



[12] 发明专利申请公开说明书

[21] 申请号 200410088388.1

[43] 公开日 2005年5月11日

[11] 公开号 CN 1614936A

[22] 申请日 2004.11.8

[21] 申请号 200410088388.1

[30] 优先权

[32] 2003.11.6 [33] US [31] 60/517776

[32] 2004.2.6 [33] US [31] 10/773543

[71] 申请人 西门子医疗健康服务公司

地址 美国宾夕法尼亚州

[72] 发明人 A·莫尼茨尔

[74] 专利代理机构 中国专利代理(香港)有限公司

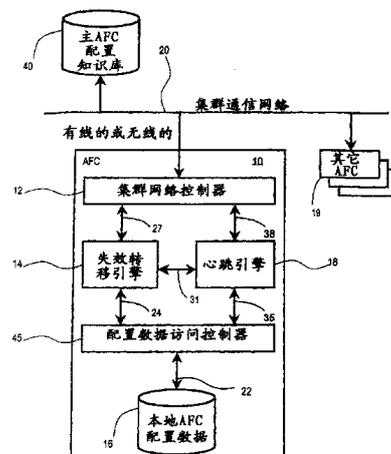
代理人 张雪梅 张志醒

权利要求书4页 说明书12页 附图7页

[54] 发明名称 处理设备管理系统

[57] 摘要

一种系统自动地自适应修改处理设备组(集群)的备份设备的失效转移配置优先级列表,以便提高可用性并降低与人工配置相关的风险与成本。网络处理设备组内的独立处理设备利用一种系统来管理该组内的设备中发生的操作故障。所述系统包括接口处理器,它用来维持标识用于响应于第一处理设备的操作故障而接管执行第一处理设备的任务的第二处理设备的转换信息,并且用来响应于出现在组内的另一处理设备中的转换信息方面的变化来更新转换信息。操作检测器检测第一处理设备的操作故障。此外,故障控制器响应于第一处理设备的操作故障的检测来启动由第二处理设备执行指定将由第一处理设备执行的任务。



1. 一种供网络处理设备组的独立处理设备使用的系统，用于管理所述组内的设备中发生的操作故障，包括：

5 接口处理器，用于维持标识用于响应于所述第一处理设备的操作故障而接管执行第一处理设备的任务的第二处理设备的转换信息，以及用于响应于出现在所述组的另一个处理设备中的转换信息方面的变化来更新所述转换信息；

操作检测器，用于检测所述第一处理设备的操作故障；和

10 故障控制器，用于响应于所述第一处理设备的操作故障的检测来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任

2. 根据权利要求1所述的系统，其中：

所述网络处理设备组的每个独立的处理设备包括有存储转换信息的知识库，以及

15 所述独立处理设备都保持通信以便在所述独立处理设备中维持一致的转换信息。

3. 根据权利要求1所述的系统，其中：

所述转换信息包括用于响应于所述第一处理设备的操作故障而承担执行第一处理设备的任务的处理设备的优先表，

20 响应于来自所述组的另一处理设备的通信而动态地更新所述优先表，以及

所述优先表表示用于响应于所述第一处理设备的操作故障而承担执行第一处理设备的任务的被动非操作处理设备。

4. 根据权利要求3所述的系统，其中：

25 响应于多个因素而动态地更新所述优先表，所述多个因素包括下列因素中的至少一个：(a)所述组内的另一处理设备的操作故障的检测，和(b)低于预定阈值的所述组内的另一处理设备的可用存储器的检测，以及

30 所述多个因素包括下列中的至少一个：(a)超过预定阈值的所述组内的另一处理设备的工作负载的检测、(b)超过预定阈值的所述组内的另一个处理设备的CPU(中央处理器)资源的使用的检测，和(c)在预定时间周期内超过预定阈值的所述组内的另一处理设备的多个

I/O (输入-输出) 操作的检测。

5. 根据权利要求3所述的系统, 其中:

所述优先表被动态地更新成表示至少下列其中之一状态信息:

- 5 (a) 已检测到的所述组内另一处理设备的状态从可用到不可用的变化, 和 (b) 已检测到的所述组内另一处理设备的状态从不可用到可用的变化, 包括:

所述接口处理器根据由所述组内的不同处理设备提供的状态信息来判断所述组内处理设备的状态。

6. 根据权利要求1所述的系统, 包括:

- 10 所述接口处理器询问所述组内的其它处理设备以便识别出现在所述组内另一处理设备中的转换信息方面的变化, 其中:

处理设备的操作故障包括 (a) 软件执行故障和 (b) 硬件故障中的至少一个。

7. 根据权利要求1所述的系统, 其中:

- 15 处理设备包括下列中至少一个: (a) 服务器、(b) 计算机、(c) PC、(d) PDA、(e) 电话、(f) 经由无线通信进行通信的处理设备、(g) 电视、(h) 机顶盒、和 (i) 包括可执行软件的网络设备并且
所述组内的独立处理设备包括有与所述组内其它处理设备相类似的软件。

20 8. 根据权利要求1所述的系统, 其中:

所述组包括集群, 而处理设备包括节点。

9. 一种供网络处理设备组内的独立处理设备使用的系统, 用于管理所述组内的设备中发生的操作故障, 包括:

独立处理设备, 包括:

- 25 知识库, 包含标识用于响应于所述第一处理设备的操作故障而接管执行指定将由第一处理设备执行的任务的第二处理设备的转换信息;

接口处理器, 用于响应于出现在所述组内另一个处理设备中的转换信息方面的变化而维持和更新所述转换信息;

30 操作检测器, 用于检测所述第一处理设备的操作故障; 和

故障控制器, 用于响应于所述第一处理设备的操作故障的检测来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任

务。

10. 根据权利要求9所述的系统，其中：

所述接口处理器与所述组内的其它处理设备进行通信，以便在所述独立的处理设备转换信息知识库中维持一致的转换信息。

5 11. 一种供网络处理设备组内的独立处理设备使用的系统，用于管理所述组内的设备中发生的操作故障，包括：

独立的处理设备，包括：

知识库，包含标识用于响应于所述第一处理设备的操作故障而接管执行指定将由第一处理设备执行的任务的第二当前非操作的处理设备的转换信息；

接口处理器，用于响应于下列其中至少之一来维持和更新所述转换信息：(a)所述组内的另一处理设备的操作故障的检测、和(b)低于预定阈值的所述组内的另一处理设备的可用存储器的检测；

15 操作检测器，用于检测所述第一处理设备的操作故障；和故障控制器，用于响应于所述第一处理设备的操作故障的检测来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任

