用不同方法考虑以上描述的项来计算推荐级别。操作管理程序 124 可以参考以上描述的五项的仅部分计算推荐级别。例如操作管理程序 124 可以仅根据最小带宽这一项或者根据最小带宽和中转连接装置数目计算推荐级别。

[0193] 各项可以被分离地加权,并且可以根据项不同地递减计数。例如可以向最小带宽和中转连接装置数目给予最大加权因子,可以向平均带宽使用率和峰值带宽使用率给予第二最大加权因子,并且可以向延时给予最小加权因子。操作管理程序 124 可以参考与以上五项不同的其它项。在前述示例中,每项具有一个阈值,但是操作管理程序 124 可以将多个阈值用于每项。例如操作管理程序 124 可以根据卷、路径或者存储系统使用不同阈值。

[0194] 在以上描述的配置示例中,管理计算机 1 具有存储拓扑阈值表 210。每个存储系统 2 可以具有用于它自己的存储系统的存储拓扑阈值表 210 并且响应于来自管理计算机 1 的请求向管理计算机 1 发送阈值。

[0195] 以上描述的示例具有五个推荐级别;然而级别数目可以是不少于二的任何值。例如在定义两个推荐级别的情况下,如果五项中的任一项为负(在用于递减计数的状态中),则推荐级别可以是低;否则,如果所有项为正,则推荐级别可以是高的。

[0196] 管理计算机 1 可以从每个存储系统 2 获得在存储拓扑管理表 200 中保持的信息并且创建和保持管理计算机本身将引用的用于以上描述的处理和其它处理的存储拓扑管理表。这样的存储拓扑管理表是通过合并所有存储系统 2 的存储拓扑管理表 200 而获得的表。管理计算机 1 响应于预定事件从每个存储系统 2 获得更新信息以更新它自己的拓扑管理表。这同样适用于稍后描述的主机拓扑管理表。

[0197] 接着参照图 24,将说明更新(创建)存储拓扑管理表 200 的示例。图 24 的序列图图示响应于在计算机系统中安装新存储系统来更新(创建)存储拓扑管理表 200 的示例。

[0198] 在图 24 中,新存储系统 2C 的端口首先连接到连接装置 40 的端口。存储系统 2C 已经新连接到的连接装置 40 的管理程序 422 利用存储系统 2C 的操作管理程序 2214 的协商来获得存储系统 2C 和连接端口的 ID 并且向连接装置管理表 400 添加它们(S401)。在存储系统 2C 的多个端口连接到分离连接装置 40 的情况下,个体连接装置 40 在相应连接的装置管理表 400 中注册信息。

[0199] 在完成在连接装置管理表 400 中注册时,在连接装置 40 中的管理程序 422 向存储系统 2C 通知完成。存储系统 2C 的操作管理程序 2214 向在连接装置组 4 中的每个连接装置 40 请求拓扑信息。拓扑信息是在连接装置管理表 400 中包括的信息。操作管理程序 2214

从在连接装置组 4 中的每个连接装置 40 获得在连接装置管理表 400 中包括的所有信息。

[0200] 在连接的存储系统 2C 中的操作管理程序 2214 用获得的信息更新存储拓扑管理表 200 (S402)。在这一示例中,更新意味着存储新数据。操作管理程序 2214 从在连接装置组 4 中的连接装置 40 获得的信息获得在存储拓扑管理表 200 中的列 2004 至 2007 中的值。操作管理程序 2214 还测量在路径中的延时并且在延时列 2008 中存储它们。

[0201] 虽然在图 24 中未示出,但是其它存储系统 2 周期性地或者响应于来自管理计算机 1 的指令或者物理线路连接配置改变从存储系统 2C 的端口连接的连接装置 40 获得关于存储系统 2C 的信息、即关于存储系统 2C 的连接端口的信息并且更新它们的存储拓扑管理表 200。

[0202] 例如在接收到对于卷迁移的请求时,在管理计算机 1 中的操作管理程序 124 向特定卷的存储系统、在这一示例中为存储系统 2C 请求关于在特定卷的存储系统端口与其它存储系统端口之间的路径的拓扑信息以计算推荐级别。拓扑信息是在存储拓扑管理表 200 中保持的关于路径的信息。

[0203] 响应于来自管理计算机 1 的请求,在存储系统 2C 中的操作管理程序 2214 向每个连接装置 40 请求用于拓扑信息的更新信息(在连接装置管理表 400 中的信息)。更新信息例如是关于装置或者与端口相反的端口的更新或者更新的测量的流量数量的信息。

[0204] 操作管理程序 2214 用从连接装置 40 获得的更新信息更新它自己的存储拓扑管理表 200(S403)并且向管理计算机 1 发送从存储拓扑管理表 200 获得的拓扑信息。在管理计算机 1 中的操作管理程序 124 使用获得的路径信息来计算路径(目的地端口)的推荐级别。

[0205] 前述示例在计算推荐级别时参考最小带宽。在这一实施例中计算推荐级别的方法可以参考关于路径的带宽的不同信息。例如管理计算机 1 参考路径的平均带宽以确定推荐级别。

[0206] 如以上描述的那样,管理计算机 1 可以保持和管理用于管理存储拓扑的表。存储系统管理程序 121 从每个存储系统 2 接收用于存储拓扑管理表 200 的所有信息或者更新的信息。存储系统管理程序 121 使用接收的信息来创建或者更新存储拓扑管理表。

[0207] 在前述示例中,管理计算机 1 经由存储系统 2 收集关于存储拓扑的信息。管理计算机 1 可以从连接装置 40 直接收集关于存储拓扑的信息。例如管理计算机 1 可以从存储系统 2 获得在存储拓扑管理表 200 中的关于延时的信息并且从连接装置 40 获得其它信息。关于计算推荐级别的其它方法的说明适用于下文将描述的其它实施例。

[0208] 前述示例计算在实况卷迁移中的候选目的地端口的推荐级别。被计算推荐级别的候选卷位置端口不限于用于现有卷的候选目的地端口并且可以是用于对将新创建的卷定位的候选端口。

[0209] 第二实施例

[0210] 下文将说明本发明的第二实施例。第二实施例基于将传送的卷的 I/O 数量和路径的使用来计算目的地端口的推荐级别。这一实施例实现根据将选择的卷的访问条件选择适当目的地端口。在以下描述中,将主要说明在第二实施例中的与第一实施例不同的点。

[0211] 图 25 图示在第二实施例中的存储系统管理信息存储部 221 中包括的信息。作为与第一实施例不同的一点,添加卷访问条件管理表 220。卷访问条件管理表 220 是用于管理向每个卷的访问条件的表。

[0212] 图 26 图示卷访问条件管理表 220 的配置示例。卷访问条件管理表 220 具有端口 ID 列 2201、卷 ID 列 2202、在卷访问中使用的平均带宽列 2203 和在卷访问中使用的峰值带宽列 2205。端口 ID 和卷 ID 与在第一实施例中的说明的端口 ID 和卷 ID 相同；卷 ID 列 2202 指示从用端口 ID 标识的端口连接的卷。

[0213] 在卷访问中使用的平均带宽和峰值带宽是代表卷的 I/O 数量的值。在卷访问中使用的平均带宽是在预定时间(例如一小时)内访问卷时使用的带宽的平均值。在卷访问中使用的峰值带宽是在以上提到的预定时间内访问卷时使用的带宽的最大值。在这一示例中，以 bps 为单位表达这些。在每个存储系统 2 中的数据 I/O 程序 2211 测量关于每个卷的这些值并且用确定的时间间隔更新在卷访问条件管理表 220 中的值。

[0214] 参照图 27A 和 27B 的流程图，将说明在第二实施例中的确定推荐级别的方法。图 27A 和 27B 是流程图的部分，并且它们由连接符 A 连结。在以下描述中，将主要说明与在第一实施例中的图 23 的流程图不同的点。首先，管理计算机 1 的操作管理程序 124 设置推荐级别项计数为项数(S501)。在这一示例中，项数为七。

[0215] 接着，操作管理程序 124 选择将被计算推荐级别的候选目的地端口并且从卷传送源的存储系统 2 获得关于在存储拓扑管理表 200 中的用于将传送的卷的源端口和选择的候选目的地端口的条目的信息。另外，操作管理程序 124 从卷属于的存储系统 2 获得关于将传送的卷的访问条件的信息。具体而言，操作管理程序 124 获得在卷访问条件管理表 220 中保持的在卷访问中使用的平均带宽和峰值带宽的值(S502)。

[0216] 操作管理程序 124 计算在路径中的平均带宽使用率的值与最小带宽的值的乘积并且计算获得的乘积与在卷访问中使用的平均带宽的值的求和。操作管理程序 124 比较获得的求和与用于路径的用于最小带宽的阈值(S503)。用于最小带宽的阈值是最小带宽的值与平均带宽使用率的乘积。阈值表 140 可以存储与用于最小带宽的这一阈值不同的用于最小带宽的阈值。

[0217] 如果获得的求和高于用于最小带宽的阈值(在 S503 为是)，则操作管理程序 124 将计数递减一(S504)。如果获得的求和等于或者低于用于最小带宽的阈值(在 S503 为否)，则维持计数。

[0218] 操作管理程序 124 计算在路径中的峰值带宽使用率与最小带宽的值的乘积并且计算获得的乘积与在卷访问中使用的卷的峰值带宽的值的求和。操作管理程序 124 比较获得的求和与用于峰值带宽的阈值(S505)。用于峰值带宽的阈值是最小带宽的值与用于峰值带宽使用率的阈值的乘积。阈值表 140 可以存储与用于峰值带宽的这一阈值不同的用于峰值带宽的阈值的信息。

[0219] 如果获得的求和高于用于峰值带宽的阈值(在 S505 为是)，则操作管理程序 124 将计数递减一(S506)。如果获得的求和等于或者低于用于峰值带宽的阈值(在 S505 为否)，则维持计数。

[0220] 在图 27A 和 27B 的流程图中的步骤 S507 至 S516 与在图 23 的流程图中的步骤 S203 至 S212 相同；因而省略说明。在图 27B 中的步骤 S517，操作管理程序 124 使用预定公式来计算推荐级别。公式与在第一实施例中的公式(图 23 中的步骤 S213)相同。然而在第二实施例中，项数为七；用于推荐级别的最大值为七。

[0221] 这一实施例考虑用于将传送的卷的附加 I/O 数量对用于路径的当前使用条件(在

前述示例中由在卷访问中使用的平均带宽和峰值带宽指定)引起的影响来计算推荐级别。这一计算实现根据卷的 I/O 特性选择更适合的目的地端口。

[0222] 这一示例向在第一实施例中的项添加代表关于卷的访问条件的两项。如同在第一实施例中,操作管理程序 124 可以通过与这一方法不同的方法计算推荐级别。例如操作管理程序 124 可以使用代表关于卷的访问条件的两项中的任一项;否则,它可以参考其它五项的部分或者无需参考它们中的任一项。

[0223] 第三实施例

[0224] 下文将说明本发明的第三实施例。第三实施例基于在主机计算机 3 和存储系统 2 之中的数据通信网络(在前述示例中为 SAN)的拓扑确定用于对虚拟机定位的端口(主机计算机 3 的端口)的优先级级别。

[0225] 不同于计算在存储系统之间的路径的推荐级别(优先级)的第一实施例,第三实施例基于在主机计算机 3 与存储系统 2 之间的路径的配置和使用来计算虚拟机的目的地的推荐级别(优先级)。第三实施例也基于在主机计算机 3 与存储系统 2 之间的路径的配置(连接装置配置和带宽)和使用来计算用于卷的目的地端口的推荐级别(优先级)。

[0226] 在第三实施例中,一个或者多个虚拟机在主机计算机 3 上运行。主机计算机个体执行虚拟化软件;虚拟机在虚拟化软件上运行。虚拟机是软件(程序)并且除了关于虚拟机的配置信息之外还包括 OS(操作系统)。例如虚拟机是虚拟服务器并且除了 OS 之外还包括用于虚拟服务器的应用程序和用于程序操作的数据。

[0227] 图 28 图示在管理计算机 1 中的非易失性存储装置 12 中保持的信息。与第一实施例不同的点是添加管理表 190、主机拓扑管理阈值表 195 和虚拟机位置推荐级别表 197。主机管理表 190 是用于管理在计算机系统的主机计算机 3 的表。虚拟机位置推荐级别表 197 是用于管理端口的推荐级别以对虚拟机定位的表。如稍后将描述的那样,传送范围管理表 130 还包括定义虚拟机可以被传送到的端口的信息。

[0228] 图 29 是主机管理表 190 的配置示例。主机管理表 190 具有主机 ID 列 1901、管理端口 IP 地址列 1902、安装的存储器数量列 1903、存储器使用数量列 1904、CPU 芯数目列 1905、CPU 频率列 1906、CPU 使用率列 1907 和端口 ID 列 1908。主机管理表 190 包括关于主机计算机 3 的配置、性能和使用的信息。

[0229] 主机 ID 是用于唯一标识在主机系统中的主机计算机 3 的标识;管理计算机 1 可以通过主机计算机 3 的管理端口 IP 地址与主机计算机 3 通信。关于除了存储器使用和 CPU 使用之外的值,管理员在计算机系统中安装主机计算机 3 时注册它们。存储器意味着主机计算机 3 的主存储器,并且端口 ID 列 1908 存储个体主机计算机 3 的端口的标识符。

