(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110912780 B
(45)授权公告日 2021.08.27

(21)申请号 201911281240.2

(22)申请日 2019.12.13

(65)同一申请的已公布的文献号
申请公布号 CN 110912780 A

(43)申请公布日 2020.03.24

(73)专利权人 华云数据控股集团有限公司
地址 214000 江苏省无锡市滨湖区科教软件园6号

(72)发明人 过育红 朱正东 仇大玉 张银滨

(74)专利代理机构 苏州友佳知识产权代理事务所(普通合伙) 32351
代理人 储振

(51)Int.Cl.
H04L 12/26(2006.01)
H04L 12/721(2013.01)
H04L 12/931(2013.01)
H04L 12/947(2013.01)
H04L 29/08(2006.01)

审查员 牛爽

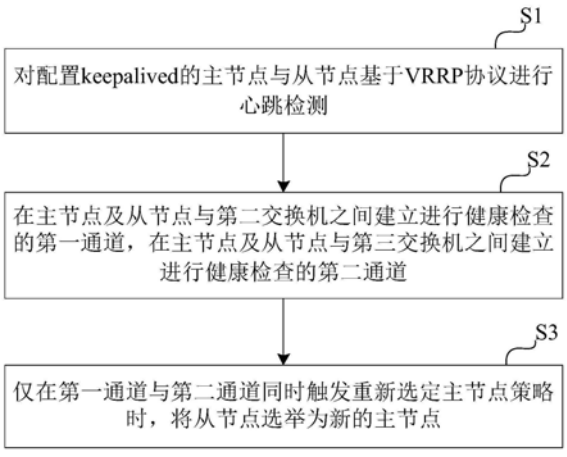
权利要求书2页 说明书9页 附图4页

(54)发明名称

一种高可用集群检测方法、系统及受控终端

(57)摘要

本发明提供了一种高可用集群检测方法,以及基于该方法的一种高可用集群检测系统及受控终端,该高可用集群检测方法,对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测,在主节点及从节点与第二交换机之间建立进行健康检查的第一通道,在主节点及从节点与第三交换机之间建立进行健康检查的第二通道,仅在第一通道与第二通道同时触发重新选定主节点策略时,将从节点选举为新的主节点。通过本申请所揭示的一种高可用集群检测方法、高可用集群检测系统,显著地改善了现有的主从节点之间的keepalived心跳检测机制,避免了因主节点由于业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象,确保了高可用集群的可靠性与服务的高可用性。



CN 110912780 B

1.一种高可用集群检测方法,用于区分主节点所发生的实质性宕机与非实质性宕机,其特征在于,

对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测,

在主节点及从节点与第二交换机之间建立进行健康检查的第一通道,在主节点及从节点与第三交换机之间建立进行健康检查的第二通道,

仅在第一通道与第二通道同时触发重新选定主节点策略时,将从节点选举为新的主节点;

在主节点及从节点与第二交换机之间建立进行健康检查的第一通道,在主节点及从节点与第三交换机之间建立进行健康检查的第二通道,

在主节点与从节点之间建立BFD会话,并仅在第一通道与第二健通道同时触发重新选定主节点策略时,将从节点选举为新的主节点。

2.根据权利要求1所述的高可用集群检测方法,其特征在于,

对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测。

3.根据权利要求1或者2所述的高可用集群检测方法,其特征在于,所述主节点与从节点均通过第二交换机及第三交换机与高可用集群中的集群节点建立会话;

所述第一通道为主节点及从节点与控制节点、计算节点、网络节点或者存储节点中的至少一个节点通过第二交换机所建立的会话通道,

所述第二通道为主节点及从节点与控制节点、计算节点、网络节点或者存储节点中的至少一个节点通过第三交换机所建立的会话通道。

4.根据权利要求3所述的高可用集群检测方法,其特征在于,还包括:

复用对已经建立会话的主节点与从节点之间发送的基于VRRP协议的报文中所包含的认证数据字段,以确定第一通道与第二通道是否同时触发重新选定主节点策略,以在第一通道与第二通道同时触发重新选定主节点策略时,以在从多个从节点中所确定新的主节点与主节点之间建立BFD会话。

5.根据权利要求4所述的高可用集群检测方法,其特征在于,所述重新选定主节点策略由优先级及权重值共同描述,以从多个从节点中确定新的主节点。

6.根据权利要求3所述的高可用集群检测方法,其特征在于,所述健康检查包括TCP检测、HTTP检测、检查脚本检测、超时检测或者负载检测。

7.根据权利要求3所述的高可用集群检测方法,其特征在于,还包括:

在多个从节点中所确定新的主节点后,将新的主节点的状态信息同步至从节点,并将虚拟IP漂移至新的主节点;

将新的主节点的状态信息同步配置至挂载至第三交换机的集群节点中;

其中,

所述集群节点包括控制节点、计算节点、网络节点与存储节点。

8.一种高可用集群检测系统,其特征在于,用于区分主节点所发生的实质性宕机与非实质性宕机,包括:

心跳检测单元,用以对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测;

第一健康检查单元,对由主节点及从节点与第二交换机之间建立的第一通道进行健康

检查；

第二健康检查单元，对由主节点及从节点与第三交换机之间建立的第二通道进行健康检查；

决策单元，仅在第一通道与第二通道同时触发重新选定主节点策略时，将同时触发选定主节点策略的从节点选举为新的主节点；

在主节点及从节点与第二交换机之间建立进行健康检查的第一通道，在主节点及从节点与第三交换机之间建立进行健康检查的第二通道，

在主节点与从节点之间建立BFD会话，并仅在第一通道与第二健通道同时触发重新选定主节点策略时，将从节点选举为新的主节点。

9.根据权利要求8所述的高可用集群检测系统，其特征在于，所述高可用集群检测系统运行于Zookeeper集群中。

10.一种受控终端，其特征在于，包括：处理器，存储装置，以及在处理器与存储装置之间建立通信连接的通信总线；

所述处理器用于执行存储装置中存储的一个或者多个程序，以实现如权利要求1至8中任一项所述的高可用集群检测方法。

一种高可用集群检测方法、系统及受控终端

技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及一种高可用集群检测方法、系统及受控终端。

背景技术

[0002] 随着互联网的快速发展，用户的业务量的不断增加，对业务的可靠性和性能要求越来越高。为满足用户的需求，在实际的应用环境中往往会采用HA(High Availability)集群来实现业务的处理。在高可用集群中，需要各节点之间协同一致来保证集群对业务处理的有效性。如果集群中某个节点出现问题，就会影响到整个集群的工作性能，因此需要集群具有能够快速对问题节点进行处理的功能，从而保证集群的可靠性和对业务处理的有效性。