12. 根据权利要求11所述的系统，其中：

20 所述转换信息包括用于响应于所述第一处理设备的操作故障而承担执行指定将由所述第一处理设备执行的任务的处理设备的优先表，并且

响应于来自所述组内的另一个处理设备的通信而动态地更新所述优先表。

25 13. 一种供网络处理设备组内的独立处理设备使用的方法，用于管理所述组内的设备中发生的操作故障，包括下列动作：

维持标识用于响应于所述第一处理设备的操作故障而接管执行第一处理设备的任务的第二处理设备的转换信息，并且响应于出现在所述组内的另一处理设备中的转换信息方面的变化来更新所述转换信息；

30 检测所述第一处理设备的操作故障；以及

响应于所述第一处理设备的操作故障的检测，来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任

14. 一种供网络处理设备组内的独立处理设备使用的方法，用于管理所述组内的设备中发生的操作故障，包括下列动作：

存储标识用于响应于所述第一处理设备的操作故障而接管执行指定将由第一处理设备执行的任务的第二处理设备的转换信息；

5 响应于出现在所述组内的另一个处理设备中的转换信息方面的变化来维护和更新所述转换信息；

检测所述第一处理设备的操作故障；以及

响应于所述第一处理设备的操作故障，来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任务。

10 15. 一种供网络处理设备组内的独立处理设备使用的方法，用于管理所述组内的设备中发生的操作故障，包括下列动作：

维持标识用于响应于所述第一处理设备的操作故障而接管执行指定将由第一处理设备执行的任务的第二当前非操作处理设备的转换信息；

15 响应于下列其中至少之一来更新所述转换信息：(a) 所述组内的另一处理设备的操作故障的检测、(b) 低于预定阈值的所述组内的另一处理设备的可用存储器的检测；

检测所述第一处理设备的操作故障；以及

20 响应于所述第一处理设备的操作故障的检测，来启动由所述第二处理设备执行指定将由所述第一处理设备执行的任务。

处理设备管理系统

本申请是 A.Monitzer 于 2003 年 11 月 6 日提交的临时申请
5 60/517,776 号的非临时申请。

技术领域

本发明涉及一种用于管理网络处理设备组内的处理设备中发生的
操作故障的系统。

10

背景技术

在各种行业（电信、保健、金融等）中，利用计算平台来向顾客
提供高利用率在线网络访问服务。这些服务的运行时间（正常运行时
间）是重要的，并且影响着顾客接受度、顾客满意度以及进行中的顾
15 客关系。典型来讲，服务水平协议（SLA）是网络服务提供商与服务顾
客之间的契约，该协议定义了服务可用（可用性）的保障时间百分率。
如果终端用户不能在已提供的用户接口上执行定义好的功能性，则该
服务就被视作为是不可用的。现有的计算网络实现方案采用失效转移
（failover）集群结构，一旦设备集群（组）中的第一处理设备发生
20 操作故障，所述失效转移集群结构就指定备份处理设备来承担第一处
理设备的功能。已知的失效转移集群结构典型地使用处理设备（网络
的节点）的静态列表（受保护对等节点列表），所述静态列表指定了
用于承担遭受操作故障的设备的功能的备份处理设备。将列表预
先配置成能确定集群内独立主动节点的备份节点的优先级。如果发生
25 主动节点故障，则集群典型地试图按照列表上的最高优先级失效转移
至第一可用的节点。

这种已知系统的一个问题就是：多个节点可能会失效到相同的备
份节点上，这进一步导致因过载的计算机资源而引发的故障。此外，
对于多个节点集群而言，现有的方法需要进行相当大量的配置操作来
30 人工地配置备份处理设备。万一两个主动节点在多个节点集群中发生
故障，而且给这两个主动节点都配置了与它们失效转移列表中的最高
优先级相同的可用备份节点，那么这两个节点就会都失效转移到这个

相同的备份节点上。这对于备份节点而言要求更高的计算机资源容量，并也增加了失效转移配置的成本。在现有的系统中，也许可以通过按照单个节点故障由用户人工再配置失效转移配置优先级列表来避免这种多个节点发生故障的情况。然而，这种操作节点集群的人工再配置不是直接了当的，而是包含了造成进一步的服务破坏的另一个主动节点故障的风险。此外，在现有的系统中，典型地将节点专用作主服务器，而将其它节点专用作从服务器。还可以把集群进一步分成更小的集群组。因此，如果由集群中的主组和从组或者独立的组共享的盘或存储器损坏时，该集群可能再也不能工作了。同样，在现有的系统中通常采用负载均衡操作来共享群集当中的设备中的工作负载，并且这包括增加风险的动态应用和复杂应用。根据本发明原理的系统提供了一种针对解决所提出的问题和缺陷的处理设备故障管理系统。

发明内容

一种根据下列因素自动地自适应修改处理设备组（集群）的备份设备的失效转移配置优先级列表以提高可用性并降低与人工配置相关的风险与成本的系统，所述因素包括例如：该组的当前负载状态、该组内设备的存储器使用率和该组内被动备份处理设备的可用性。网络处理设备组内的各个处理设备使用了一种用于管理出现在该组内的设备中的操作故障的系统。所述系统包括接口处理器，它用来维护标识用于响应于第一处理设备的操作故障而接管执行第一处理设备的任务的第二处理设备的转换信息，并且用来响应于出现在该组的另一个处理设备中的转换信息的变化而更新转换信息。操作检测器检测第一处理设备的操作故障。此外，故障控制器响应于第一处理设备的操作故障的检测来启动由第二处理设备执行指定将由第一处理设备执行的任