[0230] 例如代理 322 周期性地获得代理 322 在其上操作的主机计算机 3 的存储器使用和 CPU 使用并且响应于来自管理计算机 1 的请求向它发送那些值。管理计算机 1 的主机计算机管理程序 123 分别在主机管理表 190 中的存储器使用数量列 1904 和 CPU 使用率列 1907 中存储在给定的时段中获得的存储器使用和 CPU 使用的值或者其对于多个时段的平均值。

[0231] 图 30 图示主机拓扑管理阈值表 195 的配置示例。主机拓扑管理阈值表 195 具有用于最小带宽的阈值列 1951、用于中转连接装置数目的阈值列 1952、用于最低带宽路径的平均带宽使用率的阈值列 1953、用于最低带宽路径的峰值带宽使用率的阈值列 1954 和用于延时的阈值列 1955。

[0232] 主机拓扑管理阈值表 195 的配置与存储拓扑管理阈值表 140 的配置相同。主机拓扑管理阈值表 195 与存储拓扑管理阈值表 140 不同在于存储用于在每个主机计算机 3 的端口与每个存储系统 2 的端口之间的路径的阈值。

[0233] 图 31 图示在存储系统管理信息存储部 221 中保持的信息。与第一实施例不同的是添加主机拓扑管理表 230。主机拓扑管理表 230 是用于管理关于在自己的存储系统 2 的端口与主机计算机 3 之间的路径的信息的表。

[0234] 包括存储拓扑管理表 200 和主机拓扑管理表 230 管理的路径的网络是相同网络, 该网络是用于在主机计算机 3 与存储系统 2 之间的数据通信的网络并且通常是 SAN。存储拓扑管理表 200 管理在存储系统 2 的端口之间的路径, 并且主机拓扑管理表 230 管理在主机计算机 3 的端口与存储系统 2 之间的路径。

[0235] 图 32 图示主机拓扑管理表 230 的列的配置示例。主机拓扑管理表 230 具有端口 ID 列 2301、相反主机计算机 ID 列 2302、相反端口 ID 列 2303、最小带宽列 2304、中转连接装置数目列 2305、最低带宽路径的平均带宽使用率列 2306、最低带宽路径的峰值带宽使用率列 2307 和延时列 2308。

[0236] 主机拓扑管理表 230 的配置与存储拓扑管理表 200 的配置基本上相同; 从存储系统 2 的端口连接的相反端口是主机计算机 3 的端口而不是存储系统 2 的端口。在其它列中的值的含义与在存储拓扑管理表 200 中说明的含义相同。

[0237] 用于存储系统 2 获得将在主机拓扑管理表 230 中存储的信息的方法与用于存储拓扑管理表 200 的方法相同。在连接装置组 4 中的每个连接装置 40 在它的连接装置管理表 400 中具有条目, 这些条目在与它自己的端口相反的装置和连接的端口的 ID 中保持主机计算机 ID 和它们的端口 ID。连接装置 40 响应于来自存储系统 2 的请求发送从连接装置管理表 400 获得的所需信息。

[0238] 现在参照图 33 的流程图, 将说明确定虚拟机的位置(主机计算机 3 的端口)的推荐级别的方法。例如进行确定虚拟机的位置以将在操作中的虚拟机从当前主机计算机 3 向不同主机计算机 3 传送或者在主机计算机 3 中安装新虚拟机。

[0239] 例如管理计算机 1 的操作管理程序 124 或者管理员参考在主机管理表 190 中的主机计算机 3 的使用条件确定高负荷的主机计算机 3 并且从在主机计算机 3 上运行的虚拟机选择将向不同主机计算机 3 传送的虚拟机。在确定虚拟机的位置时, 预定将参考的卷的端口。例如它是在操作中的虚拟机访问的卷的端口或者保持虚拟机的 OS 的卷的端口。

[0240] 操作管理程序 124 或者管理员参考传送范围管理表 130 选择可以从卷的端口连接的主机计算机 3 的端口作为用于定位的候选端口。传送范围管理表 130 除了图 10 中所示信息之外还包括定义从存储系统 2 的每个端口可连接的主机计算机 3 的端口的信息。

[0241] 操作管理程序 124 或者管理员可以在选择用于定位的候选端口时参考在主机管理表 190 中的主机计算机 3 的性能和使用条件。选择值落在预定范围中的主机计算机 3 的端口作为候选位置。操作管理程序 124 对于选择的候选位置端口中的每个候选位置端口执行图 33 的流程。

[0242] 首先, 管理计算机 1 的操作管理程序 124 设置推荐级别项计数为项数(S601)。在这一示例中的项数为 5。操作管理程序 124 对于在主机计算机 3 的选择的端口与卷的端口之间的路径执行步骤 S603 至 S613。

[0243] 步骤 S603 至 S613 与在第一实施例中的图 23 的流程图中的步骤 S203 至 S213 相同。虽然第一实施例为在存储系统 2 的端口之间的路径计算卷位置端口的推荐级别,但是在第三实施例中的管理计算机 1 为在主机计算机 3 的端口与存储系统 2 的端口之间的路径计算候选虚拟机位置端口的推荐级别。

[0244] 操作管理程序 124 选择候选虚拟机位置端口并且从卷属于的存储系统 2 获得关于在卷的主机端口与存储系统端口之间的路径的信息(S602)。存储系统 2 向管理计算机 1 发送关于在拓扑管理表 230 中的由存储系统端口和主机端口确定的条目的信息。

[0245] 操作管理程序 124 参考从存储系统 2 获得的路径信息对于用于定位的候选端口(确定的路径)执行步骤 S603 至 S613。操作管理程序 124 在虚拟机位置推荐级别表 197 中存储计算的推荐级别。位置推荐级别表 197 例如包括主机计算机 ID 列、端口 ID 列和位置推荐级别列。

[0246] 操作管理程序 124 从位置推荐级别表 197 获得候选主机计算机和用于定位的端口的标识符以及它们的推荐级别的值,并将它们输出到映像输出装置。管理员可以参考显示的推荐级别的值确定虚拟机位置端口。如同在第一实施例中,管理计算机 1 可以基于计算的优先级别(在这一示例中为推荐级别)自动确定用于对虚拟机定位的端口。

[0247] 接着参照图 34,将描述计算其在主机计算机 3 中的连接端口已经被确定的虚拟机中的卷的连接端口的推荐级别的示例。这一示例未从当前存储系统 2 向不同存储系统 2 传送卷、但是计算在相同存储系统 2 中的多个端口的推荐级别。

[0248] 由于在存储系统 2 中的不同端口用不同路径连接到主机计算机 3 的端口,所以它们的相应路径配置(包括连接装置配置和带宽)可以互不相同。因而,在存储系统 2 中选择适当端口造成选择可以表现向主机计算机 3 的端口的良好传送性能的端口。

[0249] 在图 34 的流程图中,步骤 S701 和 S703 至 S713 与在图 33 的流程图中的步骤 S601、S603 至 S613 相同。如同图 33 的流程,管理计算机 1 的操作管理程序 124 计算在主机计算机 3 与存储系统 2 之间的路径的推荐级别。

[0250] 在步骤 S702,操作管理程序 124 选择存在卷的存储系统 2 的端口之一。操作管理程序 124 依次逐个选择存在卷的存储系统 2 的端口以执行步骤 S703 至 S713。

[0251] 操作管理程序 124 在推荐级别表(未示出)中存储该存储系统的端口及其计算的推荐级别并且向管理员呈现它们。这与在计算推荐级别的其它流程中相同。如同在其它处理中,管理计算机可以自动确定卷的连接端口而未呈现推荐级别。

[0252] 在这一示例中,已经说明计算在单个存储系统 2 中的端口的推荐级别的示例,但是管理计算机 1 可以计算卷连接到的存储系统端口、包括其它存储系统 2 的端口的推荐级别。虽然第一实施例为在存储系统端口之间的路径计算端口的推荐级别,但是第三实施例基于在主机计算机 3 的端口与存储系统 2 的端口之间的路径计算卷的连接端口的推荐级别。

[0253] 如前文阐述的那样,已经参照附图具体描述本发明。然而本发明不限于具体配置、但是在所附权利要求的范围内包括各种等效修改和等效配置。在这一实施例中的配置的部分可以被添加、删除或者替换为在其它配置中的部分。优选的是系统包括所有以上描述的功能,但是系统可以仅包括功能的部分。

[0254] 以上描述的配置和功能对于它们的全部或者部分可以例如由通过设计集成电路

而获得的硬件实施。可以在存储设备、比如非易失性半导体存储器、硬盘驱动或者 SSD 或者计算机可读非瞬态数据存储介质、比如 IC 卡、SD 卡或者 DVD 中存储用于实施功能的程序、表和文件的信息。

