

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2015年7月30日 (30.07.2015)



(10) 国际公布号
WO 2015/109804 A1

- (51) 国际专利分类号:
H04L 29/06 (2006.01)
- (21) 国际申请号: PCT/CN2014/083113
- (22) 国际申请日: 2014年7月28日 (28.07.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410029760.5 2014年1月22日 (22.01.2014) CN
- (71) 申请人: 上海交通大学 (SHANGHAI JIAO TONG UNIVERSITY) [CN/CN]; 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。
- (72) 发明人: 管海兵 (GUAN, Haibing); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 马汝辉 (MA, Ruhui); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 李健 (LI, jian); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 戚正伟 (QI, Zhengwei); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。 钱正宇 (QIAN, Zhengyu); 中国上海市闵行区东川路800号, Shanghai 200240 (CN)。

- (74) 代理人: 上海旭诚知识产权代理有限公司 (SUN-RAY INTELLECTUAL PROPERTY ATTORNEYS); 中国上海市浦东新区东方路710号汤臣金融大厦1212室, 郑立, Shanghai 200122 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

[见续页]

(54) Title: DUAL-SERVER HOT-BACKUP DISASTER RECOVERY SYSTEM FOR NETWORK SERVICE IN VIRTUALIZATION ENVIRONMENT AND METHOD THEREFOR

(54) 发明名称: 一种虚拟化环境下针对网络服务的双机热备份容灾系统及其方法

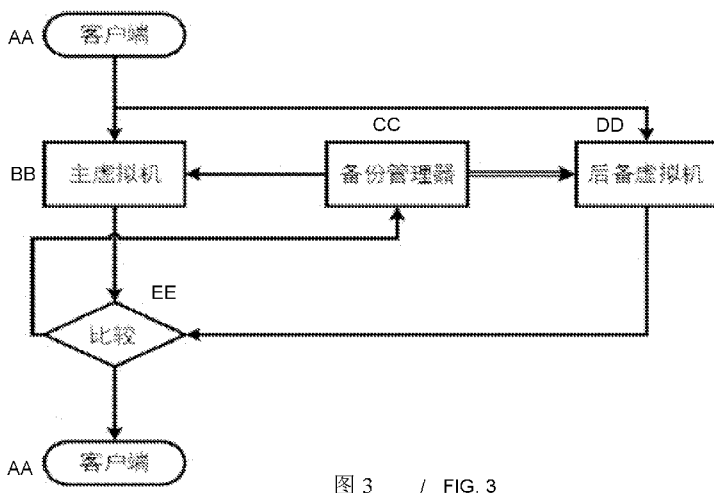


图3 / FIG. 3

AA CLIENT DD BACKUP VIRTUAL MACHINE
 BB MAIN VIRTUAL MACHINE EE COMPARISON
 CC BACKUP MANAGER

(57) Abstract: Provided are a dual-server hot-backup disaster recovery system for a network service in a virtualization environment and a method therefor. The dual-server hot-backup disaster recovery system comprises a main server and a backup server, wherein the main server is connected to the backup server via a network, a main virtual machine is operated on the main server, a backup virtual machine is operated on the backup server, the backup virtual machine is in a replacement state on an application-layer semantics of the main virtual machine, and the replacement state on the application layer semantics refers to the fact that, on the application layer semantics, the backup virtual machine can replace the main virtual machine to conduct service and produce a correct output for any customer request. By comparing the output of a main virtual machine with the output of a backup virtual machine by means of a replaceability rule, whether backup is required is judged, so that backup frequency is effectively reduced and system performance is improved on the basis that quick recovery is guaranteed. The present invention greatly reduces system overheads and increases system throughput.

(57) 摘要:

[见续页]

WO 2015/109804 A1

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

本发明提供一种虚拟化环境下针对网络服务的双机热备份容灾系统及其方法，双机热备份容灾系统包括主服务器与备份服务器，主服务器与备份服务器通过网络连接，主服务器上运行主虚拟机，备份服务器上运行后备虚拟机，后备虚拟机处于主虚拟机的应用层语义上的替代状态，应用层语义上的替代状态是指在应用层语义上后备虚拟机能够代替主虚拟机进行服务，对任何客户请求产生正确的输出。通过可替代性规则比较主虚拟机与后备虚拟机的输出，判断是否需要进行备份，从而有效的降低了备份频率，在保证快速恢复的基础上提升了系统的性能，本发明极大降低了系统开销，提高了系统的吞吐量。

一种虚拟化环境下针对网络服务的双机热备份容灾系统 及其方法

[1] 技术领域

[2] 本发明涉及虚拟化环境下的高可靠容灾技术，具体地，涉及一种虚拟化环境下针对网络服务的双机热备份容灾系统及其方法。

[3] 背景技术

[4] 当前，网络化服务是云计算与数据中心的主要服务形式。然而，由于断电、机器硬件故障、灾难或者人为因素（统称故障）的影响，使得这些网络应用停止对外服务、丢失数据的情况时有发生，不仅影响用户的使用，还会带来经济上的损失。因此，如何提高网络服务器的容灾，在故障后快速恢复对外的服务成了许多学者和公司的研究热点。