附图说明

图 1 示出根据本发明原理、网络处理设备组所使用的系统的框图，所述系统用于管理该组内的设备中发生的操作故障。

图 2 示出根据本发明原理、图 1 的用于管理网络处理设备组内的设备中发生的操作故障的系统所使用的过程的流程图。

图 3 示出根据本发明原理、由图 1 的系统管理的网络处理设备组的网络图。

图 4 示出根据本发明原理、由图 1 的系统管理的网络处理设备组的示例性配置。

5 图 5-9 示出根据本发明原理、举例说明如果发生设备操作故障则承担处理设备功能的备份处理设备的自动故障管理的优先表。

图 10 示出根据本发明原理、图 1 的用于管理网络处理设备组内的设备中发生的操作故障的系统的 AFC 10 所使用的过程的流程图。

10 具体实施方式

图 1 示出了包括用于管理经由通信网络 20 访问的网络处理设备组（未示出）内的处理设备（节点）中发生的操作故障的自动失效转移控制器（AFC）10 的系统的框图。所述系统实现了多个节点的分组（集群）并且提高了整体集群可用性。系统中的集群内的独立节点具有为
15 每个受保护的主动节点标识备份节点的优先级列表。所述列表包括主动节点的优先级列表，并可被称为受保护对等节点列表。在已知的现有故障系统实现方案中，受保护对等节点列表是静态的，因此在一个主动节点发生故障的情况下，故障管理系统在不依赖所发现的备份节点的当前资源利用率的情况下在优先级列表中搜索第一可用备份节点。
20 与此相反，图 1 的系统针对集群中处理设备（节点）的当前状态，自动地适应并优化受保护对等节点列表。图 1 的系统有利于在集群配置中工作的多个节点的故障管理。节点是经由通信网络（例如网络 20、LAN、内联网或因特网）连接于其它节点的单个处理设备或拓扑实体。这里所使用的处理设备包括：服务器、PC、PDA、笔记本、膝上型 PC、
25 移动式电话、机顶盒、TV 或响应于已存储的代码化机器可读指令而提供功能的任何其它设备。应当注意，在此术语“节点”和“处理设备”以及术语“集群”和“组”可互换使用。

集群是连接于群聚网络且共享某些功能的节点组。集群所提供的功能是用软件或硬件实现的。参予集群的独立节点并入了故障处理功能并且向备份节点提供了故障管理（失效转移）能力。在图 1 的系统中，独立的节点还提供支持集群管理的处理器实现的功能，包括向集
30 群添加节点以及从集群中删除节点。这里所使用的处理器是用于执行

任务的设备和/或机器可读指令集。这里所使用的处理器包括硬件、固件和/或软件中的任何一个或其组合。处理器通过操纵、分析、修改、转换或发送可执行程序或信息设备所使用的信息和/或通过把信息路由至输出设备来处理信息。例如，处理器可以使用或包括控制器或微处理器的能力。

图 1 的系统根据对集群中节点的状态改变（例如，从可用到不可用）的检测对集群结构进行重新配置（并更新备份优先级列表）。这种重新配置功能例如是利用失效转移引擎 14 来实现的，所述失效转移引擎使用网络控制器 12 和网络 20 来通知主自动失效转移控制器（AFC）配置知识库 40 及其它 AFC 集群处理设备 19 关于状态改变的情况。失效转移引擎 14 还对网络控制器 12 所转发的配置变化和同步消息作出响应，以及对心跳引擎 18 所传送的通知作出响应。响应于接收到的消息，失效转移引擎 14 启动知识库 16 中存储的本地失效转移配置的修改，并经由网络控制器 12 和网络 20 向主配置知识库 40 及其它 AFC 单元 19 传送表明已修改的配置的数据。

图 1 的系统结构提供了能管理一个组内操作设备的多重故障的稳健配置。所述系统为包含多个处理设备的组动态地优化预定的处理设备备份列表中指明的配置。所述系统易于伸缩，以便万一发生多个节点故障，能适应节点数目的增加并降低 AFC 10 与其它 AFC 19 之间所需的数据通信量。此外，如果据优先级列表表明不同备份节点都是可用的，那么一旦在节点组内发生两个节点故障，这两个节点就不会失效转移到相同的备份节点上。所述系统降低或消除了对这样的人工干预和广泛测试的需要，所述人工干预和广泛测试是为了确保在特定节点承担发生故障节点或组的的任务的操作之后其它主动节点能失效转移到不同于该特定节点的备份节点。这还减小了与修复和人工再配置相关的风险，也减少了集群配置的维护成本。

一个组内的独立节点包括自适应失效转移控制器（例如，AFC 10），它包括下面所述的提供功能和连接的各种模块。AFC 10 的失效转移引擎 14 控制和配置 AFC 10 的其它模块，包括经由配置数据访问控制器 45 来配置的心跳引擎 18、集群网络控制器 12 以及本地 AFC 配置数据知识库 16。失效转移引擎 14 还初始化、维护和更新 AFC 10 所使用的状态机，并且采用和维护包含使用参数在内的其它相关数据。这些使

用参数标识了用来执行特定计算机操作任务的资源以及在管理处理设备备份优先级列表过程中失效转移引擎 14 所使用的资源（例如，处理设备、存储器、CPU 资源、IO 资源）。将使用参数存储在本地 AFC 配置数据知识库 16 中。失效转移引擎 14 优选地利用状态和使用参数信息来优化处理设备组（例如，包括并入 AFC 10 的设备及独立地包含诸如其它 AFC 19 之类的 AFC 的其它设备）的受保护对等节点列表。失效转移引擎 14 按同步化方式从本地 AFC 配置数据知识库 16 中导出使用参数信息。此外，引擎 14 采用集群网络控制器 12 来更新一个组内的处理设备的状态和使用参数信息，这些状态和使用参数信息都保留在主 AFC 配置知识库 40 中，并且还更新保留在其它 AFC 19 的本地 AFC 配置数据知识库中的状态和使用参数信息。

失效转移引擎 14 经由失效转移心跳接口 31 向心跳引擎 18 传送包含本地 AFC 配置数据知识库 16 的数据标识更新的消息。心跳引擎利用配置数据访问控制器 45、经由通信接口 22 从本地 AFC 配置数据知识库 16 中读取配置信息。心跳引擎 18 还利用集群网络控制器 12 来建立与使用从知识库 16 中获得的配置数据的其它 AFC 19 的心跳引擎之间的通信信道。配置数据访问控制器 45 经由接口 22 来支持对知识库 16 的读写访问，并且经由接口 24 来支持与失效转移引擎 14 之间的数据通信以及经由接口 35 来支持与心跳引擎 18 之间的数据通信。为了这个目的，配置数据访问控制器 45 采用了保护数据不在知识库 16 数据修改期间遭到误用的通信仲裁协议。

集群网络控制器 12 提供了分别支持失效转移引擎 14 和心跳引擎 18 访问网络 20 的通信接口 27 和 38。控制器 12 提供了集群通信网络 20 上的双向网络连通服务，并支持从连接源到连接目的地之间的信息传送。具体地说，控制器 12 在专用网络连接（或者例如通过经由因特网动态分配的连接）上提供了从 AFC 10 到其它 AFC 19 或到主 AFC 配置知识库 40 的下列连通服务。控制器 12 支持 AFC 10 与其它节点的网络控制器（例如，其它 AFC 19 的控制器）之间的双向通信，集群网络控制器 12 是网际协议（IP）兼容的，但是也可以采用其它的协议，包括与开放系统互连（OSI）标准相兼容的协议（例如 X.25）或与内联网标准相兼容的协议。另外，集群网络控制器 12 有利地提供了网络宽同步和数据内容自动发现机制以实现集群中处理设备的知识库中的优