[0003] 在高可用集群中通常包含一个主节点(Master)和多个从节点(Backup)，主节点与多个从节点之间通常基于Keepalived及Haproxy的组合，以确保集群的高可用性能。Keepalived是以VRRP协议(虚拟路由冗余协议)为基础实现的。主节点与各从节点之间通过心跳机制维持状态。当从节点无法接收到主节点发送的VRRP控制报文时，则认为主节点已经宕机。在此场景中则根据VRRP协议的优先级从多个从节点中选举出一个从节点并作为新的主节点。新的主节点启动资源管理模块以接管原来的主节点上运行的资源、服务或者进程。

[0004] 目前，对主从节点之间进行心跳检测的现有技术中，检测不到主节点的心跳的原因并非是主节点已经宕机，也存在主节点因主节点繁忙或者检测超时等诸多原因。如果一旦检测不到主节点的心跳就盲目的切换主节点，则会导致出现脑裂现象。脑裂(split-brain)是指在一个高可用(High Availability,HA)系统中，当联系着的两个节点断开联系时，本来为一个整体的系统，分裂为两个独立节点，这时两个节点开始争抢共享资源，从而导致系统混乱、数据损坏的现象。

[0005] 同时，公开号为CN109286525A的中国发明专利公开了一种基于MQTT通讯和主备之间心跳的双机备份方法。但是上述基于MQTT协议的心跳检测的现有技术存在以下缺陷：(1)MQTT协议没有齐备的SDK，不同的异构终端，需要有对应的与MQTT服务器通信的软件SDK包；(2)MQTT协议不支持负载均衡，无法有效防止高并发和恶意攻击；(3)不支持用户管理接口、不支持点对点通信、不支持群通信和群管理、不支持离线消息；(4)由于需要配置MQTT服务器，因此不仅增加了集群在拓扑逻辑上的复杂性，增加了集群搭建成本，也增加了后期对集群维护的难度。

[0006] 有鉴于此，有必要对现有技术中的对高可用集群的检测方法等诸多方面予以改进，以解决上述问题。

发明内容

[0007] 本发明的目的在于揭示一种高可用集群检测方法、系统及受控终端，用以解决现

有技术所存在的缺陷，尤其是为了解决基于传统的keepalived心跳检测机制中由于主节点因业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象，解决由此所导致的整个集群中发生脑裂的技术问题，确保高可用集群的可靠性与服务的高可用性。

[0008] 为实现上述一个发明目的，本申请首先提供了一种高可用集群检测方法，

[0009] 对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测，

[0010] 在主节点及从节点与第二交换机之间建立进行健康检查的第一通道，在主节点及从节点与第三交换机之间建立进行健康检查的第二通道，

[0011] 仅在第一通道与第二通道同时触发重新选定主节点策略时，将从节点选举为新的主节点。

[0012] 作为本发明的进一步改进，

[0013] 对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测，

[0014] 在主节点及从节点与第二交换机之间建立进行健康检查的第一通道，在主节点及从节点与第三交换机之间建立进行健康检查的第二通道，

[0015] 在主节点与从节点之间建立BFD会话，并仅在第一通道与第二健通道同时触发重新选定主节点策略时，将从节点选举为新的主节点。

[0016] 作为本发明的进一步改进，所述主节点与从节点均通过第二交换机及第三交换机与高可用集群中的集群节点建立会话；

[0017] 所述第一通道为主节点及从节点与控制节点、计算节点、网络节点或者存储节点中的至少一个节点通过第二交换机所建立的会话通道，

[0018] 所述第二通道为主节点及从节点与控制节点、计算节点、网络节点或者存储节点中的至少一个节点通过第三交换机所建立的会话通道。

[0019] 作为本发明的进一步改进，还包括：

[0020] 复用对已经建立会话的主节点与从节点之间发送的基于VRRP协议的报文中所包含的认证数据字段，以确定第一通道与第二通道是否同时触发重新选定主节点策略，以在第一通道与第二通道同时触发重新选定主节点策略时，以在从多个从节点中所确定新的主节点与主节点之间建立BFD会话。

[0021] 作为本发明的进一步改进，所述重新选定主节点策略由优先级及权重值共同描述，以从多个从节点中确定新的主节点。

[0022] 作为本发明的进一步改进，所述健康检查包括TCP检测、HTTP检测、检查脚本检测、超时检测或者负载检测。

[0023] 作为本发明的进一步改进，还包括：

[0024] 在多个从节点中所确定新的主节点后，将新的主节点的状态信息同步至从节点，并将虚拟IP漂移至新的主节点；

[0025] 将新的主节点的状态信息同步配置至挂载至第三交换机的集群节点中；

[0026] 其中，

[0027] 所述集群节点包括控制节点、计算节点、网络节点与存储节点。

[0028] 基于相同发明思想，本申请还提供给了一种高可用集群检测系统，其特征在于，包括：

[0029] 心跳检测单元，用以对配置keepalived的主节点与从节点基于VRRP协议进行心跳

检测；

[0030] 第一健康检查单元，对由主节点及从节点与第二交换机之间建立的第一通道进行健康检查；

[0031] 第二健康检查单元，对由主节点及从节点与第三交换机之间建立的第二通道进行健康检查；

[0032] 决策单元，仅在第一通道与第二通道同时触发重新选定主节点策略时，将同时触发选定主节点策略的从节点选举为新的主节点。

[0033] 作为本发明的进一步改进，所述高可用集群检测系统运行于Zookeeper集群中。

[0034] 最后，本申请还提供了一种受控终端，包括：处理器，存储装置，以及在处理器与存储装置之间建立通信连接的通信总线；

[0035] 所述处理器用于执行存储装置中存储的一个或者多个程序，以实现如上述任一项发明所揭示的高可用集群检测方法。

[0036] 与现有技术相比，本发明的有益效果是：

[0037] 通过本申请所揭示的一种高可用集群检测方法、高可用集群检测系统，显著地改善了现有的主从节点之间的keepalived心跳检测机制，避免了因主节点由于业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象，有效地避免了高可用集群出现脑裂现象，确保了高可用集群的可靠性与服务的高可用性。