[5] 在现有的研究成果与产品中，有一部分是在虚拟化环境下实现的。

[6] 随着计算机技术，特别是网络技术的快速发展和广泛应用，人们对软件的可移植性，特别是软件在网络中移植的需求日益迫切，软件兼容性和可移植性也变得越来越重要。但由于计算机的历史发展过程中，产生了众多不同的、不相兼容的操作系统和指令集架构（Instruction Set Architecture, ISA），从而导致了软件的可移植性被局限在相似的平台。而在一个大型网络中可能包含基于各种ISA和操作系统的计算机，这就造成了软件可移植性需求和现状之间的矛盾越来越尖锐。而虚拟机（Virtual machine, VM）技术的出现消除了对软件运行平台的这些限制，从而有可能提供更高程度的兼容性和可移植性。虚拟机技术通过在硬件执行平台上加一层软件，从而屏蔽掉平台的差异性，或者说在一个平台上模拟出另一个平台或者多个平台。

[7] 目前，基于虚拟机技术的容灾解决方案可以分为检查点（checkpointing）技术和步锁（lockstepping）技术。

[8] Checkpointing 技术利用两台物理设备形成主服务器/后备服务器模式，对同一个应用/虚拟机进行备份，通过虚拟机的迁移技术，定期的将主服务器虚拟机状

态备份到后备服务器中，从而实现灾备。后备服务器的虚拟机处于非运行状态，在主服务器故障后能快速恢复到主服务器的前一状态，并继续保留有原先的所有网络连接，这样客户端感觉不到服务端发生了故障且从故障中恢复。但是，为了保证虚拟机间的状态一致性，必须进行周期性的频繁备份（20-40ms一次），导致主服务器的吞吐量会极大降低，CPU开销过大。同时Checkpointing技术将所有服务器发送到客户端的数据包保存在一个缓冲中，只有当一次备份完成后，才能释放这些数据包，这导致网络延迟的增加。

[9] Lockstepping 技术采用双机并行运行的方式保证主服务器后备服务器的状态一致，使得在主服务器故障后客户端能直接与后备服务器连接，快速从故障中恢复。但是Lockstepping技术只能适用于给虚拟机分配单个处理器的情况，对于多处理器虚拟机的性能扩展性很差，多于双处理器的虚拟机性能降低至单处理器是的1/7。另外，对于确定的指令，主后备服务器的虚拟机可以直接并行运行，而对于非确定的指令则需要主后备服务器虚拟机间做指令级别的同步，这也增加了系统的开销。

[10] 发明内容

[11] 有鉴于现有技术的上述缺陷，本发明提供一种双机热备份容灾系统。该方案中主虚拟机和后备虚拟机并行运行，根据客户端发送的请求生成各自的输出结果，比较主虚拟机和后备虚拟机的输出结果，如果不一致，需要进行备份，这样既保证了故障后的快速恢复，又有效的降低了系统开销。

[12] 本发明提供一种双机热备份容灾系统，用于虚拟化环境下的网络服务，双机热备份容灾系统包括主服务器与备份服务器，主服务器与备份服务器通过网络连接，其特征在于：主服务器上运行主虚拟机，备份服务器上运行后备虚拟机，后备虚拟机处于主虚拟机的应用层语义上的替代状态，应用层语义上的替代状态是指在应用层语义上后备虚拟机能够代替主虚拟机进行服务，对任何客户请求产生正确的输出。

[13] 进一步地，主服务器将客户端请求分别发送给主虚拟机与后备虚拟机，主虚拟机与后备虚拟机并行运行生成各自的响应数据包。

[14] 进一步地，双机热备份容灾系统还包括运行在主虚拟机上的主备份管理器，以

及运行在后备虚拟机上的后备备份管理器，后备备份管理器用于将后备虚拟机生成的响应数据包发送到主备份管理器，主备份管理器用于比较主虚拟机与后备虚拟机的响应数据包是否一致，如果一致，后备虚拟机处于主虚拟机的可替代状态，主备份管理器将主虚拟机生成的响应数据包发给客户端；如果不一致，后备虚拟机不处于主虚拟机的可替代状态。