先级备份列表及其它信息的自动标识与更新。主 AFC 配置知识库 40 是为经由通信网络 20 网络化的处理设备提供非易失性数据存储的中央知识库。

图 2 示出图 1 的用于管理网络处理设备组内的设备中发生的操作故障的 AFC 10 所使用的过程的流程图。在起始于步骤 200 之后, AFC 10 的失效转移引擎 14 初始化并命令集群网络控制器 12 连接于集群通信网络 20。在步骤 205 中, 响应于正可访问的集群网络 20, 失效转移引擎 14 从主 AFC 配置知识库 40 中获得可用的配置信息。失效转移引擎 14 在本地 AFC 配置数据知识库 16 中存储所获得的配置信息。如果主 AFC 配置知识库 40 是不可访问的, 则在图 2 的过程的后续步骤中, 故障引擎 14 就使用从本地 AFC 配置数据知识库 16 导出的配置信息。

在步骤 210 中, 失效转移引擎 14 配置网络控制器 12 的自动发现功能以便自动检测处理设备中的其它 AFC 19 的状态和使用信息, 所述处理设备包括与经由集群通信网络 20 连接的 AFC 10 相关联的集群。失效转移引擎 14 还注册为用于从主 AFC 配置知识库 40 中获得信息、识别组内处理设备的设备状态和使用参数信息方面的变化的侦听器。当在步骤 210 中设置集群网络控制器 12 之后, 在步骤 215 中, 失效转移引擎 14 启动心跳引擎 18 的操作。心跳引擎 18 从本地 AFC 配置数据知识库 16 中获得包含受保护对等节点列表的配置信息, 并且利用集群网络控制器 12 来建立 AFC 10 与其它 AFC 19 之间的心跳通信。具体地说, 心跳引擎 18 利用集群网络控制器 12 来建立 AFC 10 与其它 AFC 19 之间的心跳通信, 其中所述其它 AFC 19 将 AFC 10 指明为其它 AFC 19 的独立受保护对等节点列表中的备份节点。心跳通信包括定期的信息交换以证实独立的对等节点仍然是工作的。如果在 AFC 10 的受保护对等节点列表中标识的节点发生故障, 那么失效转移引擎 14 就还向其它 AFC 19 及将要通知的主 AFC 配置知识库 40 进行注册。图 1 的系统有利地在步骤 215 中利用集群宽配置、同步化和发现来通知心跳引擎 18 对其它 AFC 19 的本地 AFC 配置数据知识库 16 的状态改变和更新。心跳引擎 18 还对处理设备的相关联集群中的节点的失效转移策略进行优化。

在步骤 220 中, 失效转移引擎 14 有利地利用已获得的处理设备状态和使用参数信息来更新本地 AFC 配置数据知识库 16, 并且利用集群

网络控制器 12 来使这些更新与对主 AFC 配置知识库 40 及其它 AFC 19 的更新同步化。具体地说，集群网络控制器 12 通知失效转移引擎 14 关于对主 AFC 配置知识库 40 及其它 AFC 19 进行的自动发现的更新，并且失效转移引擎 14 利用这个已获得的信息来更新本地知识库 16。同样，心跳引擎 18 通知失效转移引擎 14 受保护对等节点的可用性方面的变化，并且失效转移引擎 14 利用这个已获得的信息来更新本地知识库 16。失效转移引擎 14 使已获得的信息和通知相关，并优化本地 AFC 配置数据 16 中存储的集群宽受保护对等节点列表。图 2 的过程在步骤 230 终止。

通常在现有的系统中采用负载平衡操作来共享集群中的设备内的工作负担。为了这个目的，例如（分别地或组合地）利用已测定的 CPU（中央处理器）使用率和 IOPS（接口每秒钟操作次数）总数，来平衡从频繁使用的服务器到另一个机器的负载量。此外，在现有系统中的处理设备的集群典型地以这样一种配置进行工作，即其中节点都是主动的并且对集群的新来的负载请求被发布并跨越集群中的可用服务器平衡这些负载请求。主服务器控制跨越服务器的负载的发布和平衡。测定发布给主动节点的负载，并向主节点报告。

相反，将 AFC 10 的结构用作为主动/被动配置，在这种配置中若干主动节点接收入站负载并共享被动的失效转移节点（在没有主动节点负载平衡的情况下）。负载平衡是增加额外风险和降低设备可用性的复杂应用。从客户端设备向虚拟 IP 地址转发请求，所述虚拟 IP 地址能够经由通信网络 20 从一个物理端口被移至另一个物理端口。同已知的系统相比，专用的主单元不控制集群，并且它是根据所发布的备份节点的优先级列表来作出决策的。因此，AFC 10 中的失效转移管理是根据优先化的备份设备优先级列表来进行的。在另一个实施例中，AFC 10 的结构利用例如采用诸如 CPU 负载使用率、存储器使用率和 IOPS 总数之类参数的有效负载平衡来平衡跨越集群中主动服务器的负载。

图 3 示出了由图 1 的系统的管理的网络处理设备组的网络图。具体地说，图 3 包括主动-被动集群的网络图。主动节点 300 和 302 以及被动节点 304 和 306 都连接于客户端通信网络 60，以便向连接于这个网络的处理设备 307 和 309 提供服务和为集群内部通信提供服务。节

点 300-306 也都连接于存储系统以及相关存储区网络 311, 以便提供集群所使用的共享驱动。此外, 这些节点可以安装有相同的软件 (操作系统、应用程序等)。主动节点 (300、302) 具有与连接于客户端通信网络 60 的物理端口相关联的一个或多个虚拟 IP 地址。被动节点 (304、306) 没有与连接于客户端通信网络 60 的它们的物理端口相关联的虚拟 IP 地址。客户端设备 (307、309) 向与主动节点 300 和 302 的其中一个相关联的虚拟 IP 地址传送消息请求和数据。如果发生失效转移 (例如, 一个或多个节点 300-302 的故障), 则被动节点 (例如, 节点 304 或 306) 就获得主动节点的虚拟 IP 地址的所有权, 并把它分配给它自己的物理端口。虚拟资源失效转移到备份资源。响应于虚拟 IP 地址的分配, 被动节点变为主动的, 并且变成主动节点组。