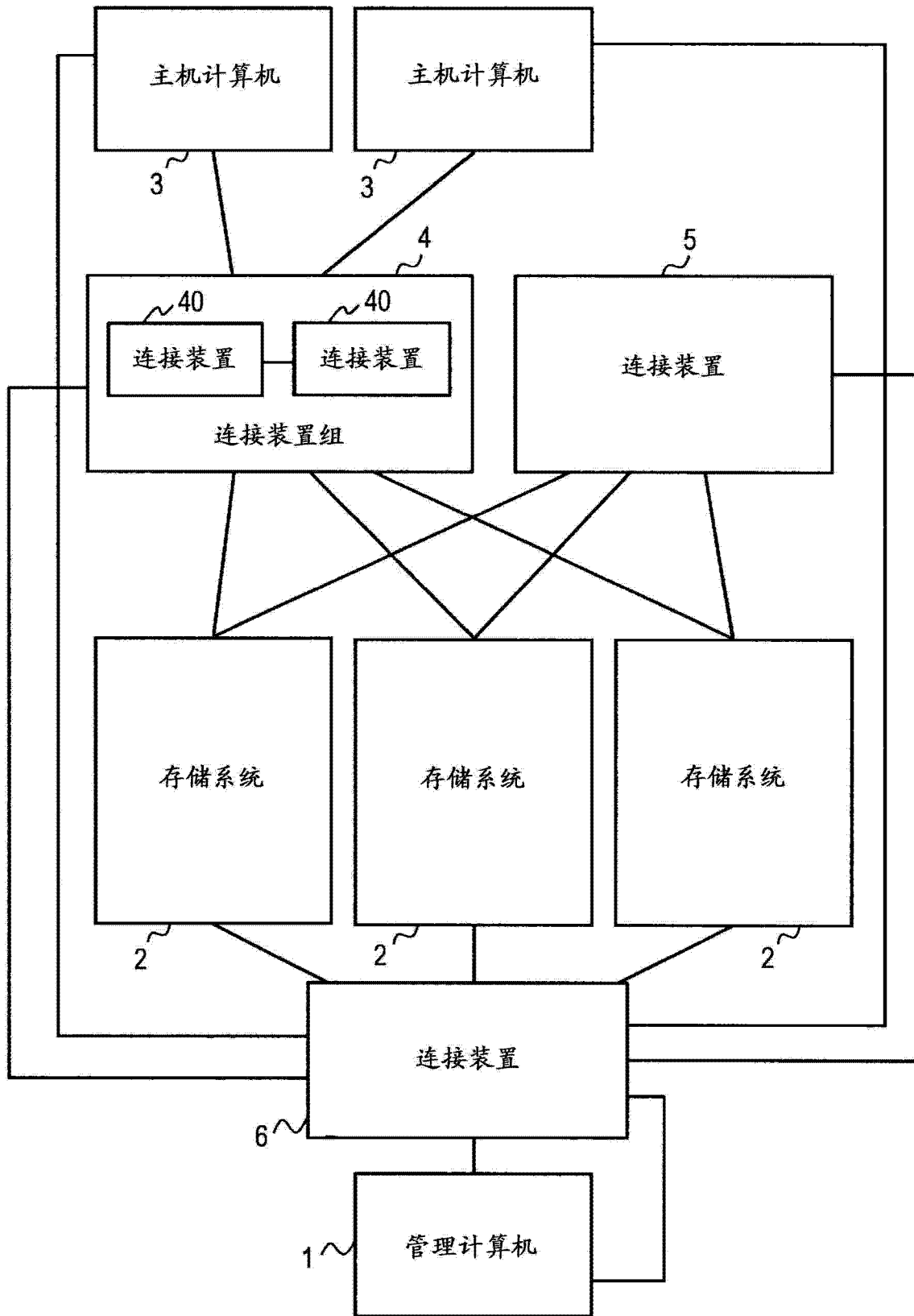


图 1

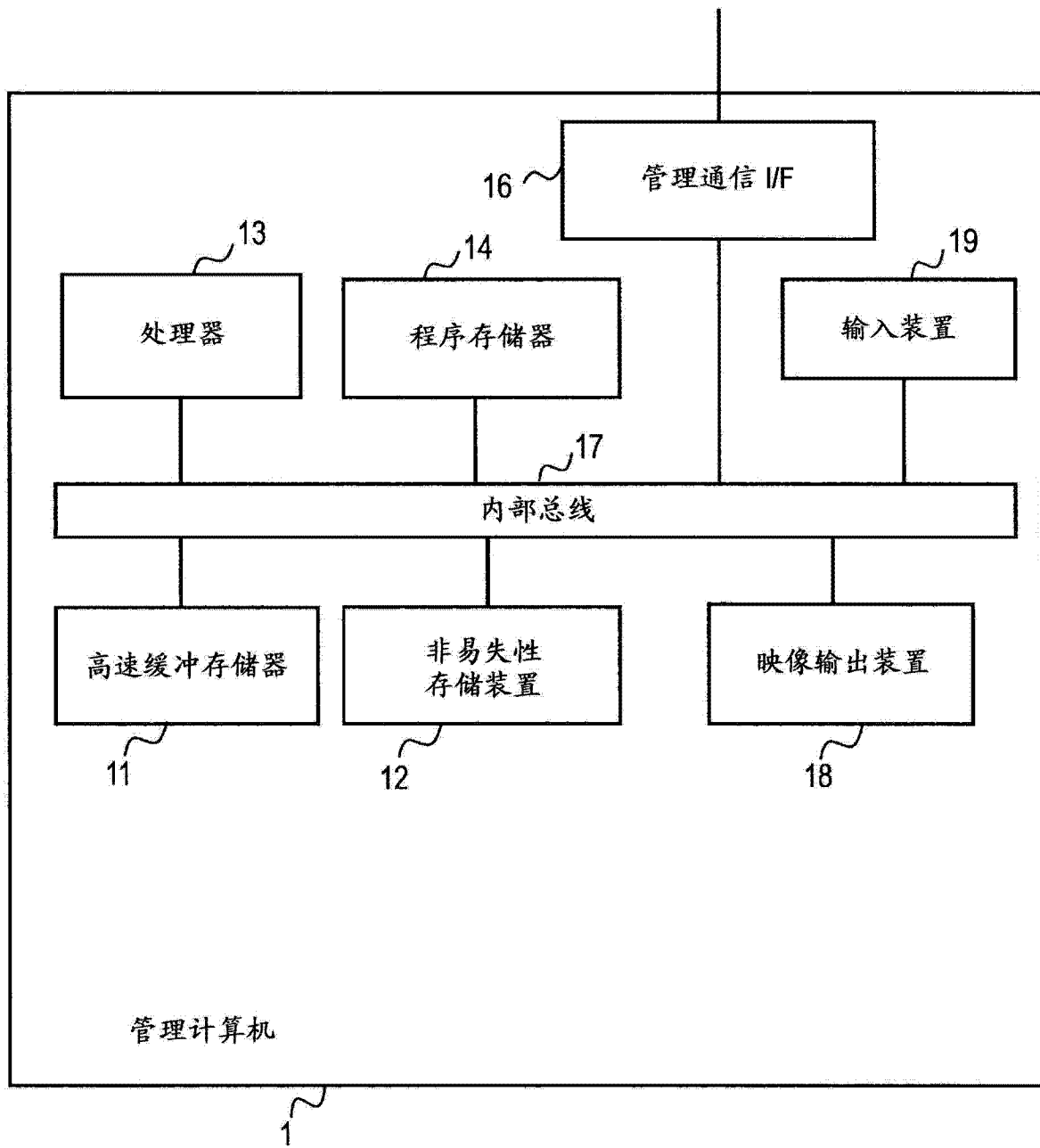


图 2

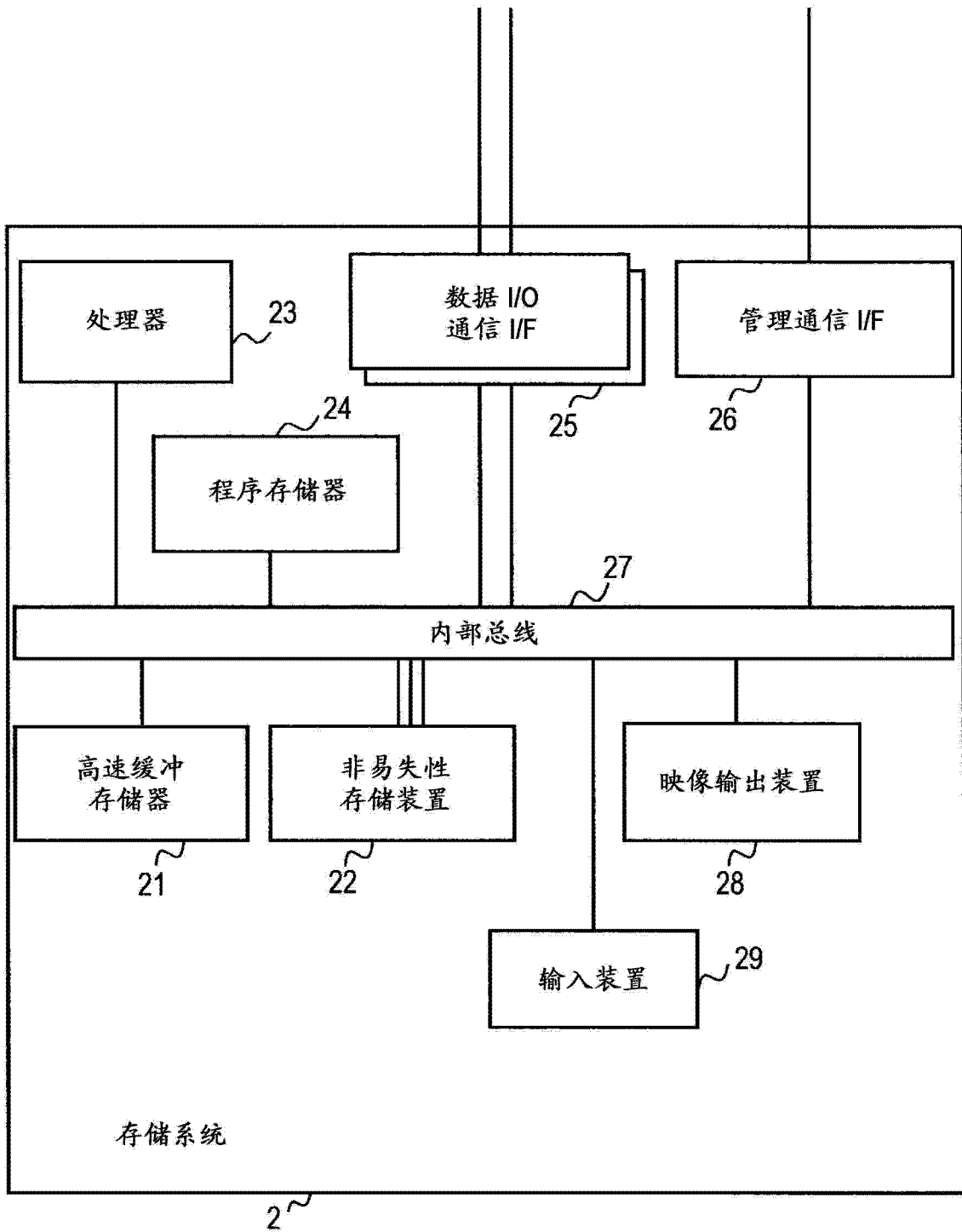


图 3

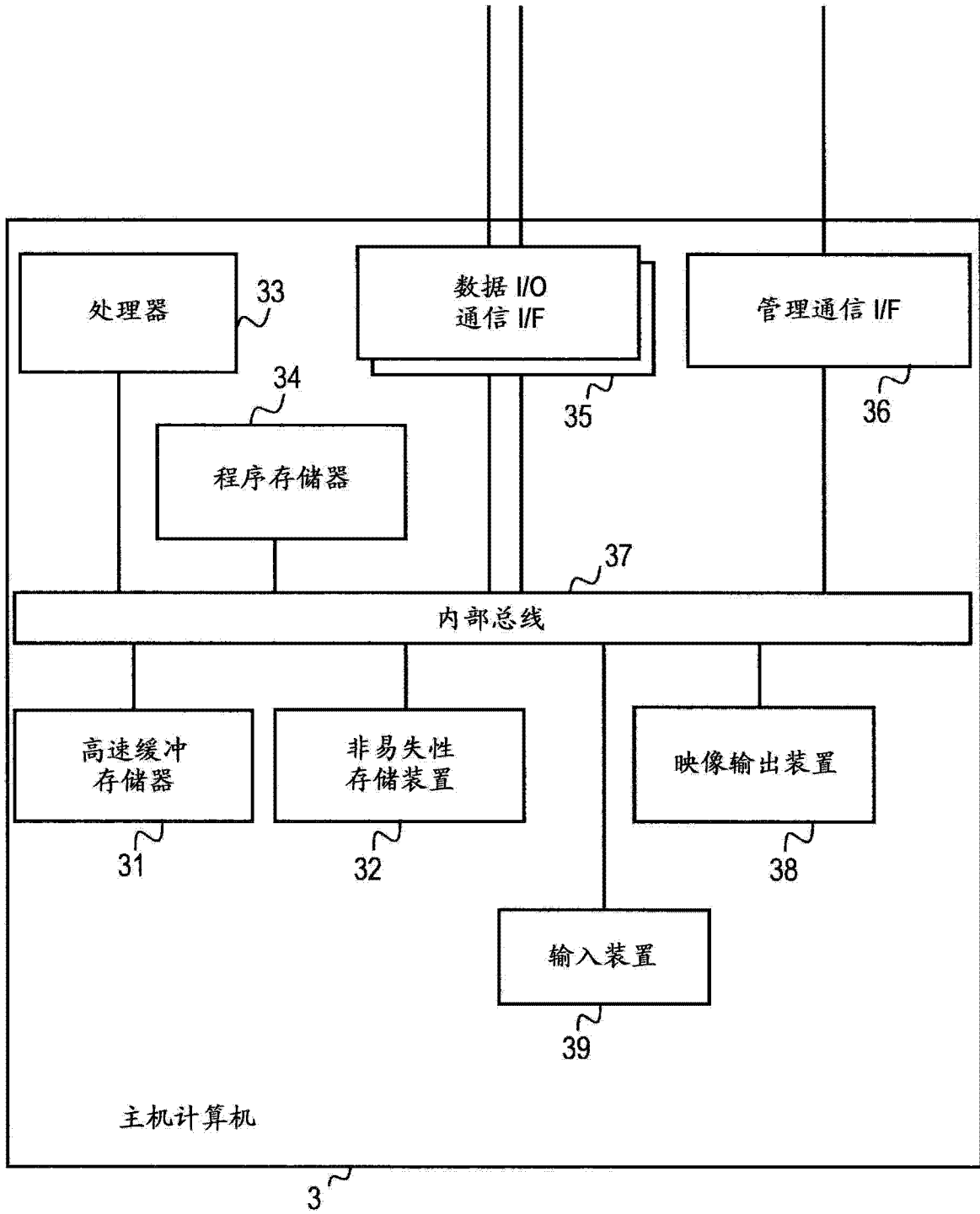


图 4

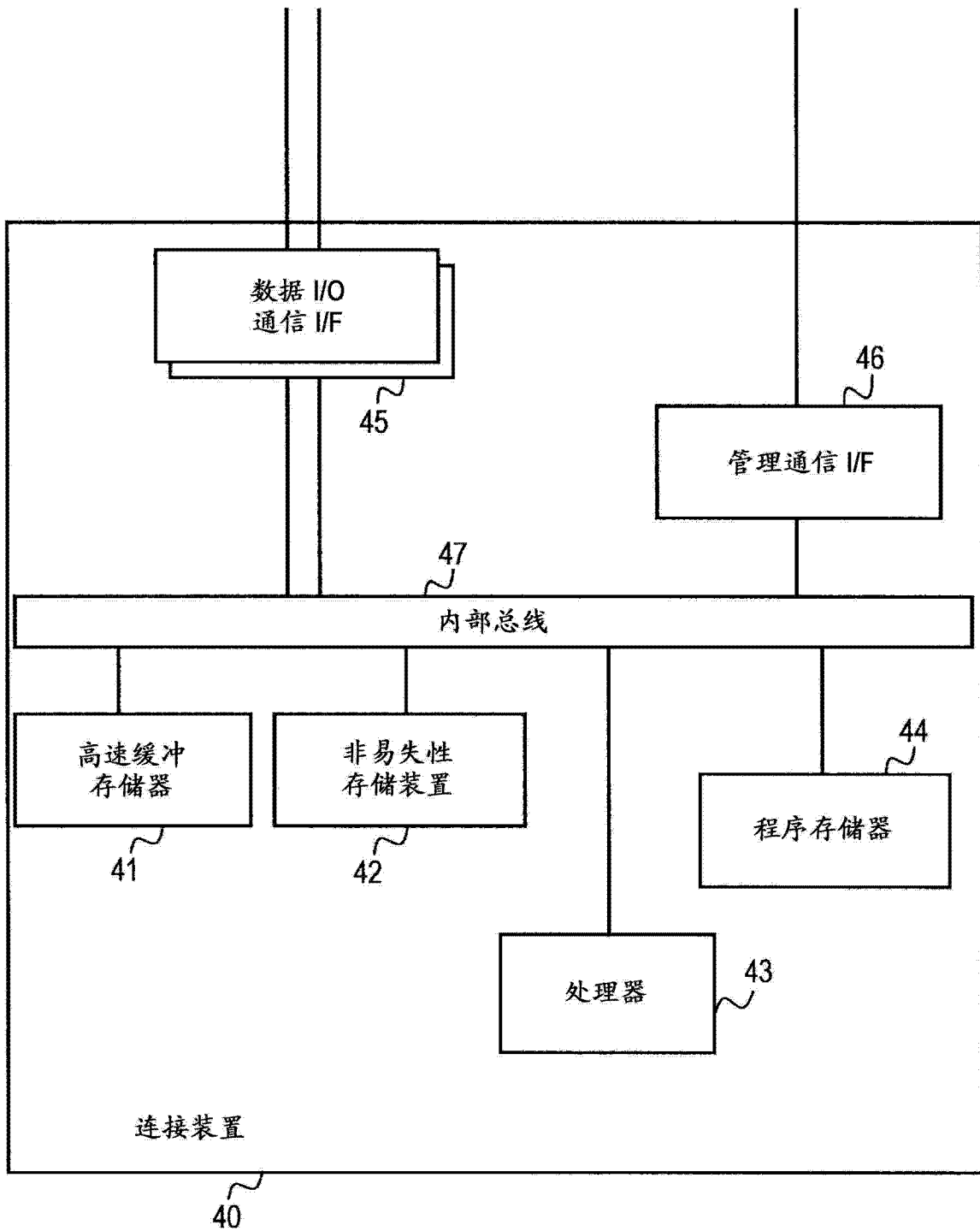


图 5

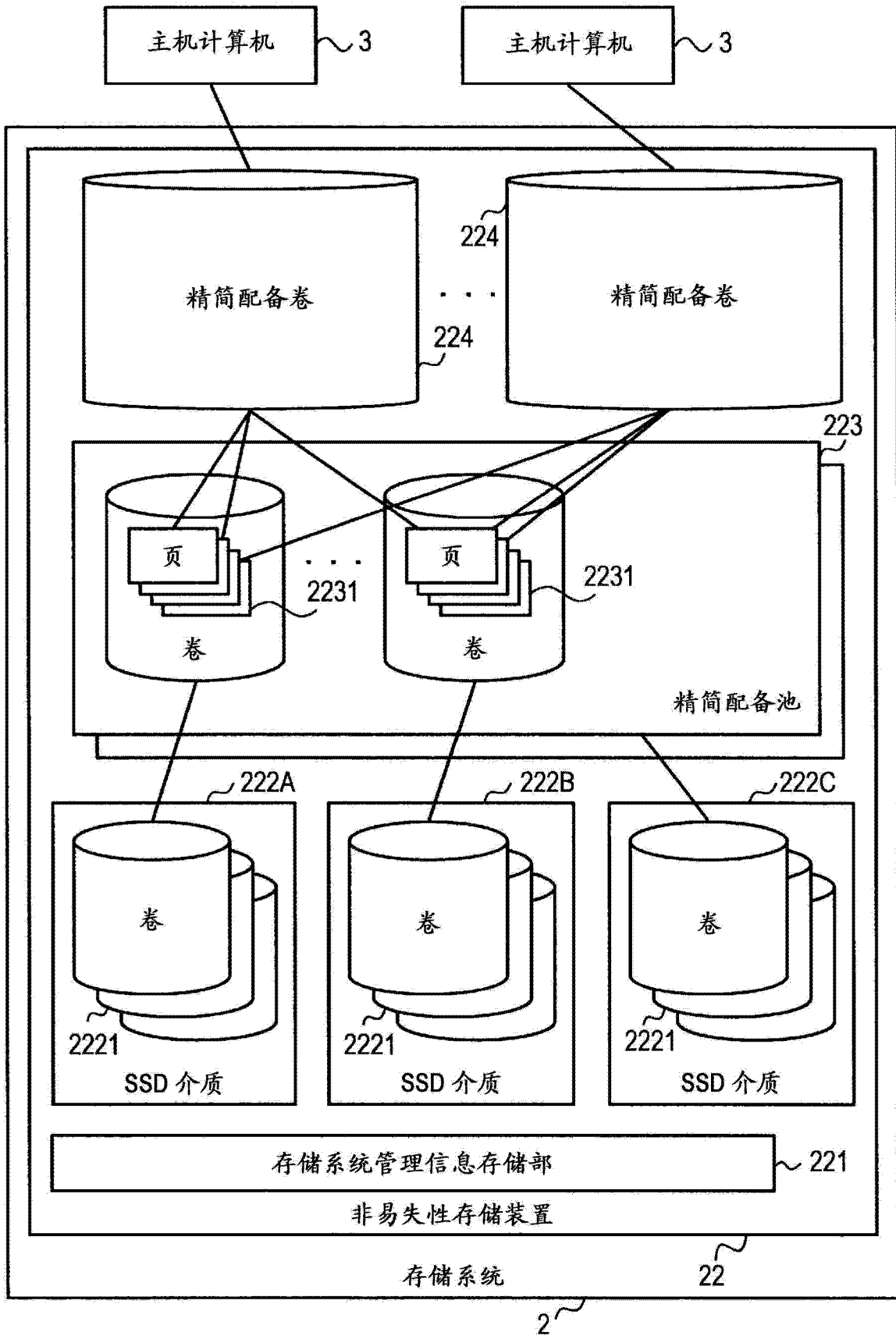


图 6

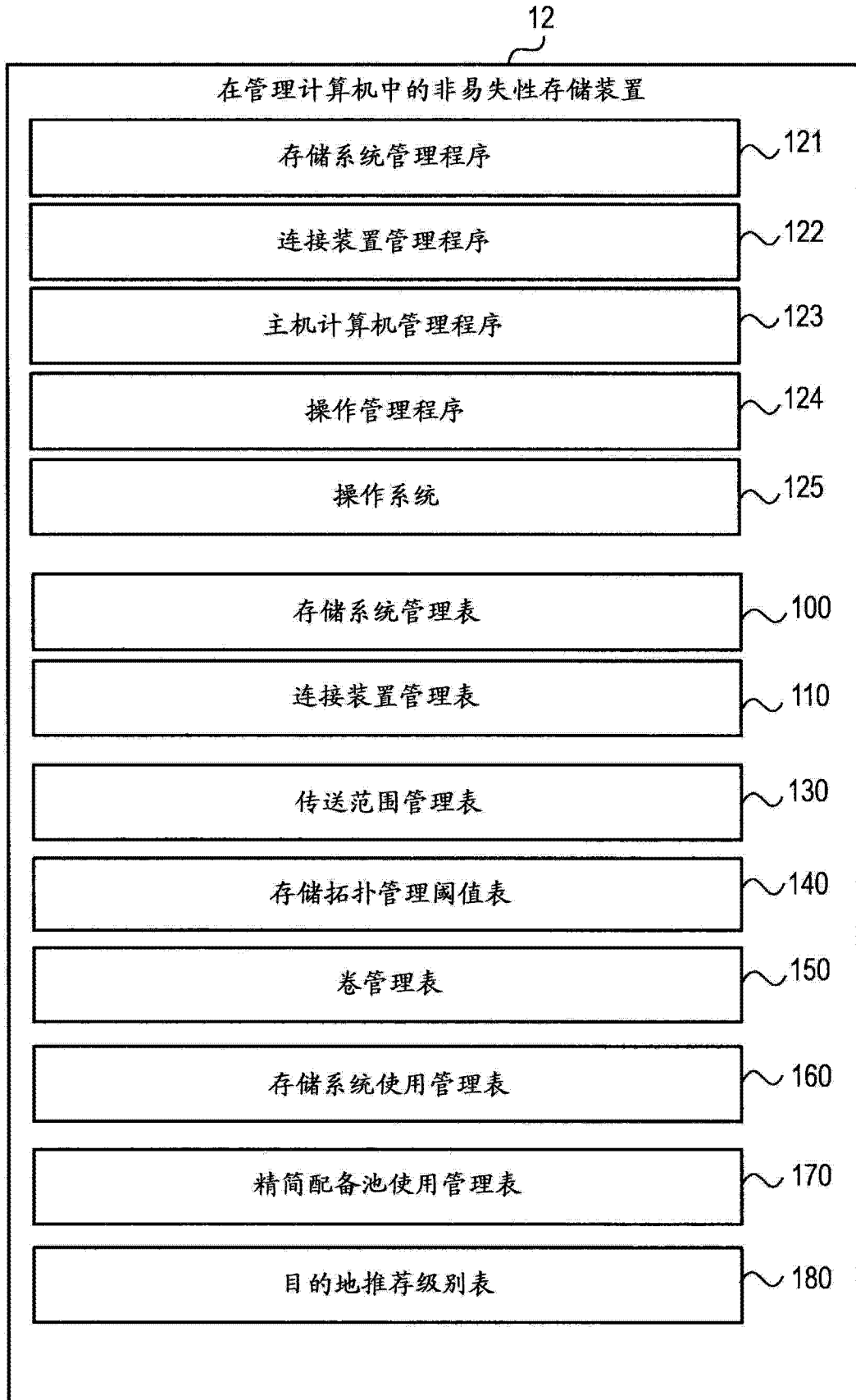


图 7

100

1001 存储系统 ID	1002 管理端口 IP 地址
ST1	192.168.1.101
ST2	192.168.1.102
ST3	192.168.1.103
ST4	192.168.1.104
.

图 8

110

1101 连接装置 ID	1102 管理端口 IP 地址
SW1	192.168.1.201
SW2	192.168.1.202
SW3	192.168.1.203
SW4	192.168.1.204
.

图 9

130

传送范围管理表		
1301 单存储池 ID	1302 存储系统 ID	1303 端口 ID
SPool1	ST1	PT1
		PT2
		PT3
		PT4
	ST2	PT1
		PT2
		PT3
		PT4
	ST3	PT1
		PT2
		PT3
		PT4
SPool2	ST4	PT1
		PT2
		PT3
		PT4
	ST5	PT1
		PT2
		PT3
		PT4
...

图 10

140

存储拓扑管理阈值表

1401	用于中转连接 装置数目的阈值	2	8Gbps
1402	用于最低带宽路径的 平均带宽使用率的阈值	50%	
1403	用于最低带宽路径的 峰值带宽使用率的阈值	80%	
1404	用于延时的 阈值	3ms	