[15] 进一步地，如果后备虚拟机不处于主虚拟机的可替代状态，主备份管理器将主虚拟机的当前状态备份到后备虚拟机。

[16] 进一步地，备份为非周期备份。

[17] 进一步地，备份到后备虚拟机为增量备份。

[18] 系统中采用了增量备份方式，以降低状态备份的开销。不同于现有的检查点技术，本发明中采用双机并行运行，因此在两次状态备份间，后备虚拟机的状态也会发生变化，这导致只备份主虚拟机状态增量是不够的。为了减少备份期间的传送内容，本发明中采用以空间换时间的方法。在主虚拟机与后备虚拟机第一次建立连接时，将主虚拟机的状态完全传输给后备虚拟机，并同时存入后备服务器的一个临时缓存中。每次备份主虚拟机状态时，仅传输在上次备份后发生改变的内容。首先将这些内容更新入后备服务器临时缓存中，然后将临时缓存中内容全备份入后备虚拟机，这就避免了两次备份间后备虚拟机状态变化对增量备份所产生的影响。

[19] 进一步地，后备备份管理器检测主虚拟机的心跳数据包，如果后备备份管理器没有接收到主虚拟机的心跳数据包，客户端请求数据包直接到达后备虚拟机，后备虚拟机产生响应数据包之后，后备备份管理器将响应数据包直接发给客户端。

[20] 系统中引入心跳包机制，用于后备虚拟机监测主虚拟机是否仍继续存活。如果后备虚拟机没有接收到心跳数据包，就认为主虚拟机已经故障，将采取故障恢复措施以替代主虚拟机继续提供服务。在这种情况下，客户端发送过来的请求数据包将直接到达后备虚拟机，后备虚拟机产生响应数据包之后，不再发送给主虚拟机，而是直接发给客户端。该情况下，客户端收到的数据包来源从主虚拟机改为后备虚拟机，并不会发现服务端进行了一次快速的故障恢复。

[21] 进一步地，在内存备份方面，启用了虚拟机监控器提供的影子页表机制，获取在上次状态备份之后被修改的页面。其基本原理是将所有虚拟机的页面都改成写保护，这样，一旦对某个页面进行写操作，就会触发一个异常并进入异常处理程序。

[22] 本发明还提供一种双机热备份容灾方法，其特征在于，包括以下步骤：

[23] (1) 主服务器通过流量控制，将客户端发出的请求分别发送到主虚拟机和后备虚拟机；

[24] (2) 主虚拟机与后备虚拟机根据客户请求并行运行，生成各自的响应数据包；

[25] (3) 后备备份管理器将后备虚拟机生成的响应数据包发送到主备份管理器；

[26] (4) 主备份管理器用于比较主虚拟机与后备虚拟机的响应数据包是否一致，如果一致，后备虚拟机处于主虚拟机的应用层语义上的替代状态，将主虚拟机的响应数据包发给客户端；如果不一致，后备虚拟机不处于主虚拟机的应用层语义上的替代状态，主备份管理器将主虚拟机的当前状态备份到后备虚拟机。

[27] 与现有技术相比，本发明提供的双机热备份容灾系统及其方法，具有以下有益的技术效果：

[28] (1) 系统实现中解决了主后备服务器双机并行情况下，存储访问的一致性，网络协议的一致性，多核状态CPU指令一致性等技术问题。

[29] (2) 基于可替代性规则，本方案中主服务器状态的备份为非周期性，备份间隔大于1秒，频率相对于现有技术降低两个数量级以上，这极大降低了系统开销，基本消除虚拟机状态备份对主服务器的性能干扰。

[30] (3) 相比于现有解决方案，本发明中主服务器不需等待备份完成即可交付输出结果，提高了系统的吞吐量。

[31] (4) 本发明方案能够提供快速灾备恢复，对于网络服务、数据库服务的灾备恢复时间快于现有技术。

[32] 附图说明

[33] 图1是现有检查点技术的流程示意图；

[34] 图2是现有步锁技术的流程示意图；

- [35] 图3是本发明的一个实施例的双机热备份容灾系统的流程示意图；
- [36] 图4是本发明的一个实施例的双机热备份容灾系统的增量备份的过程示意图。
- [37] 具体实施方式
- [38] 以下将结合附图对本发明的构思、具体结构及产生的技术效果作进一步说明，以充分地了解本发明的目的、特征和效果。
- [39] 图1是现有检查点技术的流程示意图。主虚拟机处理客户端请求并产生响应，后备虚拟机则处于非运行状态。在主服务器中有定时模块产生周期性事件，备份管理器在收到事件后获取主虚拟机状态，将在上一次备份后改变的状态备份到后备虚拟机上。
- [40] 图2是现有步锁技术的流程示意图。主虚拟机与后备虚拟机并行执行客户端发来的请求，主虚拟机将响应发回到客户端。由于非确定的指令（例如内存访问、时钟中断等）则需要在虚拟机间做指令级别的同步，以避免两边的状态发生差异。
- [41] 本发明提供一种双机热备份容灾系统，用于虚拟化环境下的网络服务，双机热备份容灾系统包括主服务器与备份服务器，主服务器与备份服务器通过网络连接，其特征在于：主服务器上运行主虚拟机，备份服务器上运行后备虚拟机，后备虚拟机处于主虚拟机的应用层语义上的替代状态，应用层语义上的替代状态是指在应用层语义上后备虚拟机能够代替主虚拟机进行服务，对任何客户请求产生正确的输出。
- [42] 客户端发送过来的请求数据包首先抵达外围交换机，交换机通过目的MAC地址来决定转发端口。当主虚拟机提供服务时，交换机学习到的虚拟机MAC地址对应端口为主服务器网卡端口，因此请求数据包被发往主服务器。
- [43] 主服务器将客户端请求分别发送给主虚拟机与后备虚拟机，主虚拟机与后备虚拟机并行运行生成各自的响应数据包。
- [44] 双机热备份容灾系统还包括运行在主虚拟机上的主备份管理器，以及运行在后备虚拟机上的后备备份管理器，后备备份管理器用于将后备虚拟机生成的响应数据包发送到主备份管理器，主备份管理器用于比较主虚拟机与后备虚拟机的响应数据包是否一致，如果一致，后备虚拟机处于主虚拟机的可替代状态，主