如果发生失效转移, 则出现故障的处理设备所正在执行的那些操作或事务就会丢失。由故障设备正在执行或者即将执行的那些记录在工作日志里的工作和事务由承担有故障设备的工作的备份设备来执行 (或重新执行)。图 4 示出了由图 1 的系统管理的网络处理设备组的示例性配置结构。具体地说, 图 4 的配置结构示出了三个主动节点 (节点 1、2 和 3) 以及两个备份节点 (节点 4 和 5), 但是可以容易地将这个配置扩展成更多的备份节点。备份节点 4 或备份节点 5 都可以充当独立的主动节点 1、2 和 3 的主备份节点或次备份节点。主动节点 1、2 和 3 执行相同应用程序的拷贝并且这些节点的相应虚拟 IP 地址分配给对应节点物理端口。被动节点 4 和 5 都处于备用模式, 并且不具有分配给它们各自的物理端口的虚拟 IP 地址。

图 5-9 示出了举例说明备份处理设备的自动故障管理的优先表, 如果发生设备操作故障则所述备份处理设备就会承担图 4 的处理设备的功能。图 5 的备份优先级列表存储在每个 AFC (图 4 的 AFC 1-5) 中。图 5 表示主备份节点 4 利用心跳引擎 (例如, 图 1 的单元 18) 来监视受保护节点。具体地说, 备份节点 4 是节点 1、节点 2 和节点 5 的主备份节点。如果受监视的节点 1、2 或 5 的其中一个发生故障, 那么节点 4 就获得特定虚拟 IP 地址的所有权以及有故障节点的虚拟服务器的所有权, 并变成不可用的状态。在图 5 的备份列表中, 节点状态: A = 可用的、N = 不可用的。

在示例性操作中, 被动节点 4 经历了操作的问题。具体地说, 例

如, 如果节点 4 需要承担由发生故障的节点 1、2 或 5 中的一个正在执行的任务, 那么存储器容量的减小将会降低它承担工作负载的能力。随后, 在解决节点 4 上的问题以前, 主动节点 1 发生故障。

在现有已知(空载平衡)的系统中, 节点 1 可能不利地重复而失败地试图失效转移到由图 5 的列表表明正处于可用状态的节点 4 上。这导致相当可观的操作中断。相反, 在图 1 的系统中, 节点 4 检测它自己的存储器容量的减小, 并更新它在其备份优先级列表中的节点状态项, 如图 6 中所示。具体地说, 存储在(图 6 中所示的)节点 4 中的备份节点列表举例说明了节点 4 的节点状态项(项 600)已经变为不可用的。然而, 如在图 7 中举例说明的那样, 起初其它节点 1-3 和 5 的备份节点列表未曾接收到更新节点 4 的可用性状态的信息。

其它节点 1-3 和 5 利用自动发现法从节点 4 的 AFC 单元中获得更新后的节点 4 可用性信息。在询问连接于网络 20 的集群中其它节点的备份列表信息的过程中, 节点 1-3 和 5 的 AGC 采用了网络控制器 12(图 1)。在另一个实施例中, 节点 4 的 AFC 单元检测备份列表信息变化, 并且经由网络 20 把更新后的信息传送给节点 1-3 和 5 以及主 AFC 知识库 40。在获得并发布已更新的备份列表信息的过程中, 网络控制器 12 采用通信和路由协议以供将节点 4 的备份列表可用性信息传送给节点 1-3 和 5。为了这个目的, 网络控制器 12 采用了包含 OSPF(开放式最短路径优先)路由协议的 IP 兼容的通信协议以及与 IETF(因特网工程任务组)相兼容的协议: 例如 RFC1131、RFC1247、RFC1583、RFC1584、RFC2178、RFC2328 和 RFC2370, 来向节点 1-3 和 5 以及主 AFC 知识库 40 发布代表节点 4 的状态信息的数据。RFC(请求说明)文档是可经由因特网获得的, 并且是由因特网标准工作组来准备的。

图 8 示出了遵循由节点 1-5 各自的 AFC 接收的表示节点 4 的状态信息的数据处理的节点 1-5 的备份优先级列表, 以及它们在本知识库(比如, 知识库 16)中各自的备份优先级列表的更新。节点 1-5 的备份优先级列表示出了在哪里节点 4 被指定为主备份节点或次备份节点(图 8 中的项 800-808), 现在响应于状态变化更新将它标记为不可用的。图 8 示出了在图 8 中举例说明的节点 4 在与图 4 的系统的 5 个节点的备份方案相对应的 5 列中都是不可用的。因此, 现在节点 5 是节点 1(主不可用: 次变为主)、节点 2(主不可用: 次变为主)和

节点 3 的主备份。

节点 5 (使用诸如图 1 的单元 18 之类的心跳引擎) 检测节点 1 中的故障, 验证已出现检测到的故障, 接管将由节点 1 执行的任务, 并且更新记录在其本地知识库里的其备份列表。节点 5 中的网络控制器
5 12 按照前面描述的方式将表示节点 5 状态方面的变化 (标识到不可用状态的变化) 的数据传送给节点 1-4。通过使用先前描述的路由和通信协议, 把状态变化信息传送给节点 1-4, 以确保一致的备份列表信息。这保证了在节点 1-5 中一致地更新所述信息。图 9 示出了遵循由节点 1-5 各自的 AFC 接收的表示节点 5 的状态信息的数据的处理的节点 1-
10 5 的备份优先级列表, 以及它们在本地知识库 (例如, 知识库 16) 中各自的备份优先级列表的更新。在判断可用备份节点的过程中, 系统失效转移策略采用了集群参数 (例如, 状态和资源使用信息)。这有利地降低了失效转移状况期间的宕机时间, 还减少了人工的系统重新配置。

15 在可选的实施例中, 由独立的节点将备份优先级列表信息传送给主 AFC 知识库 40, 并且各个节点 1-5 从知识库 40 中获得备份节点列表信息。节点 1-5 的独立节点响应于状态变化或存储在独立节点本地知识库 (例如, 知识库 16) 中的备份列表信息方面的变化的检测, 在知识库 40 中存储备份列表信息。对存储在知识库 40 中的备份列表信息进行的更新是由知识库系统 40 响应于对知识库 40 中已存储的备份
20 列表信息方面的变化的检测而传送给节点 1-5 的。在另一个实施例中, 节点 1-5 的独立节点间歇地询问知识库 40 以获得更新后的备份列表信息。

图 10 示出了图 1 的用于管理网络处理设备组 (集群) 内的设备中
25 发生的操作故障的系统的 AFC 10 所使用的过程的流程图, 所述网络处理设备采用了类似的可执行软件。在步骤 702 中, 在起始于步骤 701 后, AFC 10 在内部知识库中维护标识用于响应于第一处理设备的操作故障来接管执行指定将由第一处理设备执行的任务的第二当前不操作的被动处理设备的转换信息。处理设备的操作故障包括例如软件执行
30 故障或硬件故障。所述转换信息包括处理设备的优先化备份列表, 以供响应于第一处理设备的操作故障而承担执行第一处理设备的任务。在步骤 704 中, AFC 10 响应于下列检测来更新转换信息, 包括: (a)

