高可用性双机系统的有限自动机

王文煜，刘 勇，秦志光

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】介绍了提供高可用性的双机系统的原理、部署和体系结构；定义了双机系统各部件功能，描述了子系统间的接口，给出了该双机系统的有限自动机。该自动机具有有限数目的内部状态，模拟了双机系统的工作原理及内部各状态的相互转化，描述了服务切换、进程与服务监控、心跳侦测、网络参考点侦测等过程，为双机系统的行为描述和结构设计提供了理论依据和论证。

关 键 词 双机系统； 有限自动机； 进程与服务监控； 心跳侦测； 网络参考点侦测； 服务切换

中图分类号 TP302 文献标识码 A

Turing Machine of Dual-Hosts System Providing High Availability

Wang Wenyu，Liu Yong，Qin Zhiguang

(School of Computer Science and Engineering，UEST of China Chengdu 610054)

Abstract Dual-hosts sytem providing high availability(HA) , which is the one of most important method for data integrity and service continuous, is widely required for user.The principle, deployment and framework of a dual-hosts system is introduced, and a turing machine of the dual-hosts system is presented in this paper.The turing machine simulates the dual-hosts system and describes the states transition of the dual-hosts system and supplies a theory basis for designing architecture of the dual-hosts system.

Key words dual-hosts system; turing machine; process & service monitor; heartbeat; network referenced-node; service switch

高可用性(High Availability，HA)是指计算机系统拥有几乎为零的故障时间[1]。提供高可用性的双机系统是保护用户数据完整、提供服务连续的重要手段之一，有着广泛的用户需求。

1 高可用性双机系统

1.1 基本原理与部署

双机系统的主要功能是当主机系统出现异常，不能正常响应客户请求时，从机自动接管主机系统的工作，继续提供对外服务，确保系统的不间断运行。接管的内容包括原主机系统的网络地址与当前提供的服务。服务在主从机系统之间切换时，工作IP地址也随之漂移。主机和从机所组成的系统对外只表现为一个IP(工作IP)，同时主机和从机系统分别有自己的实体IP。当发生服务切换时，工作IP在主机和从机之间漂移，以保证该IP始终由当前的工作机占有，即始终将工作机的实际IP地址映射成双机系统的工作IP。

双机节点通过网络连线和RS232连线进行信息交换。网络参考点是与双机在同一子网中的任一可通讯节点，它用于判断发生网络故障的具体节点。

收稿日期：2003 – 02 – 25
基金项目：国家计算机网络与信息安全管理中心资助项目(2002-研3-022)
作者简介：王文煜(1978 –)，男，硕士生，主要从事信息和网络安全方面的研究.