备份管理器将主虚拟机生成的响应数据包发给客户端；如果不一致，后备虚拟机不处于主虚拟机的可替代状态。

[45] 如果后备虚拟机不处于主虚拟机的可替代状态，主备份管理器将主虚拟机的当前状态备份到后备虚拟机。

[46] 备份为非周期备份。

[47] 备份到后备虚拟机为增量备份。

[48] 系统中采用了增量备份方式，以降低状态备份的开销。不同于现有的检查点技术，本发明中采用双机并行运行，因此在两次状态备份间，后备虚拟机的状态也会发生变化，这导致只备份主虚拟机状态增量是不够的。为了减少备份期间的传送内容，本发明中采用以空间换时间的方法。在主虚拟机与后备虚拟机第一次建立连接时，将主虚拟机的状态完全传输给后备虚拟机，并同时存入后备服务器的一个临时缓存中。每次备份主虚拟机状态时，仅传输在上次备份后发生改变的内容。首先将这些内容更新入后备服务器临时缓存中，然后将临时缓存中内容全备份入后备虚拟机，这就避免了两次备份间后备虚拟机状态变化对增量备份所产生的影响。

[49] 后备备份管理器检测主虚拟机的心跳数据包，如果后备备份管理器没有接收到主虚拟机的心跳数据包，客户端请求数据包直接到达后备虚拟机，后备虚拟机产生响应数据包之后，后备备份管理器将响应数据包直接发给客户端。

[50] 系统中引入心跳包机制，用于后备虚拟机监测主虚拟机是否仍继续存活。如果后备虚拟机没有接收到心跳数据包，就认为主虚拟机已经故障，将采取故障恢复措施以替代主虚拟机继续提供服务。后备服务器将会向交换机发送一个 ARP 数据包，该数据包的源 MAC 地址就是后备虚拟机的 MAC 地址。这使得交换机学习到一个新的 MAC 地址到端口的映射条目。之后客户端发送过来的目的 MAC 地址是虚拟机的数据包，将直接发送给后备服务器的网卡。后备虚拟机产生响应数据包之后，不再发送给主虚拟机，而是直接发给客户端。该情况下，客户端收到的数据包来源从主虚拟机改为后备虚拟机，并不会发现服务端进行了一次快速的故障恢复。

[51] 在内存备份方面，启用了虚拟机监控器提供的影子页表机制。获取在上次状态

备份之后哪些页面被修改。其基本原理是将所有虚拟机的页面都改成写保护，这样，一旦对某个页面进行写操作，就会触发一个异常并进入异常处理程序。借助于'影子页表'机制，能够很容易获取在上次状态备份之后，哪些页面被修改。

[52] 图3是本实施例的双机热备份容灾系统的流程示意图，具体过程如下：

[53] 步骤1、主服务器将客户端发出的请求包分发到主虚拟机与后备虚拟机，过程如下：首先，客户端发送过来的请求数据包通过外围交换机被交换机发送到主服务器。主服务器接收数据包后送到软件网桥，在软件网桥中通过配置Linux自带的工具TC（Traffic Control）来实现对网络数据包的截获和分发，将数据包发送到主虚拟机与后备虚拟机。

[54] TC 的配置方法如下：

[55] #tc qdisc add dev vif1.0 root handle 1: prio

[56] #tc filter add dev vif1.0 parent 1: protocol ip prio 10 u32 match u32 0 0 flowid 1:2
action mirred egress mirror dev eth0

[57] #tc filter add dev vif1.0 parent 1: protocol arp prio 11 u32 match u32 0 0 flowid 1:2
action mirred egress mirror dev eth0

[58] 步骤2、主虚拟机与后备虚拟机根据应用层语义并行执行，生成各自的输出，后备虚拟机将输出发送到主服务器。通过配置TC实现后备虚拟机输出的拦截与转发，具体方法如下：