组内的另一个处理设备的操作故障的检测、或 (b) 低于预定阈值的组内的另一个处理设备的可用存储器的检测。在步骤 706 中, AFC 10 检测第一处理设备的操作故障。在步骤 708 中, 响应于对第一处理设备的操作故障的检测, AFC 10 启动由第二处理设备来执行指定将由第一处理设备执行的任

5 务。
在步骤 712 中, AFC 10 响应于来自组内的另一个处理设备的通信, 动态地更新内部存储的备份设备优先级列表信息, 以便在该组内独立的处理设备中维护一致的转换信息。具体地说, 响应于下列因素来动态地更新内部存储的优先级列表, 所述因素包括: 组内的另一个处理设备的检测、或低于预定阈值的组内的另一个处理设备的可用存储器的检测。所述因素还包括: (a) 超过预定阈值的组内的另一个处理设备的工作负载的操作故障的检测、(b) 超过预定阈值的组内的另一个处理设备的 CPU (中央处理器) 资源的使用的检测、或 (c) 在预定的时间周期内, 超过预定阈值的组内的另一个处理设备的许多 I/O (输入/输出) 操作的检测。此外, 还响应于由组内不同的处理设备提供的状态信息, 来动态地更新优先表, 所述状态信息表明已检测出的组内另一个处理设备的状态从可用到不可用的变化, 或已检测出的组内另一个处理设备的状态从不可用到可用的变化。为了这个目的, AFC 10 询问组内的其它处理设备以识别出现在组内另一个处理设备当中的转换信息方面的变化。图 10 的过程在步骤 718 终止。

图 1 的系统有利地根据参与集群的节点参数 (例如, 失效转移状态、资源使用) 来适应集群失效转移备份列表配置。所述系统还根据所述这些参数来优化备份节点列表, 并且根据更新后的备份节点列表适应心跳操作。所述系统利用自动发现功能来提供节点 1-5 中维持的参数

25 的集群宽自动同步, 以便检测节点 1-5 的本地知识库中所存储的备份列表信息方面的变化。
图 1-10 中展示的系统 and 过程都不是排他性的。可以依照本发明的原理推导出其它的系统和过程以实现相同的目的。尽管已经参照特定的实施例描述了本发明, 但是将要理解的是, 在此所示和所述的实施例和变形都仅仅是示例性的。本领域的技术人员在不背离本发明的范围的情况下, 可以实现对当前设计的修改。根据本发明原理的系统提供了高可用性的应用和操作系统软件。此外, 系统 10 (图 1) 提供的

任何功能都可以用硬件、软件或上述两者的组合加以实现，并且可以驻留在位于链接图 1 的元件的网络或另外链接的网络（包括另一个内联网或因特网）的任意位置上的一个或多个处理设备中。