图 11

卷管理表

卷编号	容量	使用量	卷类型	存储系统 ID	卷 ID	端口 ID
1	10GB	5GB	常规	ST1	Vol1	PT1
2	10GB	8GB	常规	ST1	Vol2	PT1
3	10GB	3GB	常规	ST2	Vol1	PT1
4	40GB	3GB	精简配备	ST1	Vol3	PT1
5	40GB	4GB	精简配备	ST2	Vol2	PT1
6	10GB	1GB	精简配备	ST1	Vol4	PT1
7	10GB	1GB	池	ST1	Vol1	PT2
8	10GB	1GB	池	ST1	Vol2	PT2
9	10GB	2GB	池	ST2	Vol1	PT2
10	10GB	2GB	池	ST2	Vol2	PT2

图 12

160

存储系统使用管理表			
1601 存储系统 ID	1602 总容量	1603 容量阈值	1604 使用量
ST1	14100GB	60%	6130GB
ST2	200000GB	50%	8300GB
ST3	100000GB	60%	9200GB
...

图 13

170

精简设备池使用管理表

1701	1702	1703	1704	1705	1706	1707	1708	1709
池 ID	总容量	使用量	存储系统 ID	容量	使用量	容量阈值	IOPS	性能阈值
TPool1	100 GB	15 GB	ST1	30 GB	5 GB	60%	100 IOPS	1000 IOPS
			ST2	70 GB	10 GB	50%	2000 IOPS	1500 IOPS

图 14

180

目的地推荐级别表 (ST1, PT1)			
1801 单存储池 ID	1802 存储系统 ID	1803 目的地端口 ID	1804 目的地推荐级别
SPool1	ST1	PT2	★★★★★
		PT3	★★★★★
		PT4	★★★★★
	ST2	PT1	★★★★★
		PT2	★★★★
		PT3	★★★
		PT4	★★
	ST3	PT1	★
		PT2	★
		PT3	★
		PT4	★