[59] #tc qdisc add dev vif1.0 ingress

[60] #tc filter add dev vif1.0 parent ffff: protocol ip prio 10 u32 match u32 0 0 flowid 1:2
action mirred egress redirect dev eth0

[61] 步骤3、主服务器的管理器比较主虚拟机与后备虚拟机生成各自的输出是否满足可替代性规则。具体的，在管理器中实现两个队列形式的虚拟接口，将主虚拟机与后备虚拟机的输出分别重定向入一个接口。管理器通过对两队列中的包进行逐一比较，判断后备虚拟机是否仍是主虚拟机的可替代状态。通过配置TC实现输出的重定向，具体方法如下：

[62] a) 主虚拟机输出包重定向：

[63] #tc qdisc add dev vif1.0 ingress
[64] #tc filter add dev vif1.0 parent ffff: protocol ip prio 10 u32 match u32 0 0 flowid 1:2
action mirred egress redirect dev ifb0

[65] b) 后备虚拟机输出包重定向:

[66] #tc qdisc add dev eth0 ingress
[67] #tc filter add dev eth0 parent ffff: protocol ip prio 10 u32 match u32 0 0 flowid 1:2
action mirred egress redirect dev ifb1

[68] 步骤4、将主服务器的输出作为响应包发回给客户端。

[69] 步骤5、若判断后备虚拟机不是主虚拟机的可替代状态，则将主虚拟机的当前状态备份到后备虚拟机。在主服务器上和后备服务器的管理器中上都有一个备份守护进程，负责对状态的发送、接收和对虚拟机状态的更新。

[70] 图4是本实施例的双机热备份容灾系统的增量备份的过程示意图。

[71] 步骤1、主服务器上的备份管理器获取主虚拟机在上次备份后状态变更部分。

[72] 步骤2、备份管理器将变更部分发送到后备虚拟机。

[73] 步骤3、后备虚拟机将变更部分更新临时缓存。

[74] 步骤4、将临时缓存内容全部备份入后备虚拟机。

[75] 在磁盘文件备份方面，通过修改磁盘设备的后端驱动，截获主虚拟机与后备虚拟机对磁盘的写入操作。在两次备份间主虚拟机与后备虚拟机的磁盘写入数据都被暂时保存在各自的临时缓存中。当备份时用主虚拟机临时缓存中内容替换备份虚拟机临时缓存的内容，然后分别写入磁盘。

[76] 在设备备份方面，由于设备状态涉及虚拟机监控器的前后端模型，获取较困难，因此选用丢弃主虚拟机与后备虚拟机设备驱动之前的状态。当备份完成后重新建立连接来使得设备状态保持一致。

[77] 本发明提供的双机热备份容灾系统及其方法，解决了主后备服务器双机并行情况下，存储访问的一致性，网络协议的一致性，多核状态CPU指令一致性等技术问题；基于可替代性规则，本方案中主服务器状态的备份为非周期性，备份间隔大于1秒，频率相对于现有技术降低两个数量级以上，这极大降低了系统开销，基本消除虚拟机状态备份对主服务器的性能干扰；主服务器不需等待备份

完成即可交付输出结果，提高了系统的吞吐量；提供快速灾备恢复，对于网络服务、数据库服务的灾备恢复时间快于现有技术。

[78] 以上详细描述了本发明的较佳具体实施例。应当理解，本领域的普通技术人员无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此，凡本技术领域中技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案，皆应在由权利要求书所确定的保护范围内

。

权利要求书

- [权利要求 1] 一种双机热备份容灾系统，用于虚拟化环境下的网络服务，所述双机热备份容灾系统包括主服务器与备份服务器，所述主服务器与所述备份服务器通过网络连接，其特征在于，所述主服务器上运行主虚拟机，所述备份服务器上运行后备虚拟机，所述后备虚拟机处于所述主虚拟机的应用层语义上的替代状态，所述应用层语义上的替代状态是指在应用层语义上后备虚拟机能够代替主虚拟机进行服务，对任何客户请求产生正确的输出。
- [权利要求 2] 如权利要求1所述的双机热备份容灾系统，其特征在于，所述主服务器将客户端请求分别发送给所述主虚拟机与所述后备虚拟机，所述主虚拟机与所述后备虚拟机并行运行生成各自的响应数据包。
- [权利要求 3] 如权利要求2所述的双机热备份容灾系统，其特征在于，所述双机热备份容灾系统还包括运行在所述主虚拟机上的主备份管理器，以及运行在所述后备虚拟机上的后备备份管理器，所述后备备份管理器用于将所述后备虚拟机生成的所述响应数据包发送到所述主备份管理器，所述主备份管理器用于比较所述主虚拟机与所述后备虚拟机的响应数据包是否一致，如果一致，所述后备虚拟机处于所述主虚拟机的可替代状态，所述主备份管理器将所述主虚拟机生成的所述响应数据包发给所述客户端；如果不一致，所述后备虚拟机不处于所述主虚拟机的可替代状态。
- [权利要求 4] 如权利要求3所述的双机热备份容灾系统，其特征在于，如果所述后备虚拟机不处于所述主虚拟机的可替代状态，所述主备份管理器将所述主虚拟机的当前状态备份到所述后备虚拟机。
- [权利要求 5] 如权利要求4所述的双机热备份容灾系统，其特征在于，所述备份为非周期备份。
- [权利要求 6] 如权利要求4所述的双机热备份容灾系统，其特征在于，所述备份到所述后备虚拟机为增量备份。