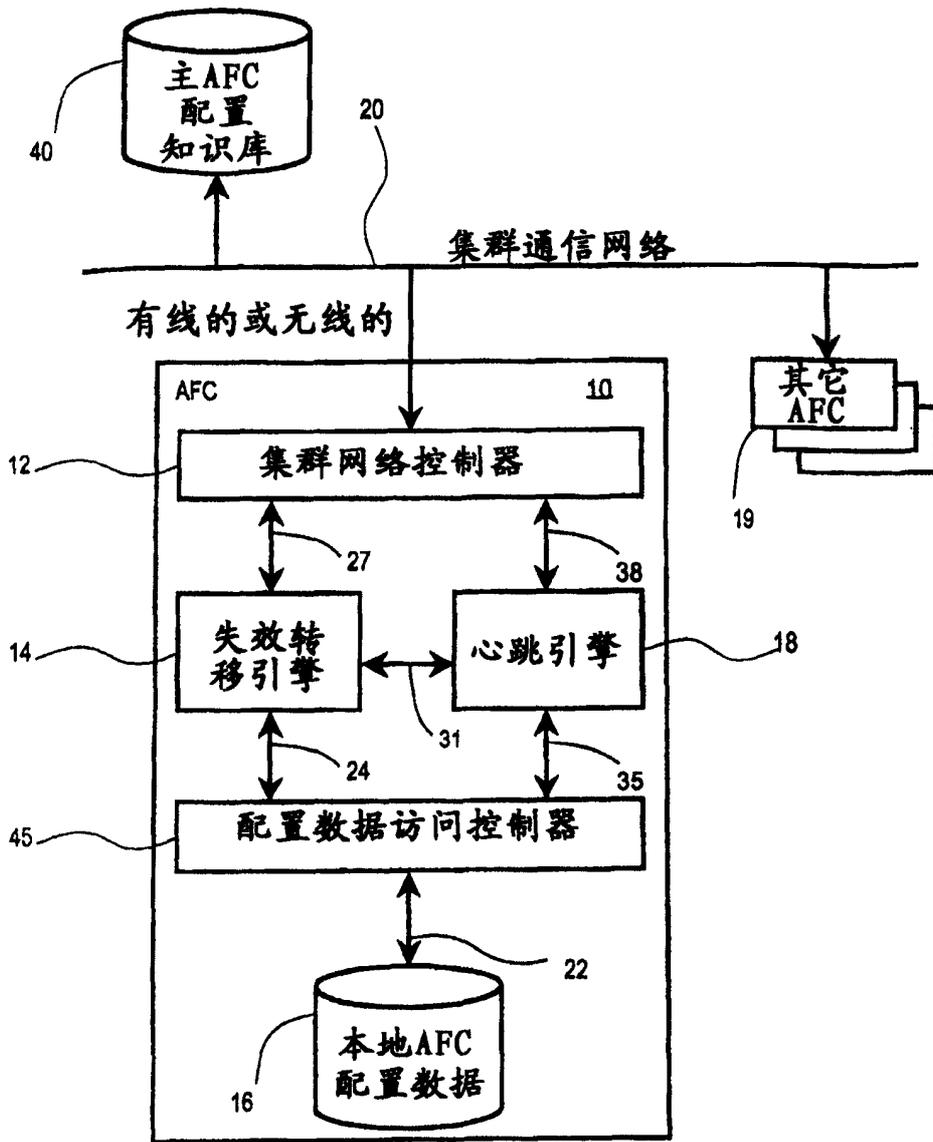


图 1

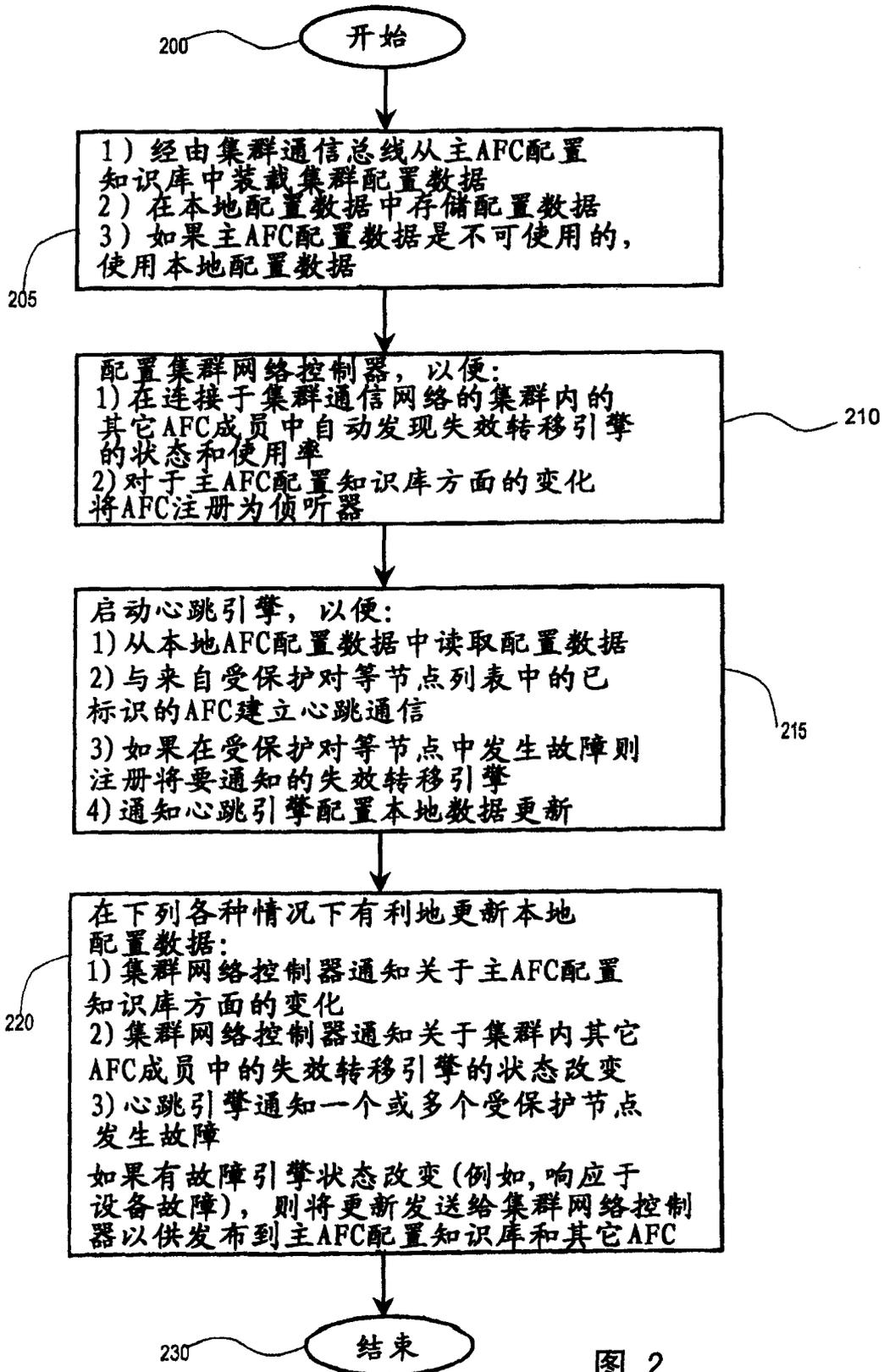


图 2

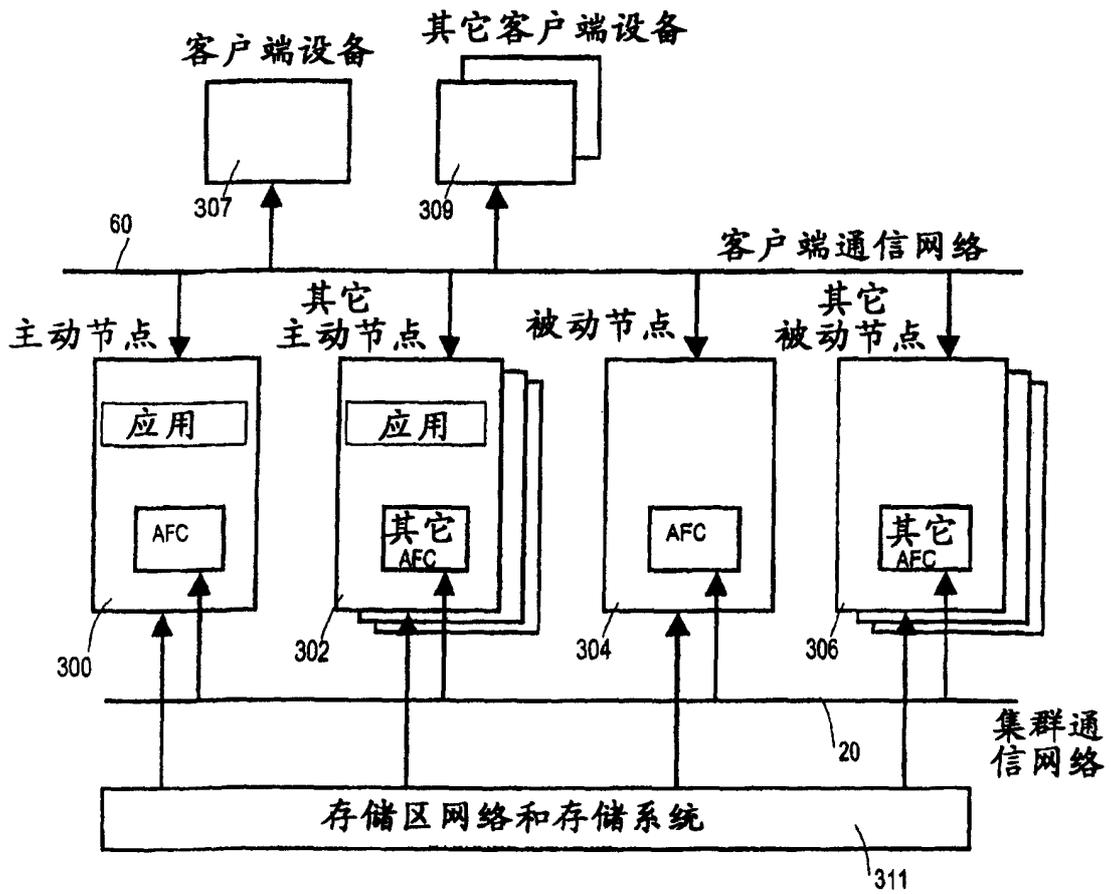


图 3

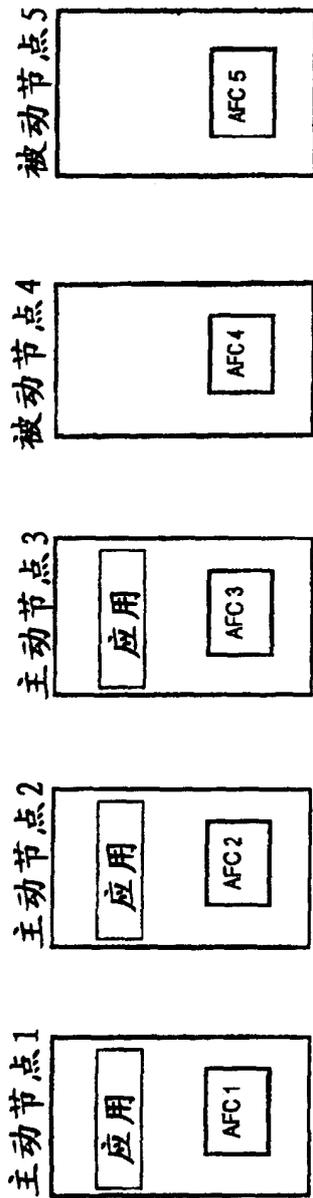


图 4

备份优先级列表(所有节点)

受保护节点	节点1	节点2	节点3	节点4	节点5
主备份	节点4	节点4	节点5	节点5	节点4
次备份	节点5	节点5	节点4	N	N
节点状态	A	A	A	A	A

图 5

备份优先级列表(节点4)

受保护节点	节点1	节点2	节点3	节点4	节点5
主备份	节点4	节点4	节点5	节点5	节点4
次备份	节点5	节点5	节点4	N	N