图 15

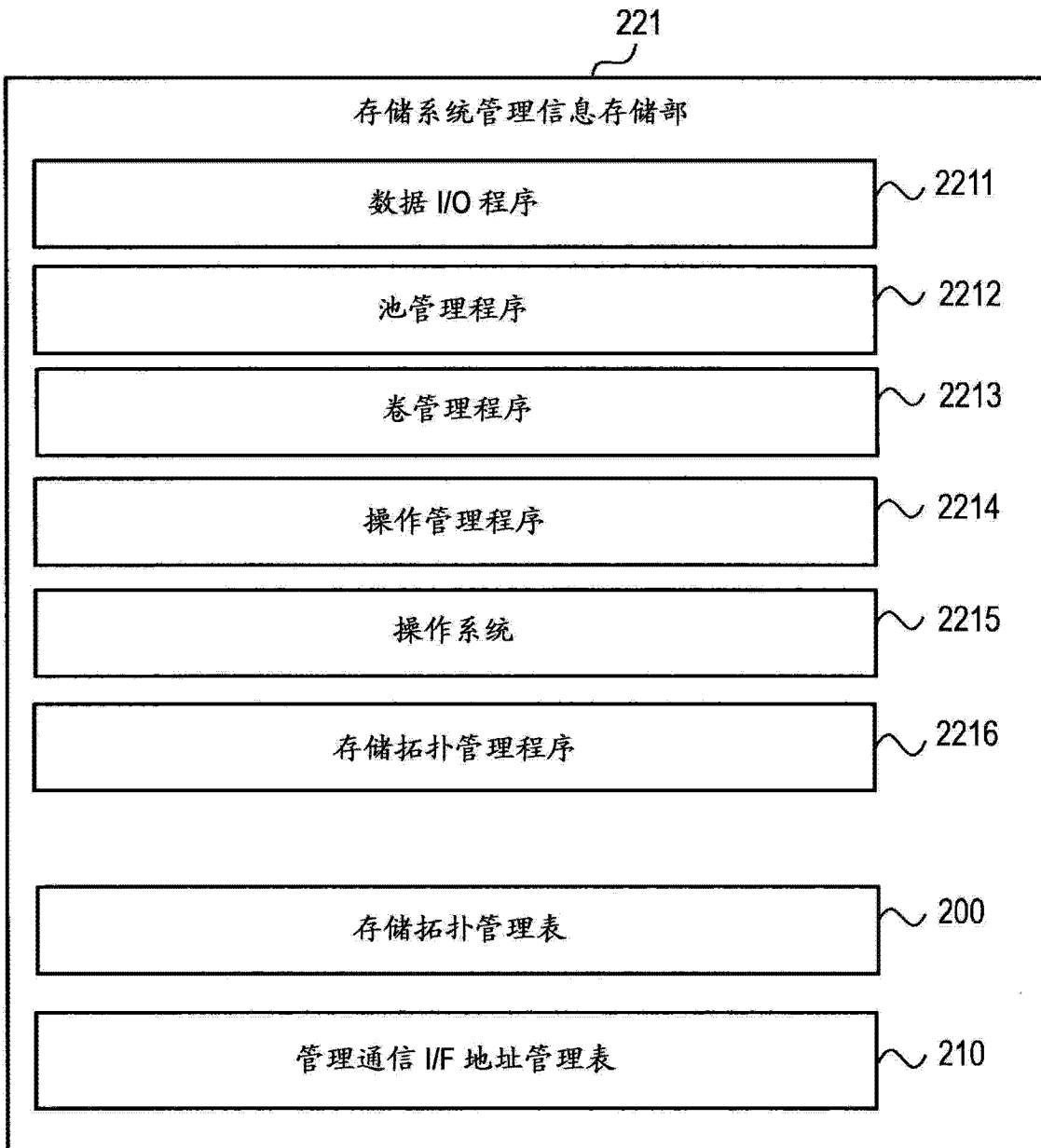


图 16

存储拓扑管理表

2001 ~ 2002	2003	2004	2005	2006	2007	2008	
端口 ID	相反存储系统 ID	相反端口 ID	最小带宽	中转连接装置数目	最低带宽路径的平均带宽使用率	最低带宽路径的峰值带宽使用率	延时
PT1	ST1	PT2	40Gbps	1	10%	50%	>1ms
		PT3	40Gbps	1	8%	40%	>1ms
	ST2	PT4	40Gbps	1	3%	25%	>1ms
		PT1	40Gbps	1	15%	20%	>1ms
PT2	ST2	PT2	2Gbps	2	80%	90%	2.4ms
		PT3	2Gbps	2	80%	90%	2.5ms
	ST3	PT4	2Gbps	2	80%	90%	2.0ms
		PT1	20Gbps	3	85%	90%	3.7ms
PT2	ST1
		PT1	40Gbps	1	15%	55%	>1ms
	ST2	PT3	40Gbps	1	12%	40%	>1ms
		PT4	40Gbps	1	10%	25%	>1ms
PT2	ST2	PT1	40Gbps	1	15%	20%	>1ms
	
	ST3
	

图 17

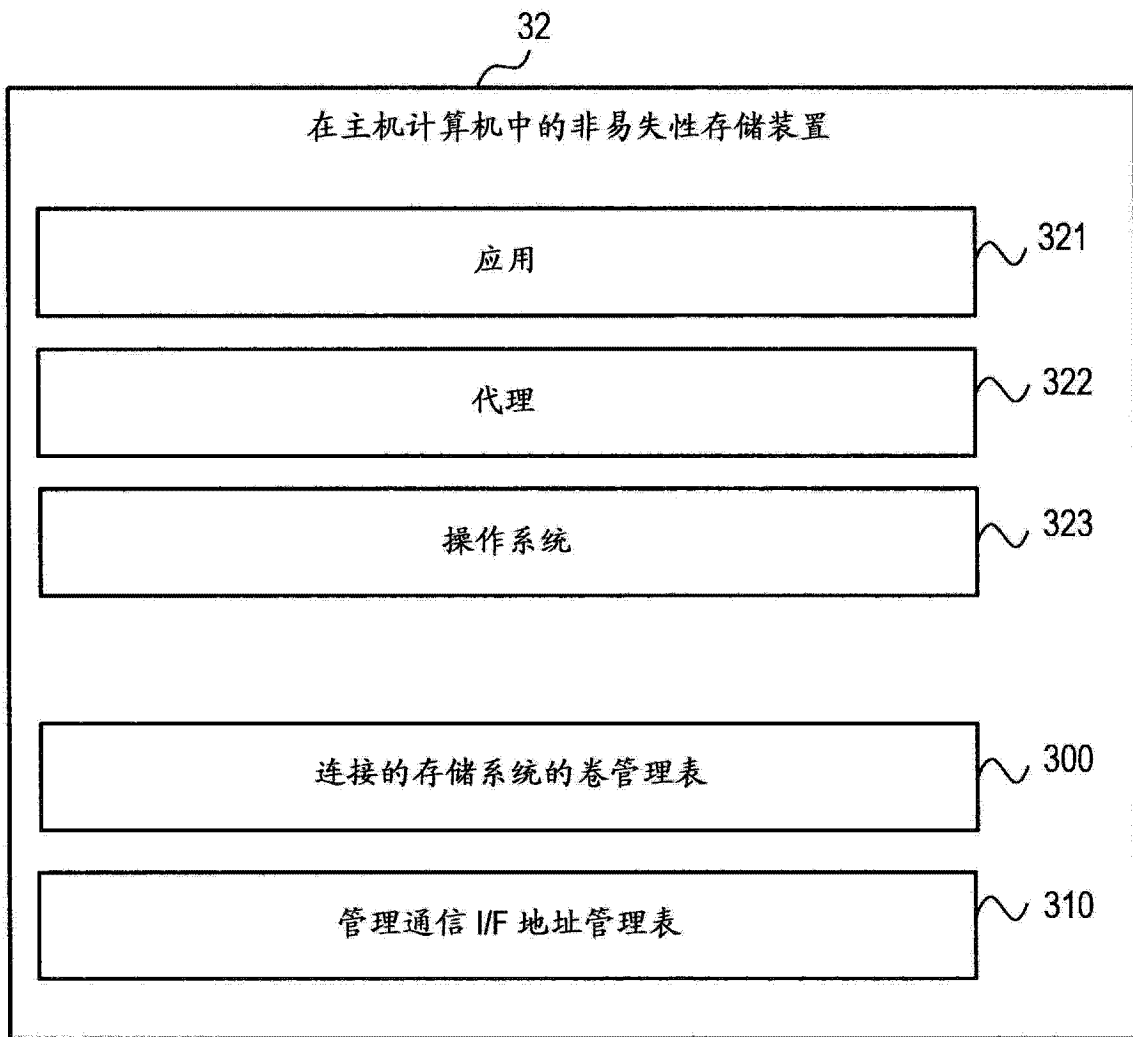


图 18

300

连接的存储系统的卷管理表

3001 存储系统 ID	3002 卷 ID	3003 端口 ID
ST1	Vol3	PT1
...

图 19

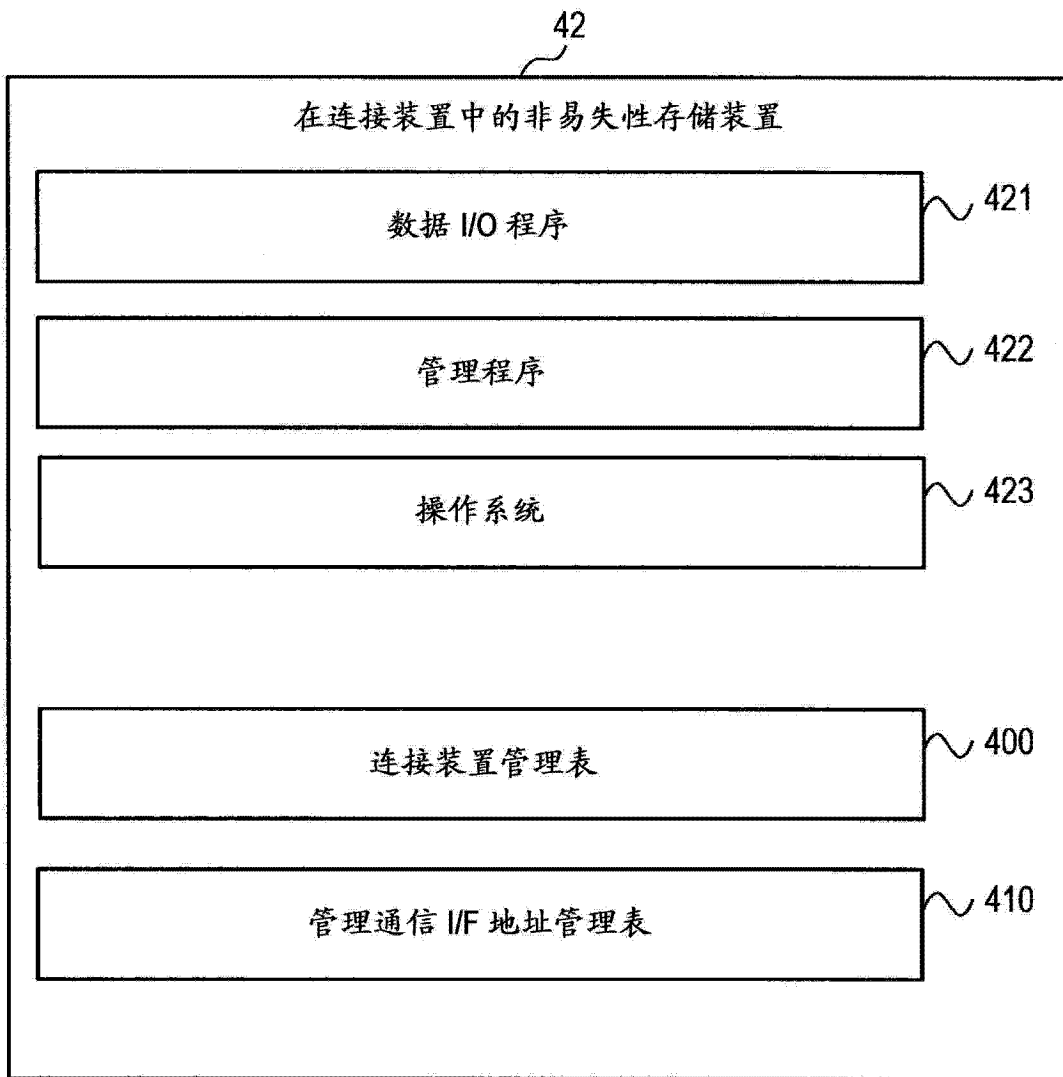


图 20

400

连接装置管理表					
4001	4002	4003	4004	4005	4006
端口 ID	相反装置 ID	连接的 端口 ID	物理带宽	流量的 平均数量	流量的 峰值数量
SPT1	ST1	PT1	40Gbps	6Gbps	22Gbps
SPT2	ST1	PT2	40Gbps	4Gbps	20Gbps
SPT3	ST1	PT3	40Gbps	3.2Gbps	16Gbps
SPT4	ST1	PT4	40Gbps	1.2Gbps	10Gbps
SPT5	ST2	PT1	40Gbps	6Gbps	8Gbps
SPT6	SW2	SPT1	2Gbps	1.6Gbps	1.8Gbps
SPT7	SW3	SPT1	8Gbps	4.2Gbps	6.0Gbps
SPT8	SW4	SPT1	8Gbps	5.0Gbps	6.4Gbps