- [权利要求 7] 如权利要求3所述的双机热备份容灾系统，其特征在于，所述后备备份管理器检测所述主虚拟机的心跳数据包，如果所述后备备份管理器没有接收到所述主虚拟机的心跳数据包，所述后备虚拟机产生响应数据包之后，所述后备备份管理器将所述响应数据包直接发给所述客户端。
- [权利要求 8] 如权利要求1所述的双机热备份容灾系统，其特征在于，在内存备份方面，启用了虚拟机监控器提供的影子页表机制，获取在上次状态备份之后被修改的页面。
- [权利要求 9] 一种如权利要求1-8所述的双机热备份容灾系统的双机热备份容灾方法，其特征在于，包括以下步骤：
- (1) 主服务器通过流量控制，将客户端发出的请求分别发送到主虚拟机和后备虚拟机；
 - (2) 所述主虚拟机与所述后备虚拟机根据所述客户端发出的请求并行运行，生成各自的响应数据包；
 - (3) 后备备份管理器将所述后备虚拟机生成的所述响应数据包发送到主备份管理器；
 - (4) 所述主备份管理器用于比较所述主虚拟机与所述后备虚拟机的响应数据包是否一致，如果一致，所述后备虚拟机处于所述主虚拟机的应用层语义上的替代状态，将所述主虚拟机的响应数据包发给所述客户端；如果不一致，所述后备虚拟机不处于所述主虚拟机的应用层语义上的替代状态，所述主备份管理器将所述主虚拟机的当前状态备份到所述后备虚拟机。