附图说明

[0038] 图1为本发明一种高可用集群检测方法的整体流程图；

[0039] 图2为应用本发明一种高可用集群检测方法的高可用集群通过第一通道与第二通道进行健康检查以确定是否发生主从切换在第一种实例中的示意图；

[0040] 图3为应用本发明一种高可用集群检测方法的高可用集群通过第一通道与第二通道进行健康检查以确定是否发生主从切换在第二种实例中的示意图；

[0041] 图4为应用本发明一种高可用集群检测方法的高可用集群通过第一通道与第二通道进行健康检查以确定是否发生主从切换在第三种实例中的示意图；

[0042] 图5为本发明一种高可用集群检测系统的拓扑图；

[0043] 图6为应用高可用集群检测方法的一种受控终端的拓扑图。

具体实施方式

[0044] 下面结合附图所示的各实施方式对本发明进行详细说明，但应当说明的是，这些实施方式并非对本发明的限制，本领域普通技术人员根据这些实施方式所作的功能、方法、或者结构上的等效变换或替代，均属于本发明的保护范围之内。

[0045] 在详细阐述本实施例之前，对本申请各个实施例所涉及的技术术语予以必要解释与定义。在本申请中，术语“集群”的含义为：通过局域网、广域网或者其他通信方式予以连接，通过一组松散集成的计算机软件或硬件连接起来高度紧密地协作完成计算工作所形成的计算机系统。同时，在本申请中“集群”与术语“计算机集群”或者术语“数据中心”具等同技术含义。术语“主从切换”的含义为：角色为主节点与角色为从节点之间的角色切换；并且在本申请中，术语“主服务器”与术语“主节点”之间具等同技术含义，术语“从服务器”与术

语“从节点”之间具等同技术含义;术语“实质性宕机”与术语“非实质性宕机”互为相反技术概念,其中,“非实质性宕机”是指因业务繁忙或者检测超时等原因所导致的主节点与一个或者多个从节点之间无法收发VRRP控制报文,并由此引发主节点被认定为“Fail”的状态。最后,在本申请中,如无特殊说明,术语“报文”特指基于VRRP协议所形成的VRRP控制报文。本申请所揭示的一种高可用集群检测方法、系统及受控终端通过下述实施例予以具体阐述。

[0046] 实施例一:

[0047] 结合图1至图4所示,本实施例揭示了一种高可用集群检测方法(以下简称“方法”)。该方法应用于计算机集群、数据中心(IDC)、云计算平台中,尤其是适用于高可用集群场景中。在本实施例中,集群的高可用性的目标之一是消除基础架构中的单点故障。单点故障是技术堆栈的一个组件,如果它变得不可用(Fail),将导致服务中断。在高可用集群中,允许灵活IP地址重映射的系统,例如浮动IP(Floating IP)。按需IP地址重新映射通过提供可在需要时轻松重新映射的静态IP地址,消除了DNS更改中固有的传播和缓存问题。域名可以保持与相同的IP地址关联,而IP地址本身也可以在服务器之间移动。

[0048] 参图1所示,该高可用集群检测方法,包括如下步骤:

[0049] 首先,执行步骤S1、对配置Keepalived的主节点与从节点基于VRRP协议进行心跳检测。本实施例所揭示的方法应用于如图2所示出的高可用集群中。该高可用集群中包含主节点A、从节点B,以及控制节点21、计算节点22、网络节点23与存储节点24(下文中涉及的诸如控制节点21、计算节点22、网络节点23级存储节点24与术语“功能性节点”具等同技术含义)。需要说明的是,该高可用集群中上述功能性节点仅作为一种范例,在实际配置中计算节点22、存储节点24的数量可以为多个,并形成分布式计算架构;尤其的,存储节点24可被配置为Ceph分布式存储架构、DAS存储架构、网络连接存储(NAS)或者存储区域网络(SAN)等架构形式。进一步的,还可将多个高可用集群通过局域网形成一个容错能力更强的高可用集群。因此,在图2中所示出的控制节点21、计算节点22、网络节点23与存储节点24的数量可为一个,也可为多个。

[0050] 控制节点21、计算节点22、网络节点23与存储节点24均耦连至第一交换机10,以通过该第一交换机10起到联通作用。第三交换机30连接路由器40,并由路由器40接入互联网50。路由器40通过其内置的浮动IP机制进行浮动IP转换,以对用户发起的访问请求进行响应。

[0051] 高可用性(HA)是指提供在本地系统单个组件故障情况下,能继续访问服务的能力,无论这个故障是业务流程、物理设施、IT软/硬件的故障。高可用集群使用Keepalived和HAproxy的组合,在主节点A与一个或者多个从节点B之间基于Keepalived心跳检测机制以确定是否需要执行故障切换、自动扩容及主从切换等操作,从而为用户(User)提供不间断的高质量的服务/响应。Keepalived心跳检测机制采用的VRRP协议,使用一个共有的VIP(虚拟IP)实现在两台(或多台)HAproxy节点上来回“漂移”,这样对外只体现了一个IP。

[0052] Keepalived是以VRRP(Virtual Router Redundancy Protocol,即虚拟路由冗余协议)为基础。VRRP协议是实现高可用的协议,即将N台提供相同功能的设备组成一个高可用集群。高可用集群里面有一个主节点A和一个或者多个从节点B,多个从节点B构成一个从节点集群40(参图3所示)。主节点A存在一个对用户(User)提供服务的VIP。主节点A和从节

点B之间使用心跳机制来维持状态，当从节点B收不到VRRP包时就认为主节点A发生宕机，这时就需要根据VRRP控制报文的优先级(Priority)从图3中的从节点集群40中选举出一个从节点，以充当新的主节点A。新的主节点A启动资源接管模块(Pacemaker集群管理服务)来接管运行在发生实质性宕机的主节点A所配置的资源或者服务。本实施例所揭示的方法，能够区分主节点A所发生的实质性宕机与非实质性宕机，以防止主节点A在因为业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象。需要说明的是，在本申请各个实施例中，术语“主从切换”是指将原本角色定义为从节点(Slave)的服务器和/或数据库定义为主节点(Master)，从而提高该高可用集群的稳定性与健壮性。