图 21

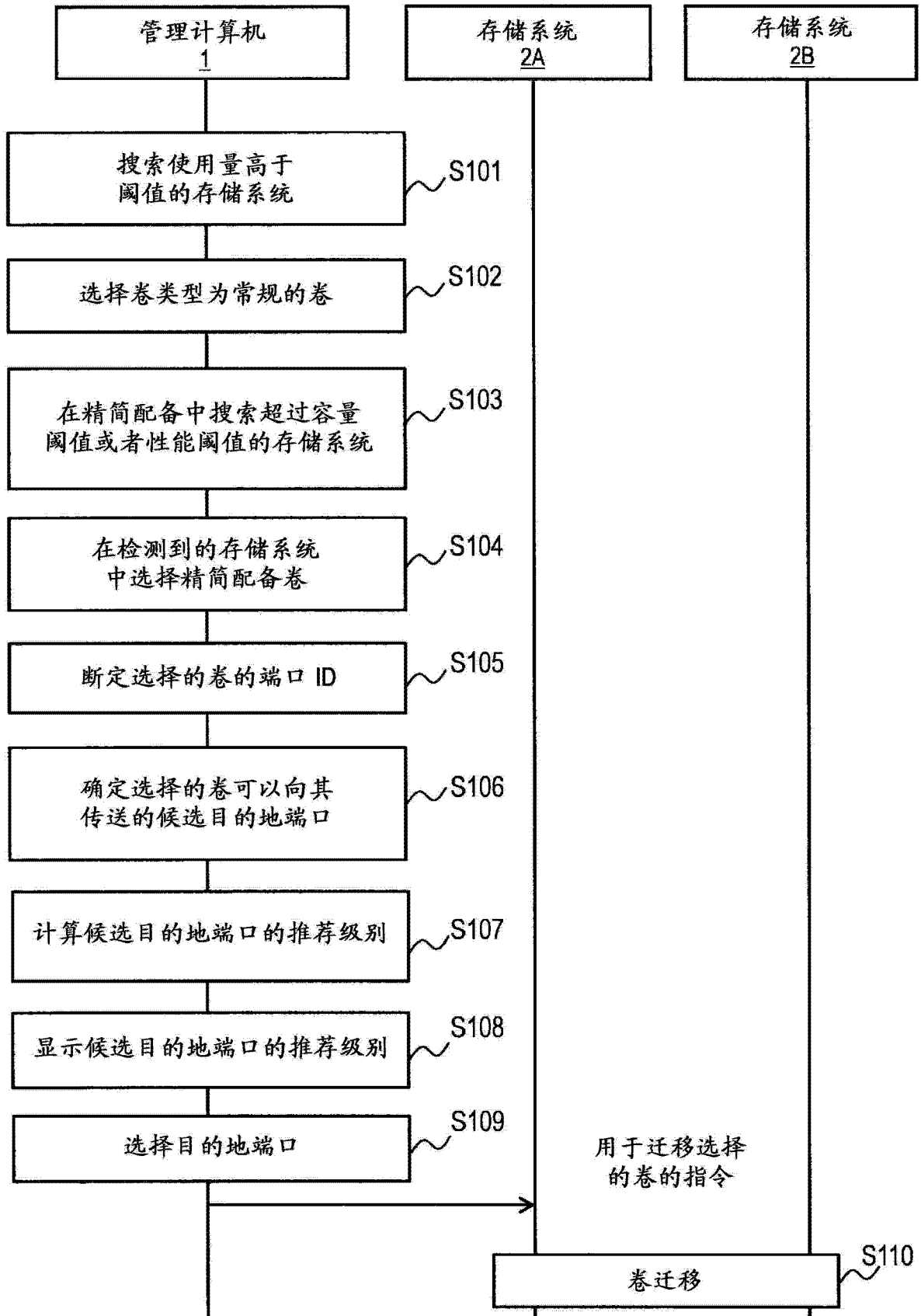


图 22

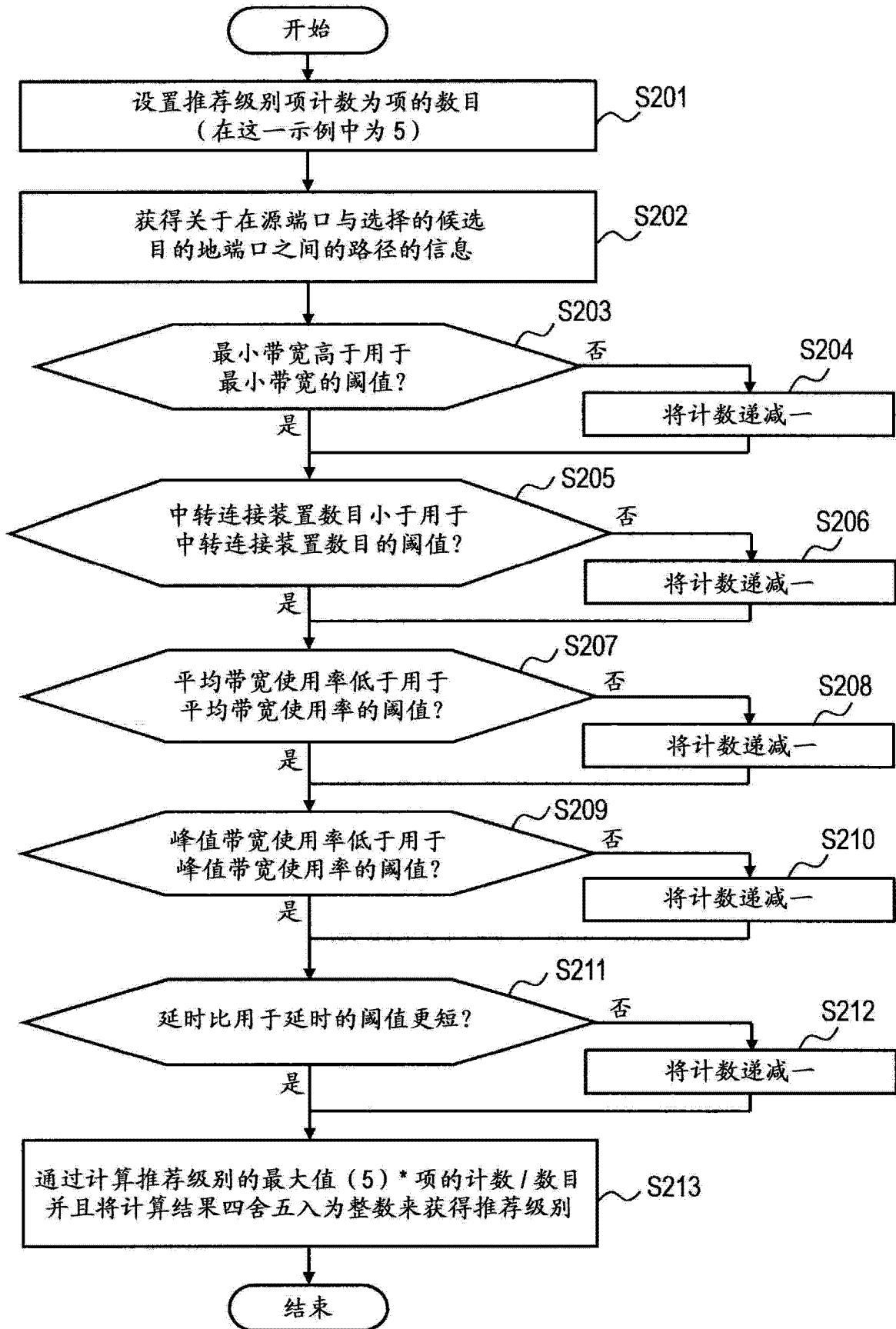


图 23

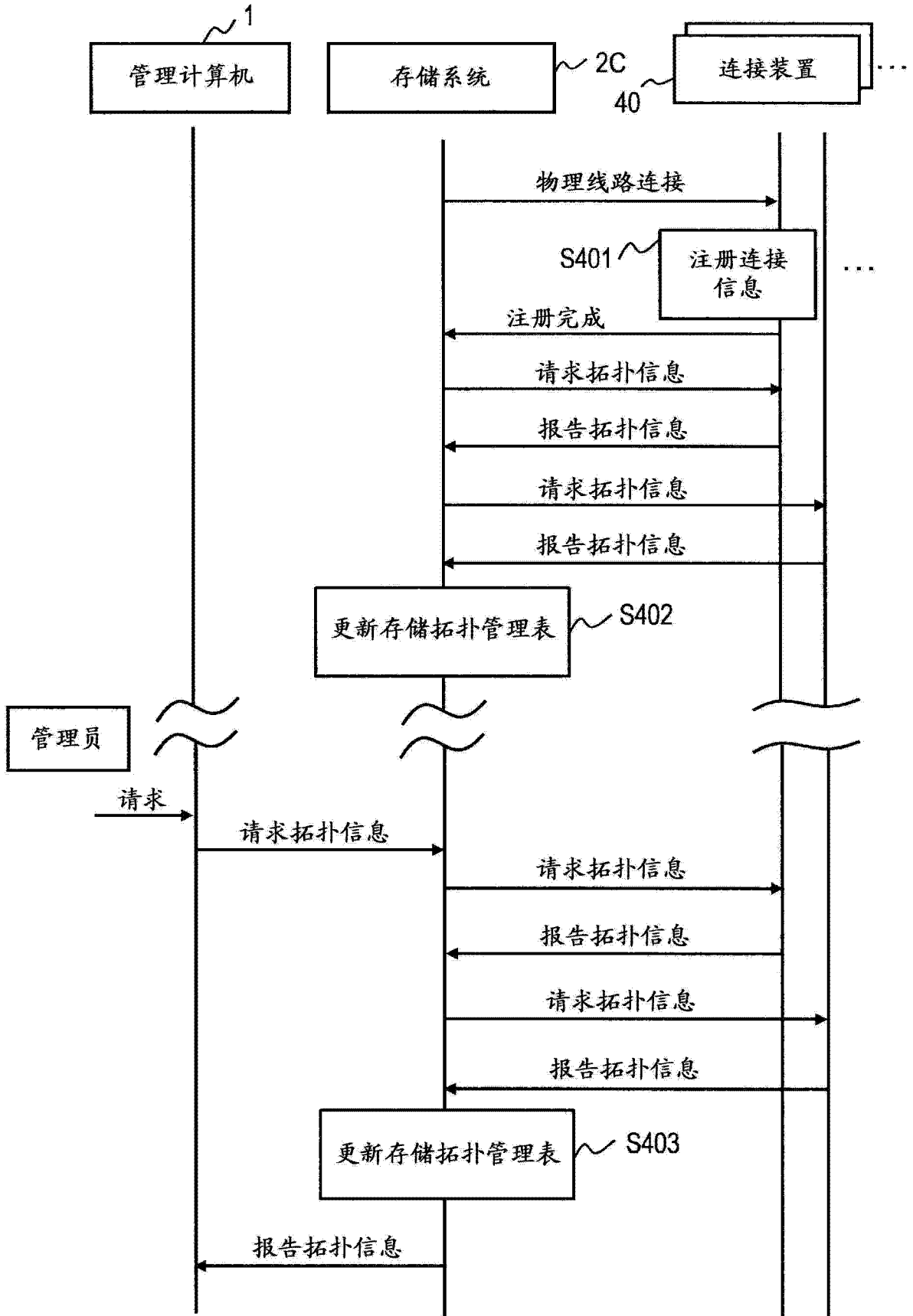


图 24

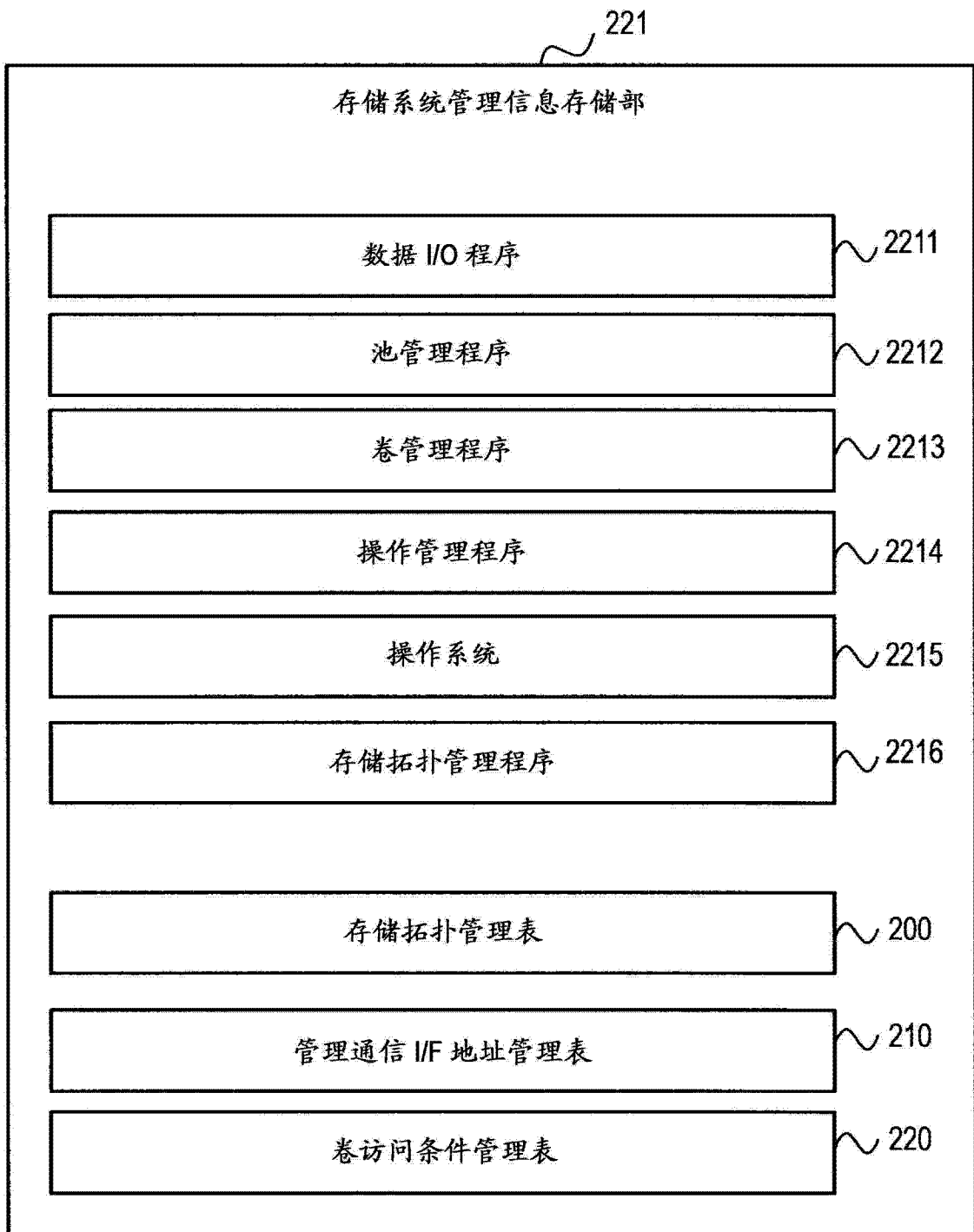


图 25

220

卷访问条件管理表			
2201 端口 ID	2202 卷 ID	2203 在卷访问中使用的 平均带宽	2204 在卷访问中使用的 峰值带宽
PT1	Vol1	300Mbps	800Mbps
	Vol2	200Mbps	300Mbps
	Vol3	50Mbps	200Mbps
	Vol4	100Mbps	150Mbps
PT2	Vol1	30Mbps	100Mbps
	Vol2	500Mbps	1Gbps
...