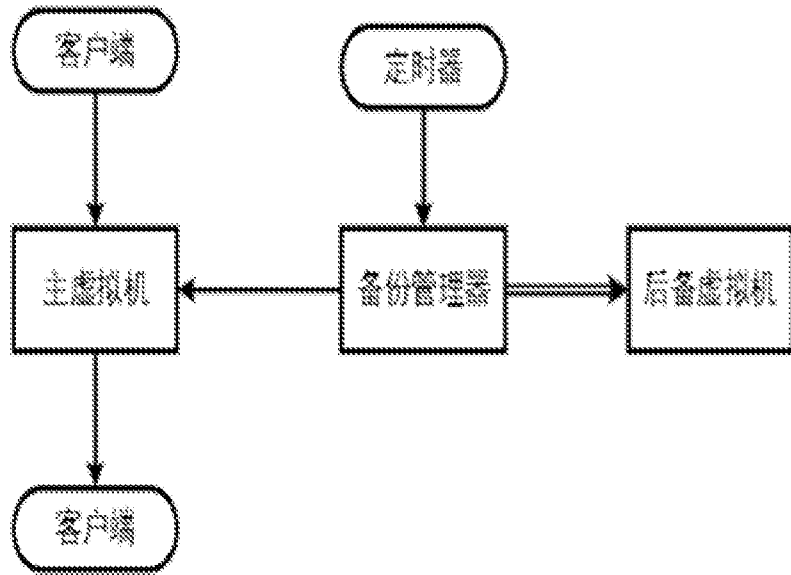


图 1

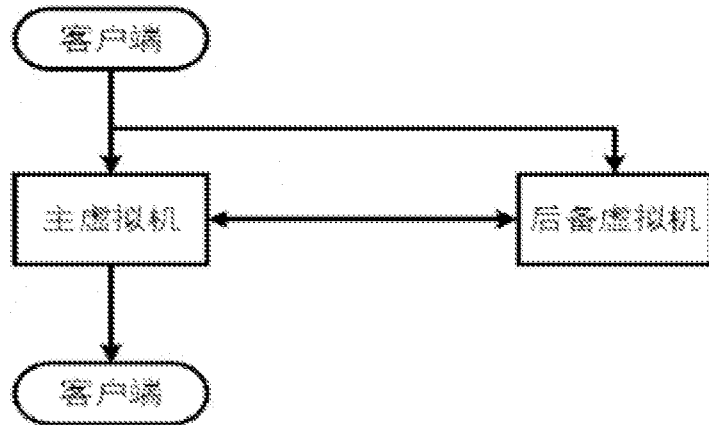


图 2

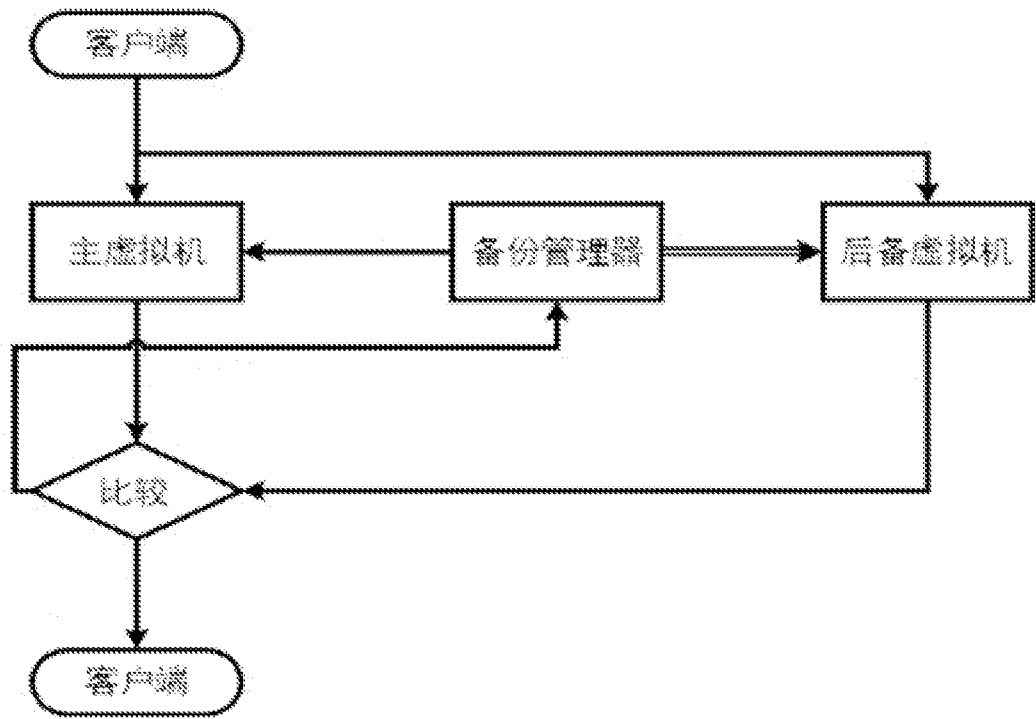


图 3

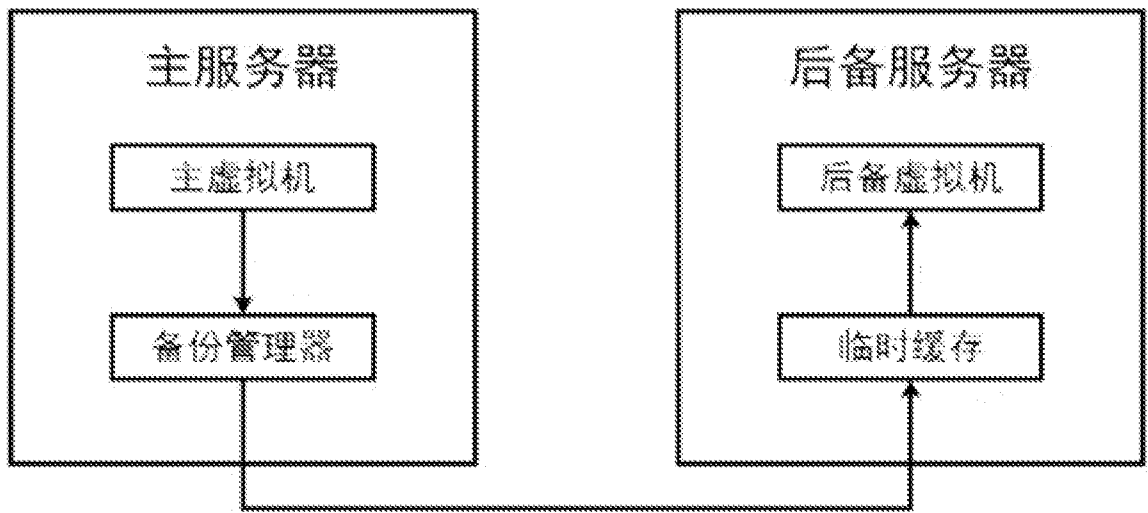


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/083113

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNKI, CNPAT: standby, hot standby, disaster recovery, Virtual machine, backup, master, prime, secondary, redundant, compare