[0053] 然后，执行步骤S2、在主节点及从节点与第二交换机之间建立进行健康检查的第一通道，在主节点及从节点与第三交换机之间建立进行健康检查的第二通道。需要说明的是，在本实施例中，所谓的第一通道与第二通道均是泛指一类通道。

[0054] 主节点A与从节点B均通过第二交换机20及第三交换机30与高可用集群中的集群节点建立会话。从而根据建立的会话确定第一通道与第二通道的数量。第一通道为主节点A及从节点B与控制节点21、计算节点22、网络节点23或者存储节点24中的至少一个节点通过第二交换机20所建立的会话通道。第二通道为主节点A及从节点B与控制节点21、计算节点22、网络节点23或者存储节点24中的至少一个功能性节点通过第三交换机30所建立的会话通道。

[0055] 具体而言，第一通道可以是图2中的主节点A与控制节点21及从节点B通过第二交换机20并基于VRRP协议所形成的VRRP控制报文的传输通道；该第一通道还可以是图2中的主节点A与控制节点21、计算节点22及从节点B通过第二交换机20并基于VRRP协议所形成的VRRP控制报文的传输通道；该第一通道还可以是图2中的主节点A与控制节点21、计算节点22、网络节点23、存储节点24及从节点B通过第二交换机20并基于VRRP协议所形成的VRRP控制报文的传输通道；该第一通道还可以是图2中的主节点A与从节点B，以及一个或者几个功能性节点(即控制节点21、计算节点22、网络节点23、存储节点24)通过第二交换机20，并基于VRRP协议所形成的VRRP控制报文的传输通道。

[0056] 同理所述，第二通道可以是图2中的主节点A与控制节点21及从节点B通过第三交换机30并基于VRRP协议所形成的VRRP控制报文的传输通道；该第二通道还可以是图2中的主节点A与控制节点21、计算节点22及从节点B通过第三交换机30并基于VRRP协议所形成的VRRP控制报文的传输通道；该第二通道还可以是图2中的主节点A与控制节点21、计算节点22、网络节点23、存储节点24及从节点B通过第三交换机30并基于VRRP协议所形成的VRRP控制报文的传输通道；该第二通道还可以是图2中的主节点A与从节点B，以及一个或者几个功能性节点(即图2所示出的控制节点21、计算节点22、网络节点23、存储节点24)通过第三交换机30，并基于VRRP协议所形成的VRRP控制报文的传输通道。

[0057] 由此可见，在本实施例中，第一通道与第二通道可分别形成为一条VRRP控制报文传输通道，也可形成为多条VRRP控制报文传输通道。

[0058] 优选的，在本实施例中，该方法的步骤S2中还包括：复用对已经建立会话的主节点A与从节点B之间发送的基于VRRP协议的报文中所包含的认证数据字段，以确定第一通道与第二通道是否同时触发重新选定主节点策略，以在第一通道与第二通道同时触发重新选定主节点策略时，以在从多个从节点中所确定新的主节点与主节点之间建立BFD会话。然后，

在建立BFD会话之后，并仅在第一通道与第二健通道同时触发重新选定主节点策略时，将从节点选举为新的主节点。选出的新的主节点从图3中的从节点集群40中所包含的多个节点B中按照重新选定主节点策略时选定新的主节点。

[0059] 尤其的，在本实施例所揭示的方法中，启用第二通道对高可用集群进行健康检查，防止仅通过第一通道进行健康检查所存在的由于主节点A所存在的业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象，有效地避免了高可用集群出现脑裂现象，从而确保了高可用集群的可靠性与服务的高可用性。本实施例所揭示的高可用集群检测方法中所披露的VRRP控制报文的格式如下所示。

[0060]

Version	Type	Virtual Rtr ID	Priority	Count IP Addrs
Auth Type		Adver Int	Checksum	
IP Address(1)				
…				
IP Address (n)				
Authentication Data (1)				
Authentication Data (2)				

[0061] 在主节点A和从节点B交互VRRP控制报文时，会将第一通道与第第二通道的VIP互相发布至各个节点(参图3中从节点集群40中所包含的多个从节点B)。由此复用VRRP控制报文中既有的Authentication Data字段(即，Authentication Data(1)与Authentication Data(2))。Authentication Data字段是用于RFC2338向后兼容，目前已被废弃，当前在发送VRRP报文时该字段被置为0，在接收VRRP控制报文时该Authentication Data字段被忽略。由此即可复用该Authentication Data字段，用以发布各自的健康检查IP，各节点在接收到带有健康检查IP的VRRP控制报文后会记录此IP地址。Authentication Data字段在本实施例中与术语“认证数据字段”具等同含义。

[0062] BFD(Bidirectional Forwarding Detection，双向转发检测)协议提供一种轻负载、快速检测两台邻接路由器/交换机之间转发路径连通状态的方法，它是一个简单的“Hello”协议。BFD协议通过三次握手机制，能提供链路来回两个方向的连通性检测。一对系统在它们之间的所建立会话通道上周期性的发送VRRP控制报文，如果某个系统在足够长的时间内没有收到对端的VRRP控制报文，则认为在这条到相邻系统的双向通道的某个部分发生了故障协议邻居通过该方式可以快速检测到转发路径的连通故障，加快启用备份转发路径，提升现有网络性能。BFD可用于检测任何形式的路径，包括直接相连的物理链路、虚电路、隧道、MPLS协议中的LSP乃至多跳的路由通道。甚至对于单向链路(如MPLS TE隧道)，只要有回来的路径，都可以检测。

[0063] BFD协议提供的检测机制与所应用的接口介质类型、封装格式、以及关联的上层协议如OSPF、BGP、RIP等无关。BFD协议在两台路由器(即图2至图4中的第二交换机20与第三交换机30)之间建立会话，通过快速发送检测故障消息给正在运行的路由协议，以触发路由协议重新计算路由表，大大减少整个网络的收敛时间。BFD协议本身没有发现邻居的能力，需

要上层协议通知与哪个邻居建立会话。

[0064] 同时，在本实施例中，所述重新选定主节点策略由优先级及权重值共同描述，以从多个从节点中确定新的主节点。具体的，该健康检查包括TCP检测、HTTP检测、检查脚本检测、超时检测或者负载检测。