图 26

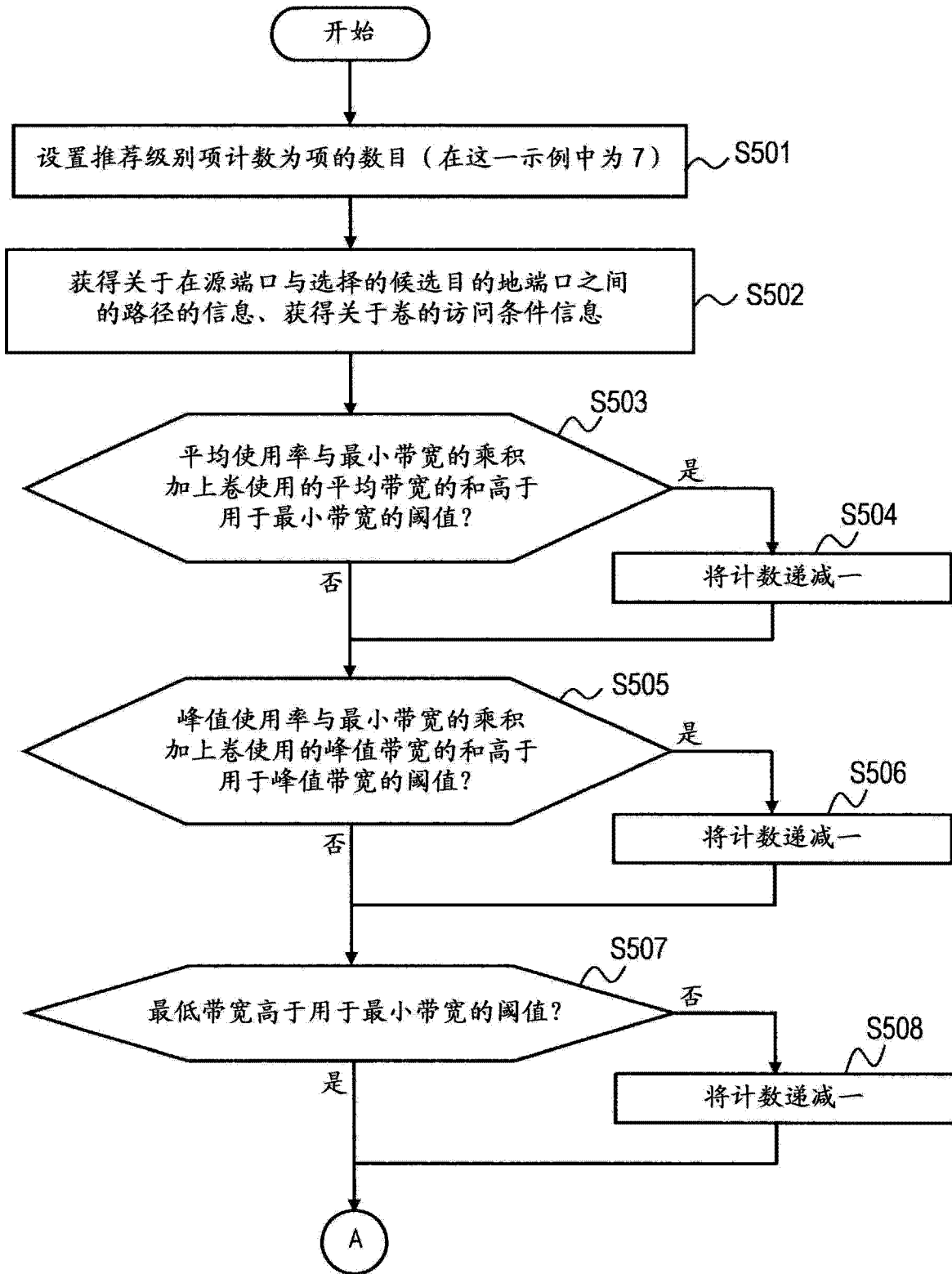


图 27A

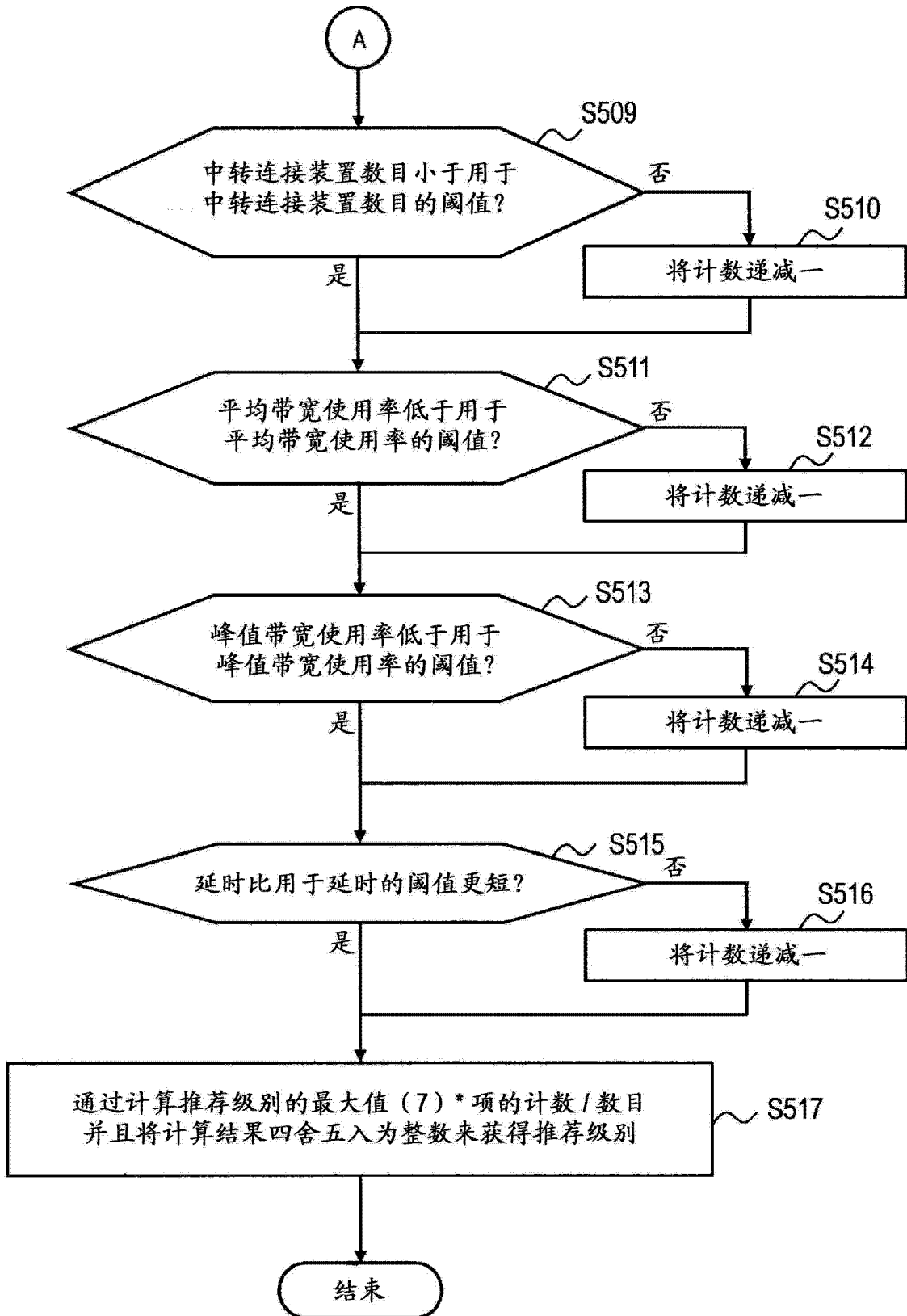


图 27B

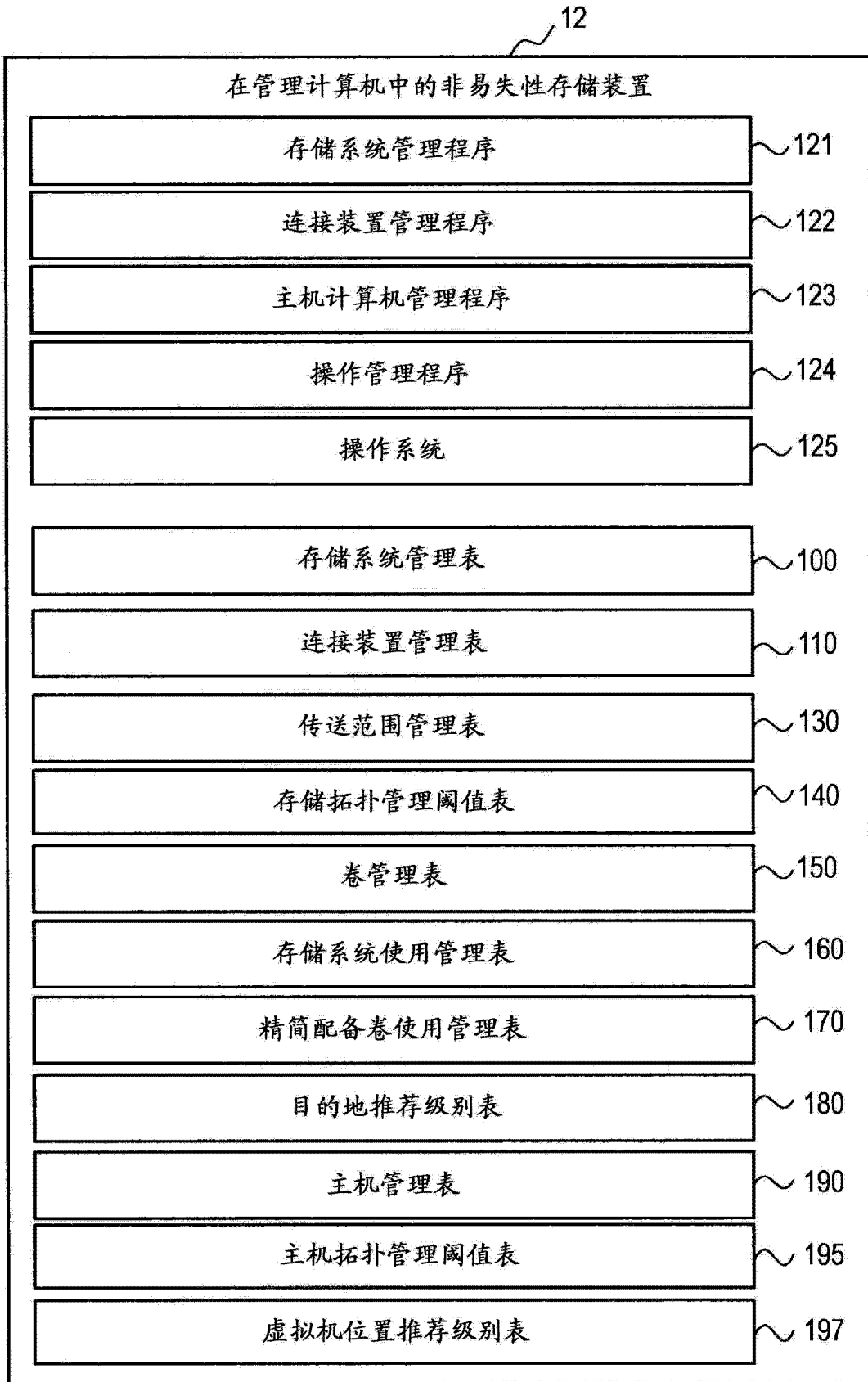


图 28

190

主机管理表

1901	1902	1903	1904	1905	1906	1907	1908
主机 ID	管理端口 IP 地址	安装的存储器数量	存储器使用量	CPU 核数目	CPU 频率	CPU 使用率	端口 ID
HS1	192.168.1.11	16GB	10GB	4	2.0GHz	20%	HBA1/HBA2
HS2	192.168.1.12	16GB	12GB	4	2.0GHz	60%	HBA1/HBA2
HS3	192.168.1.13	32GB	20GB	8	2.3GHz	30%	HBA1
HS4	192.168.1.14	32GB	28GB	8	2.3GHz	80%	HBA1
...