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103761166 A (SHANGHAI JIAO TONG UNIVERSITY), 30 April 2014 (30.04.2014), claims 1-9	1-9
X	CN 103501290 A (WONDERS INFORMATION CO., LTD.), 08 January 2014 (08.01.2014), description, paragraphs [0015]-[0020]	1-9
A	CN 103412800 A (HUAWEI TECHNOLOGIES CO., LTD.), 27 November 2013 (27.11.2013), the whole document	1-9
A	US 2010318991 A1 (VMWARE INC.), 16 December 2010 (16.12.2010), the whole document	1-9

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Date of the actual completion of the international search
11 October 2014 (11.10.2014)

Date of mailing of the international search report
27 October 2014 (27.10.2014)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer
LI, Ping
Telephone No.: (86-10) **62413398**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2014/083113

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103761166 A	30 April 2014	None	
CN 103501290 A	08 January 2014	None	
CN 103412800 A	27 November 2013	None	
US 2010318991 A1	16 December 2010	US 2012284714 A1	08 November 2012
		US 8201169 B2	12 June 2012

国际检索报告

国际申请号

PCT/CN2014/083113

<p>A. 主题的分类</p> <p>H04L 29/06(2006.01) i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																											
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04L, G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>WPI, EPODOC, CNKI, CNPAT:虚拟机, 主, 备, 热备, 冗余, 容灾, 备份, 比较, Virtual machine, backup, master, prime, secondary, redundant, compare</p>																											
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 103761166 A (上海交通大学) 2014年 4月 30日 (2014 - 04 - 30) 权利要求1-9</td> <td>1-9</td> </tr> <tr> <td>X</td> <td>CN 103501290 A (万达信息股份有限公司) 2014年 1月 08日 (2014 - 01 - 08) 说明书第[0015]-[0020]段</td> <td>1-9</td> </tr> <tr> <td>A</td> <td>CN 103412800 A (华为技术有限公司) 2013年 11月 27日 (2013 - 11 - 27) 全文</td> <td>1-9</td> </tr> <tr> <td>A</td> <td>US 2010318991 A1 (VMWARE INC.) 2010年 12月 16日 (2010 - 12 - 16) 全文</td> <td>1-9</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型:</p> <table border="0"> <tr> <td>“A” 认为不特别相关的表示了现有技术一般状态的文件</td> <td>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</td> </tr> <tr> <td>“E” 在国际申请日的当天或之后公布的在先申请或专利</td> <td>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</td> </tr> <tr> <td>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</td> <td>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</td> </tr> <tr> <td>“O” 涉及口头公开、使用、展览或其他方式公开的文件</td> <td>“&” 同族专利的文件</td> </tr> <tr> <td>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</td> <td></td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 103761166 A (上海交通大学) 2014年 4月 30日 (2014 - 04 - 30) 权利要求1-9	1-9	X	CN 103501290 A (万达信息股份有限公司) 2014年 1月 08日 (2014 - 01 - 08) 说明书第[0015]-[0020]段	1-9	A	CN 103412800 A (华为技术有限公司) 2013年 11月 27日 (2013 - 11 - 27) 全文	1-9	A	US 2010318991 A1 (VMWARE INC.) 2010年 12月 16日 (2010 - 12 - 16) 全文	1-9	“A” 认为不特别相关的表示了现有技术一般状态的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件	“E” 在国际申请日的当天或之后公布的在先申请或专利	“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性	“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)	“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性	“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件	“P” 公布日先于国际申请日但迟于所要求的优先权日的文件	
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																									
PX	CN 103761166 A (上海交通大学) 2014年 4月 30日 (2014 - 04 - 30) 权利要求1-9	1-9																									
X	CN 103501290 A (万达信息股份有限公司) 2014年 1月 08日 (2014 - 01 - 08) 说明书第[0015]-[0020]段	1-9																									
A	CN 103412800 A (华为技术有限公司) 2013年 11月 27日 (2013 - 11 - 27) 全文	1-9																									
A	US 2010318991 A1 (VMWARE INC.) 2010年 12月 16日 (2010 - 12 - 16) 全文	1-9																									
“A” 认为不特别相关的表示了现有技术一般状态的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件																										
“E” 在国际申请日的当天或之后公布的在先申请或专利	“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性																										
“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)	“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性																										
“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件																										
“P” 公布日先于国际申请日但迟于所要求的优先权日的文件																											
<p>国际检索实际完成的日期</p> <p>2014年 10月 11日</p>	<p>国际检索报告邮寄日期</p> <p>2014年 10月 27日</p>																										
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国</p> <p>传真号 (86-10)62019451</p>	<p>受权官员</p> <p>李萍</p> <p>电话号码 (86-10)62413398</p>																										

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/083113

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103761166	A	2014年 4月 30日	无			
CN	103501290	A	2014年 1月 08日	无			
CN	103412800	A	2013年 11月 27日	无			
US	2010318991	A1	2010年 12月 16日	US	2012284714	A1	2012年 11月 08日
				US	8201169	B2	2012年 6月 12日