节点状态	A	A	A		A

600

图 6

备份优先级列表(所有其它节点)

受保护节点	节点1	节点2	节点3	节点4	节点5
主备份	节点4	节点4	节点5	节点5	节点4
次备份	节点5	节点5	节点4	N	N
节点状态	A	A	A	A	A

图 7

备份优先级列表(所有节点)

受保护节点	节点1	节点2	节点3	节点4	节点5
主备份			节点5	节点5	
次备份	节点5	节点5		N	N
节点状态	A	A	A		A

800 802 804 806 808

图 8

备份优先级列表(所有节点)

受保护节点	节点1	节点2	节点3	节点4	节点5
主备份	N	N			N
次备份			N	N	N
节点状态		A	A	N	

图 9

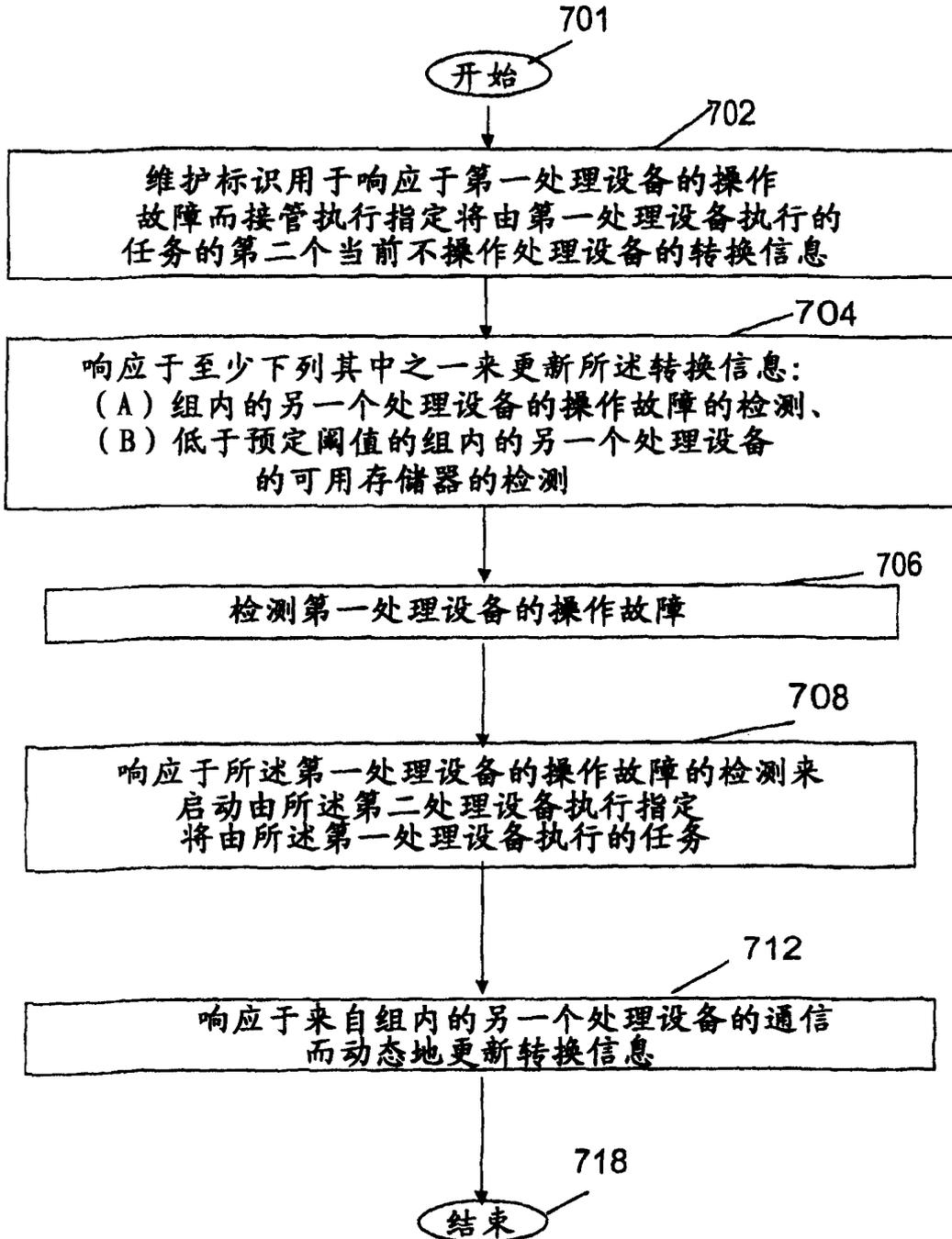


图 10