图 29

195
主机拓扑管理阈值表

1951	1952	1953	1954	1955
用于最小带宽的阈值	用于中转连接装置数目的阈值	用于最低带宽路径的平均带宽使用率的阈值	用于最低带宽路径的峰值带宽使用率的阈值	用于延时的阈值
8Gbps	2	50%	80%	3ms

图 30

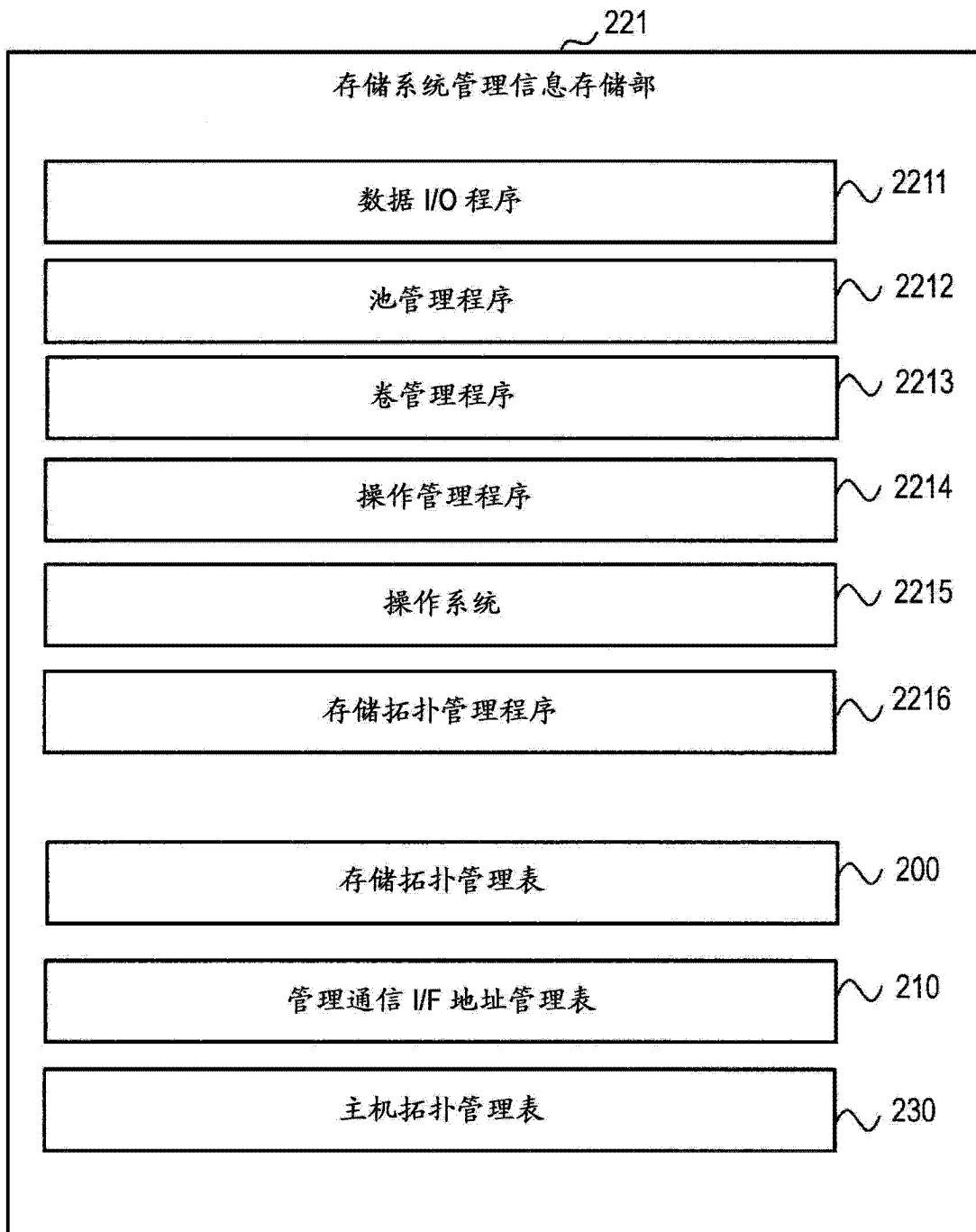


图 31

230
主机拓扑管理表

2301 端口 ID	2302 相反主机 计算机 ID	2303 相反端口 ID	2304 最小带宽	2305 中转连接 装置数目	2306 最低带宽路径的 平均带宽使用率	2307 最低带宽路径的 峰值带宽使用率	2308 延时
PT1	HS1	HBA1	2Gbps	2	80%	90%	2.3ms
		HBA2	2Gbps	2	80%	90%	2.2ms
	HS2	HBA1	2Gbps	2	80%	90%	2.8ms
		HBA2	2Gbps	2	80%	90%	2.9ms
PT2	HS3	HBA1	40Gbps	1	15%	20%	<1ms
		HBA1	40Gbps	1	20%	25%	<1ms
	HS4	HBA1	2Gbps	2	80%	90%	2.3ms
		HBA2	2Gbps	2	80%	90%	2.2ms
...	HS1	HBA1	2Gbps	2	80%	90%	2.8ms
		HBA2	2Gbps	2	80%	90%	2.9ms
	HS2	HBA1	40Gbps	1	15%	20%	<1ms
		HBA1	40Gbps	1	20%	25%	<1ms
...	

图 32

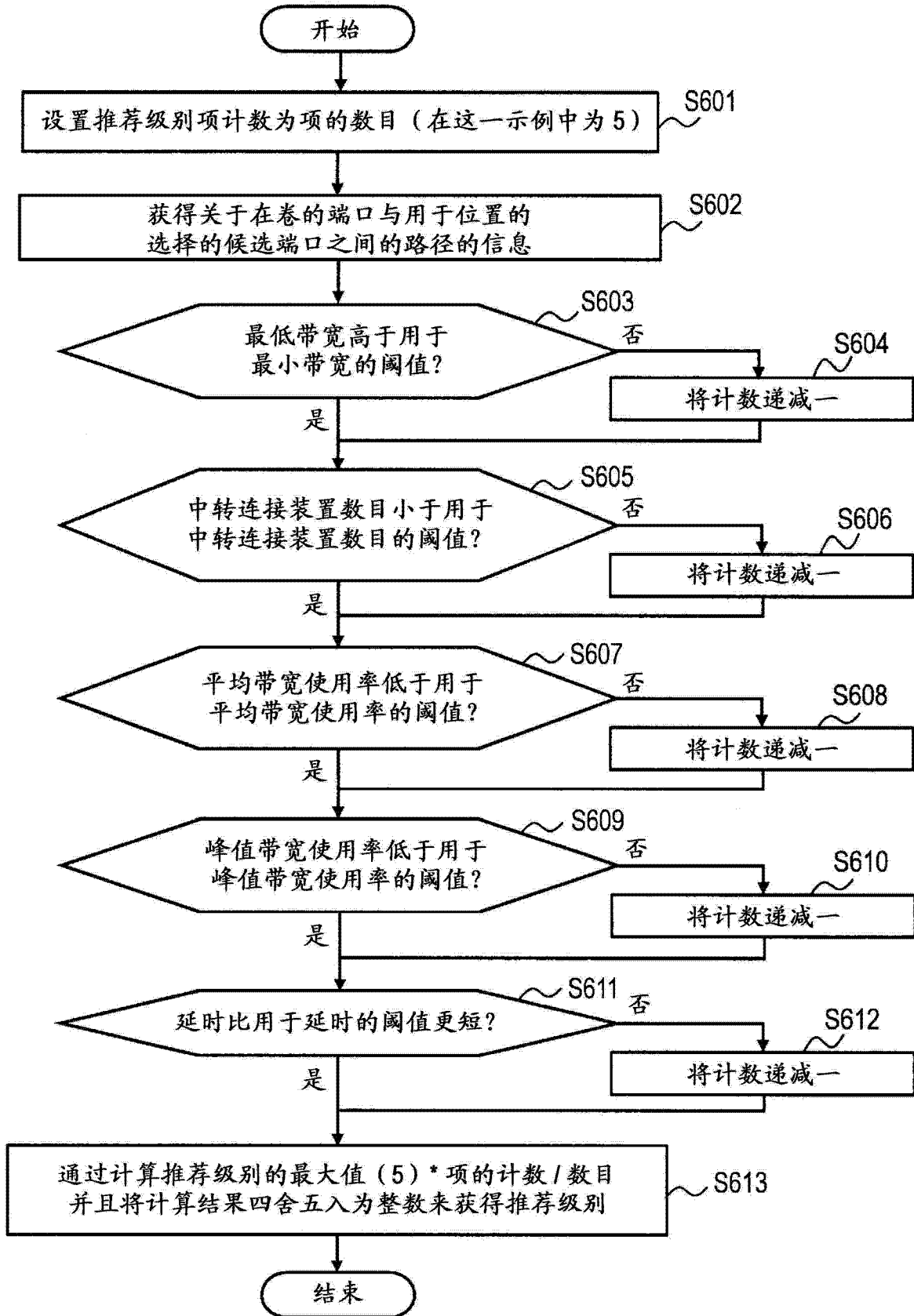


图 33

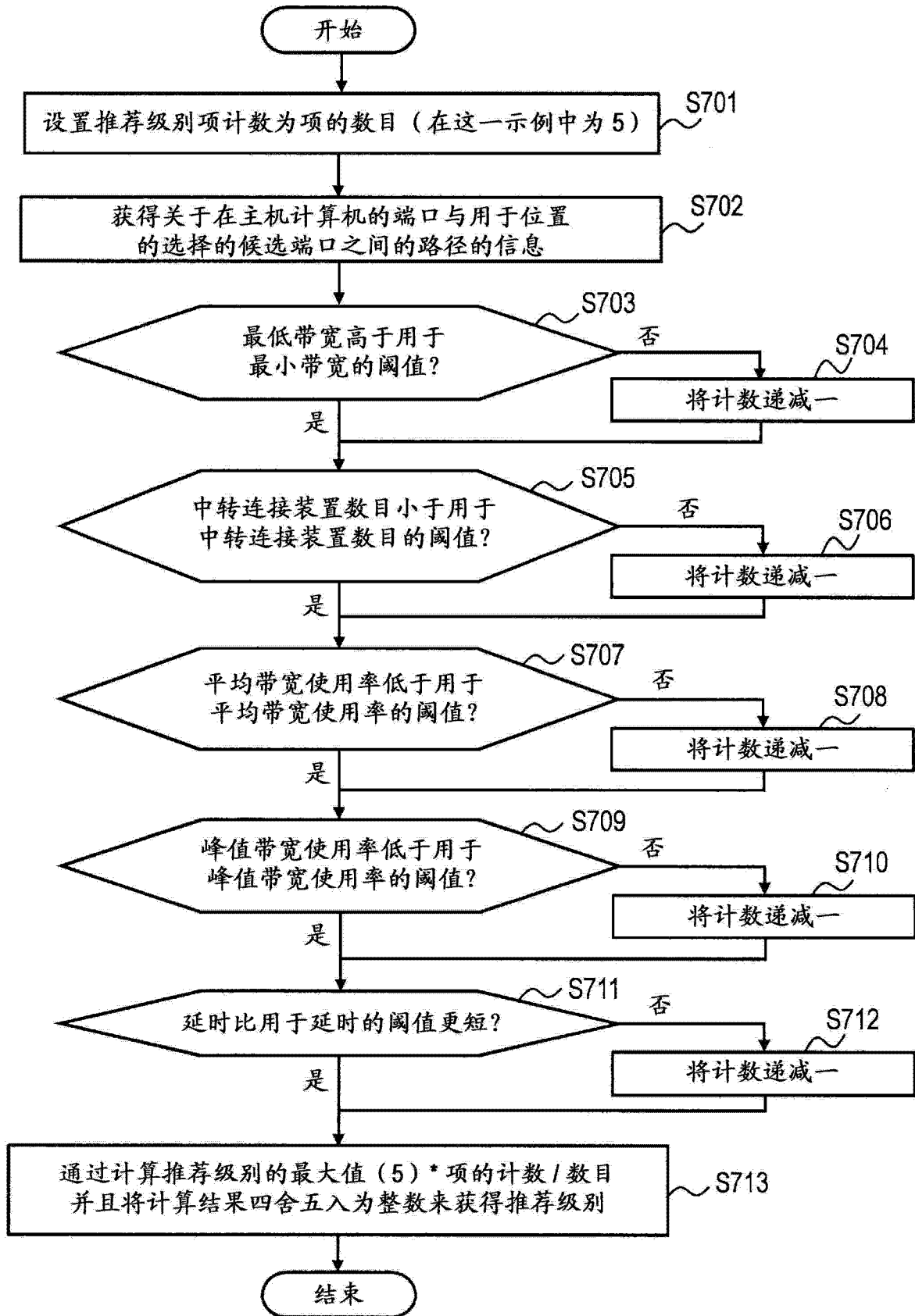


图 34