[0065] 本实施例所揭示的该高可用集群检测方法还包括：在多个从节点中所确定新的主节点后，将新的主节点的状态信息同步至从节点，并将虚拟IP漂移至新的主节点。将新的主节点的状态信息同步配置至挂载至第三交换机30的集群节点中；其中，所述集群节点包括控制节点21、计算节点22、网络节点23与存储节点24。

[0066] 最后，执行步骤S3、仅在第一通道与第二通道同时触发重新选定主节点策略时，将从节点选举为新的主节点。

[0067] 当高可用集群在运行时，必定在某个时刻已经产生了一个主节点A，以及至少一个从节点B。基于主节点A及从节点B与第二交换机20之间建立进行健康检查的第一通道，在主节点A及从节点B与第三交换机30之间建立进行健康检查的第二通道，能够在对端无法接收到VRRP控制报文时，确定是否需要选举新的主节点。基于主节点A与从节点B之间所建立的BFD会话，以在第一通道与第二通道同时触发选定新的主节点的选举策略时，才认定当前状态中的主节点A已经发生实质性宕机，从而将服务和/或数据迁移至根据选举策略所确定的新的主节点中；在此过程中，可将从节点集群40中角色为Slave的一个从节点B1(或者从节点B2)从Backup状态切换为Master状态，以完成主从切换操作。

[0068] 在本实施例中，主节点的选举策略具体如下所述。

[0069] 主节点的选举策略由优先级(Priority)及权重值(Weight)共同描述。配合参照图4与图5所示，在本实施例中，主节点A与从节点集群40中配置的多个从节点(即从节点B1与从节点B2)中均配置有一个初始优先级，并由配置文件中的优先级配置项确定。初始状态中，主节点A的初始优先级高于任何一个从节点的初始优先级。Keepalived根据vrrp_script的权重值(Weight)设定，在当需要进行主从切换并从该从节点集群40中确定一个从节点以定义为主节点时或者对主节点A与多个从节点的优先级进行调整，以对多个从节点的初始优先级执行增加或者降低操作，具体规则下述规则(1)～规则(3)顺序执行。

[0070] 规则(1)：当权重值(Weight)大于0时，vrrp_script执行返回0(成功)时优先级为Priority+Weight，否则为Priority。当作为后备节点的从节点(例如从节点B1)发现该从节点B1优先级大于主节点A通告的优先级时，进行主从切换。

[0071] 规则(2)：当权重值(Weight)小于0时，vrrp_script执行返回非0(失败)时优先级为Priority+Weight，否则为Priority。当作为后备节点的从节点(例如从节点B1)发现该从节点B1的优先级大于主节点A通告的优先级时，进行主从切换。

[0072] 规则(3)：当两个从节点(例如从节点集群40中所包含的从节点B1与从节点B2)的优先级相同时，以从节点发送VRRP通告的IP作为比较对象，IP较大者选举为新的主节点。VRRP优先级的取值范围为0到255，数值越大表明优先级越高。

[0073] 存储节点24通过FC协议或者iSCSI协议连接并映射至控制节点21和计算节点22，控制节点21与计算节点22之间通过第一交换机10耦合连接，以实现上述各个功能性节点之间的数据和/或报文的转发与互通。控制节点21通过pvcreate命令和vgcreate命令共同将存储节点24创建为卷组(VG)。在控制节点21和计算节点22上分别部署Pacemaker集群管理

服务，多个Pacemaker集群管理服务共同构成Pacemaker集群。在计算节点22上通过计算服务将逻辑卷(LV)挂载至虚拟机(VM)。Pacemaker集群管理服务将逻辑卷(LV)的元数据(Metadata)对其他任意一个控制节点21和/或计算节点22执行同步更新操作。该同步更新操作能够进一步改善因既有的主节点A的角色被剥夺后所导致对控制节点21出现性能瓶颈的缺陷，并实现对逻辑卷的集群化管理，并使得整个高可用集群中的资源状态信息进行同步更新，以确保整个高可用集群的高可用性与稳定性。

[0074] 通过本实施例所揭示的一种高可用集群检测方法，显著地改善了现有高可用集群中主从节点之间的keepalived心跳检测机制，有效地避免了因主节点A由于业务繁忙或者检测超时等非实质性宕机所引发的主从切换现象，有效地避免了高可用集群出现脑裂，从而保证了高可用集群的可靠性与服务的高可用性与稳定性。

[0075] 实施例二：

[0076] 参图5所示，本实施例基于实施例一所揭示的一种高可用集群检测方法的技术方案，还揭示了一种高可用集群检测系统100(以下简称“系统”)。

[0077] 该系统100，包括：心跳检测单元31，用以对配置keepalived的主节点A与从节点B基于VRRP协议进行心跳检测。第一健康检查单元32，对由主节点A及从节点B与第二交换机20之间建立的第一通道进行健康检查。第二健康检查单元33，对由主节点A及从节点B与第三交换机30之间建立的第二通道进行健康检查。决策单元34，仅在第一通道与第二通道同时触发重新选定主节点策略时，将同时触发选定主节点策略的从节点选举为新的主节点，即将图5中从节点集群40中所包含的从节点B1与从节点B2根据实施例一所揭示的高可用集群检测方法所揭示的技术方案，将从而将主节点A的服务迁移至根据选举策略所确定的新的主节点(例如从节点B1或者从节点B2)中；在此过程中，可将从节点集群40中的一个从节点B1从Backup状态切换为Master状态以将其角色定义为主节点，以完成主从切换操作。

[0078] 同时，本实施例中的高可用集群检测系统100运行于Zookeeper集群或者其他等同类型的分布式集群系统中。重新选定主节点策略由优先级及权重值共同描述，以从多个从节点中确定新的主节点。

[0079] 本实施例所揭示的系统100与实施例一中相同部分的技术方案参实施例一所述，在此不再赘述。

[0080] 实施例三：

[0081] 参图6所示，本实施例揭示了一种受控终端200，该受控终端200包括：处理器51，存储装置52，以及在处理器51与存储装置52之间建立通信连接的通信总线53。处理器51用于执行存储装置52中存储的一个或者多个程序，以实现如实施例一所述的高可用集群检测方法。存储装置52中包含多个存储单元，即存储单元521至存储单元52i，其中，参数i取大于或者等于2的正整数。该受控终端200可被视为计算机、数据中心、裸金属服务器或者便携式电子设备等。本实施例所揭示的受控终端200与实施例一和/或实施例二中相同部分的技术方案参上文实施例一和/或实施例二所述，在此不再赘述。

[0082] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0083] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用

时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)或处理器(processor)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0084] 上文所列出的一系列的详细说明仅仅是针对本发明的可行性实施方式的具体说明,它们并非用以限制本发明的保护范围,凡未脱离本发明技艺精神所作的等效实施方式或变更均应包含在本发明的保护范围之内。

[0085] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化囊括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。

[0086] 此外,应当理解,虽然本说明书按照实施方式加以描述,但并非每个实施方式仅包含一个独立的技术方案,说明书的这种叙述方式仅仅是为清楚起见,本领域技术人员应当将说明书作为一个整体,各实施例中的技术方案也可以经适当组合,形成本领域技术人员可以理解的其他实施方式。

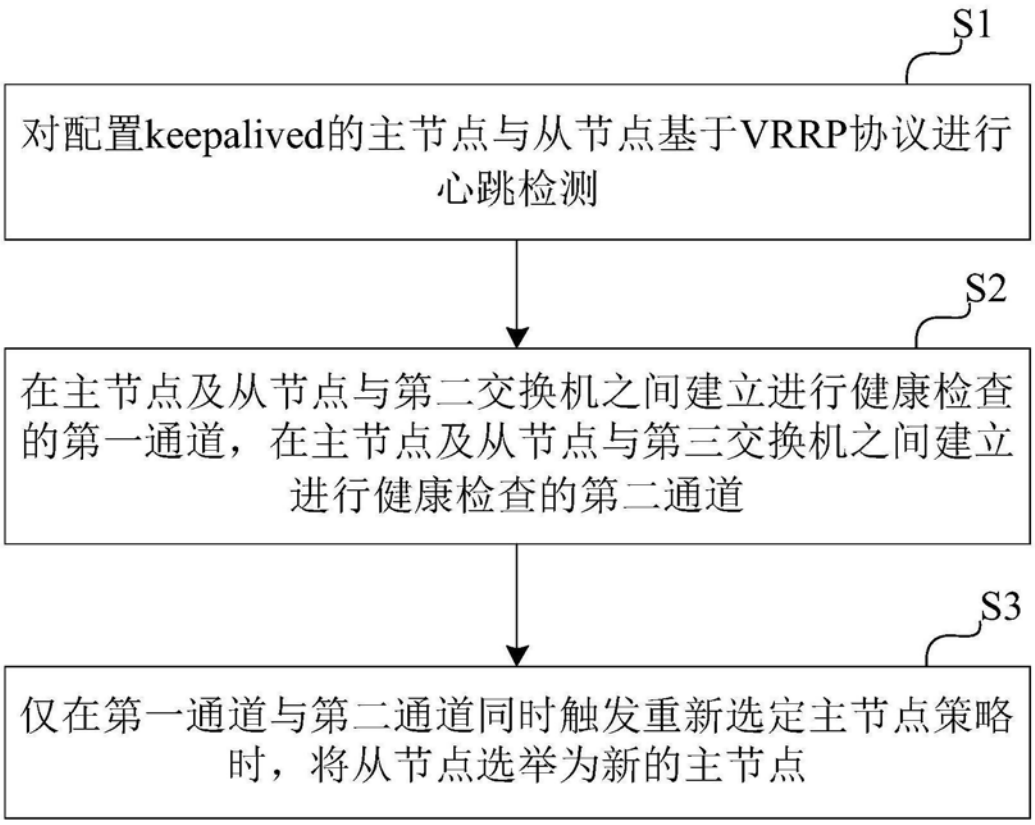

S1
对配置keepalived的主节点与从节点基于VRRP协议进行心跳检测
S2
在主节点及从节点与第二交换机之间建立进行健康检查的第一通道，在主节点及从节点与第三交换机之间建立进行健康检查的第二通道
S3
仅在第一通道与第二通道同时触发重新选定主节点策略时，将从节点选举为新的主节点

图1

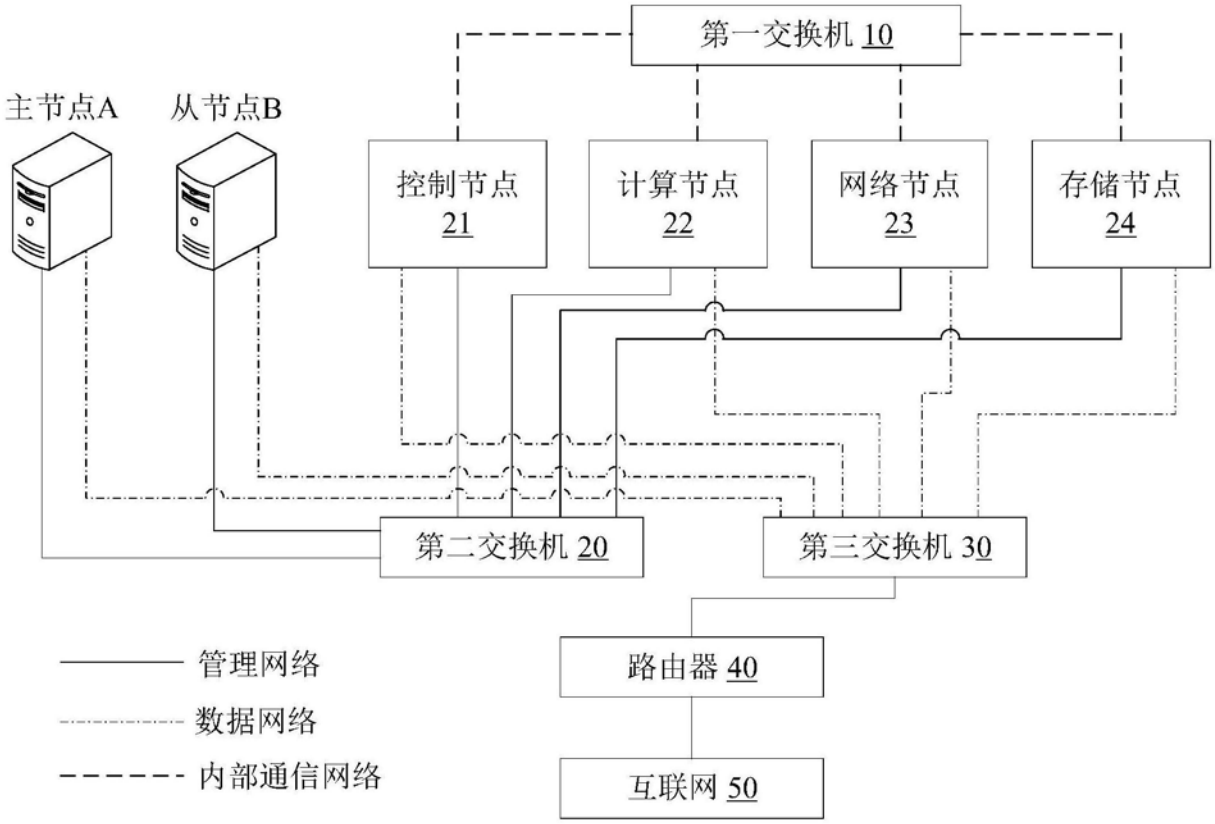

第一交换机 10
主节点A
从节点B
控制节点 21
计算节点 22
网络节点 23
存储节点 24
第二交换机 20
第三交换机 30
路由器 40
互联网 50
管理网络
数据网络
内部通信网络

图2

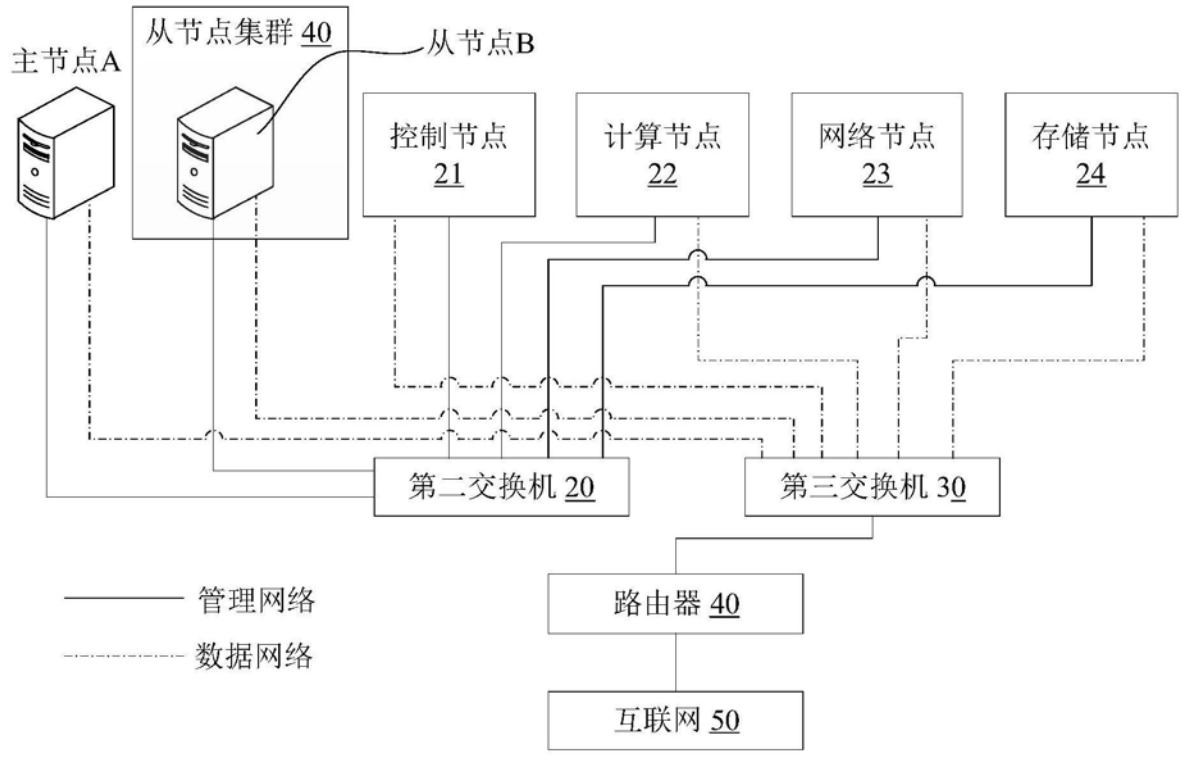

从节点集群 40
从节点B
主节点A
控制节点 21
计算节点 22
网络节点 23
存储节点 24
第二交换机 20
第三交换机 30
路由器 40
互联网 50
管理网络
数据网络

图3

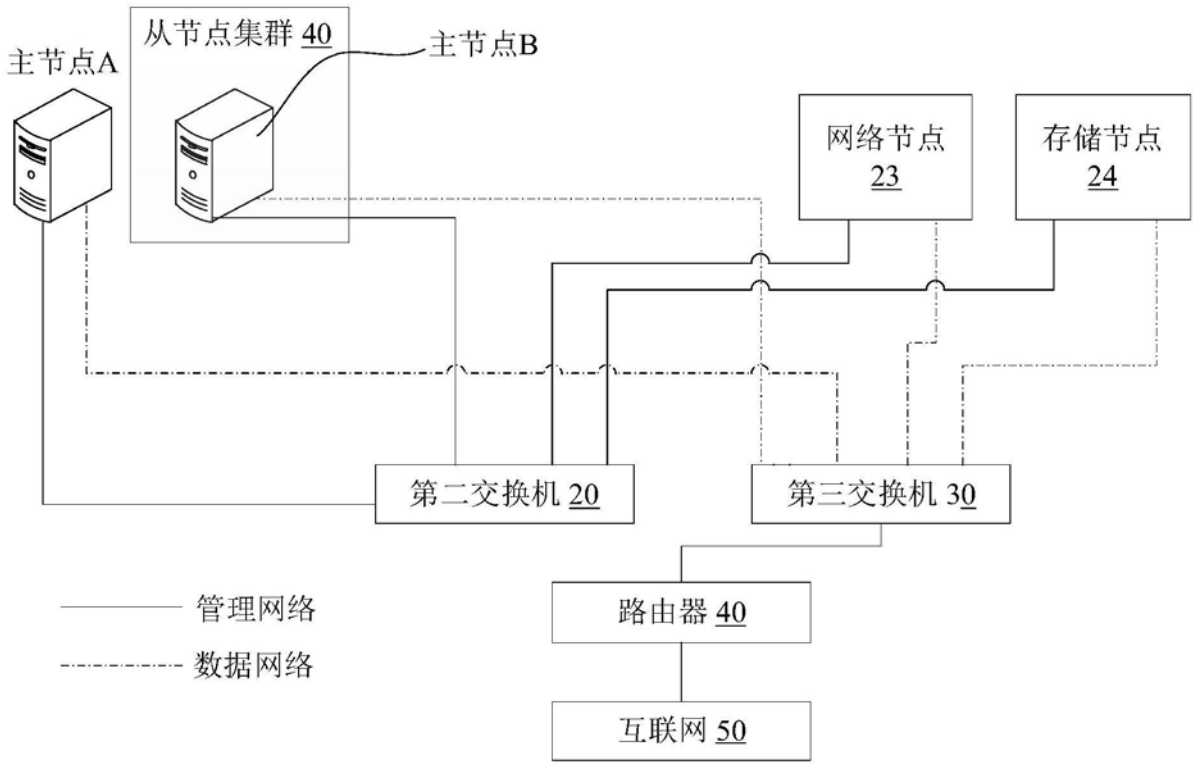

主节点A
从节点集群 40
主节点B
网络节点
23
存储节点
24
第二交换机 20
第三交换机 30
管理网络
数据网络
路由器 40
互联网 50

图4

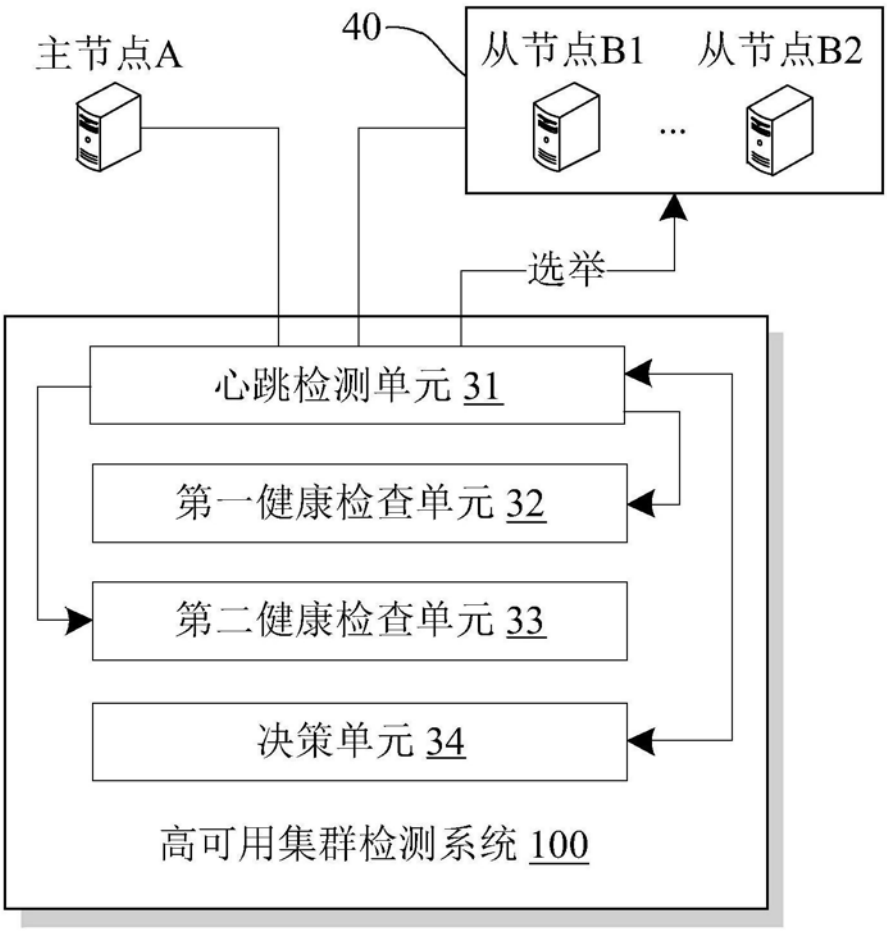

主节点A
40
从节点B1
从节点B2
...
选举
心跳检测单元 31
第一健康检查单元 32
第二健康检查单元 33
决策单元 34
高可用集群检测系统 100

图5

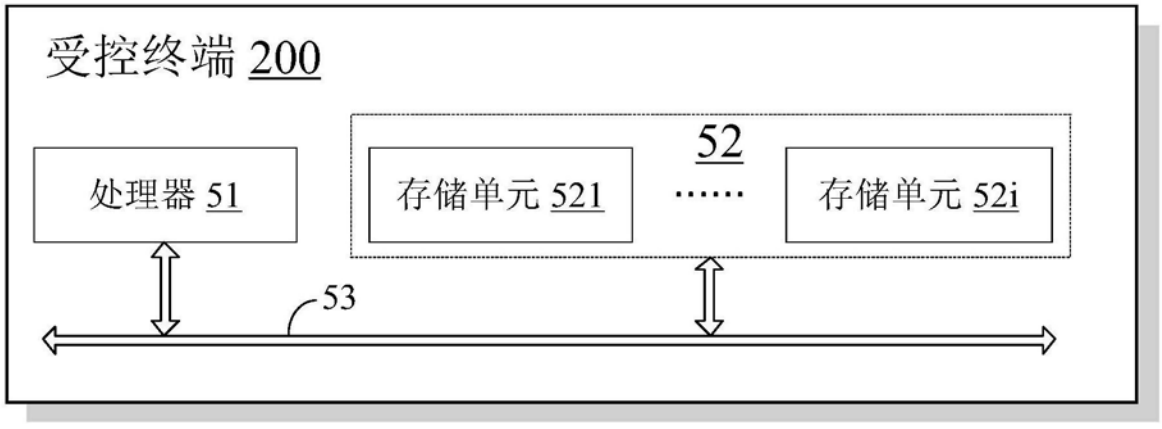

受控终端 200
处理器 51
52
存储单元 521
.....
存储单元 52i
53

图6