双机部署示意如图1所示。

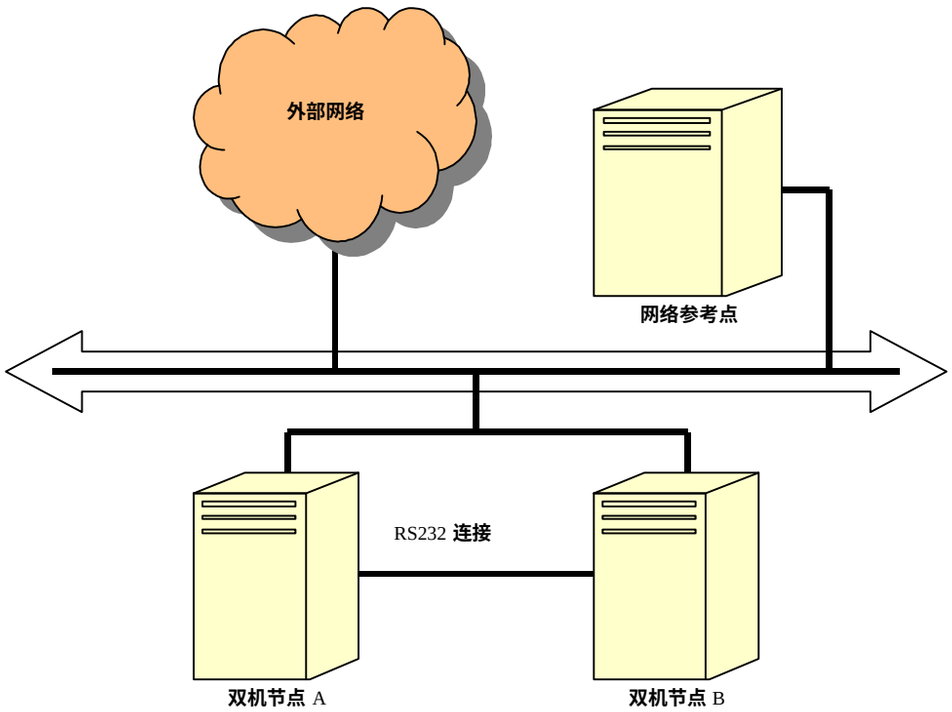

图1 双机系统部署示意图

1.2 系统体系结构

双机系统软件实现方式中主要包括服务切换子系统与监控子系统两部分，前者是在服务切换条件满足时保证服务顺利过渡，而后者实时监控满足服务切换条件的事件是否发生。服务切换子系统与监控子系统的实现粒度将影响双机系统所提供的高可用性的最终性能。

服务切换子系统通过IP地址漂移与地址解析协议（Address Resolution Protocol, ARP）地址刷新，使服务的切换对用户透明，它为监控子系统提供3个接口：

1) 由当前主机提出切换请求的主动切换接口；

2) 由当前从机提出切换请求的被动切换接口；

3) 主机释放主机身份，降级为从机的主动释放接口。

监控子系统通过进程与服务监控、心跳侦测[2]、网络参考点侦测3个模块完成对双机状态的监控，它们的功能是：

1) 进程监控完成对关键进程与服务的状态监控。当主机中关键进程与服务不能完成预定义的功能时，调用主动切换接口，由主机发起切换。在监控过程中，主机产生的信息将通过心跳线路传递到从机，使从机能够对主机当前服务有必要的了解，以保证在满足服务切换条件时顺利接管主机服务。

2) 心跳侦测是监视计算机运行状态的通用方式。它通过两条心跳线路即网络连线和RS232串口连线完成信息传递(本文中我们称前者为心跳线路1，后者为心跳线路2)。每一台计算机都运行一心跳侦测程序，该程序每秒发出一次“ alive ”信息。每台计算机向对方发出该信息并接受对方发来的信息，根据信息是否到达及信息内容判断对方主机当前状态；当从机中两条心跳线路均未收到指定信息时，由从机调用被动切换接口接管服务。

3) 网络参考点用于确定节点的网络接口和网络连线是否正常。通过检查当前提供服务的节点(即当前主机)与网络的连接来判断连接状态，若连接不通，从机调用被动切换接口，使从机接管服务，而主机释放相应资源。

双机系统的体系结构如图2所示，图中虚线箭头表示信息传递；实线箭头表示调用关系。

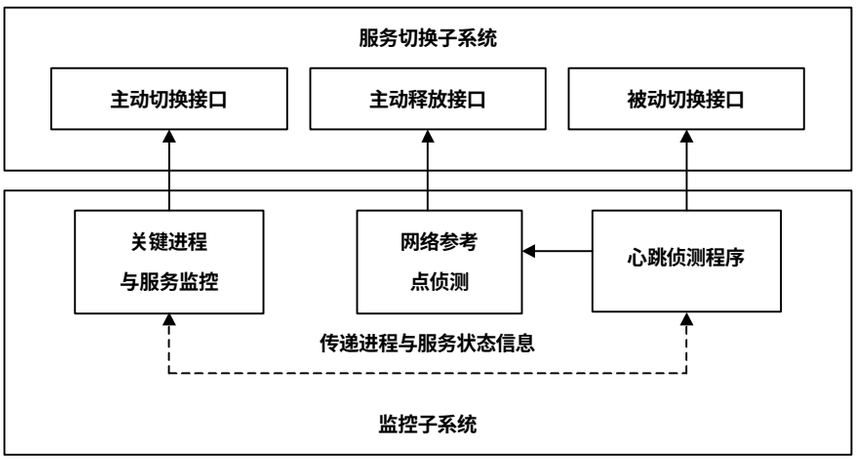

图2　双机系统体系结构

2　有限自动机

确定型(不确定型)有限自动机M是一个五元组[3]

$$M=(K,\Sigma,d,q_0,F)$$

式中　K 是状态的有限集合；Σ 是有限输入字母表；d 是 $K\times\Sigma$ 到K的一种映射；q_0 是初始状态，$q_0\subseteq K$；F是结束状态集合，$F\subseteq K$。自动机 M 处于状态q，输入字符a时，根据指令 M 将转到状态p，记为$d\ (q,a)=p$。若后继状态唯一，则称 M 为确定型有限自动机；若后继状态多于一个，则称 M 为不确定型有限自动机。有限自动机的工作情况可用状态转换图来描述。

双机系统的有限自动机的表达式如下

$$M=(K,\Sigma,d,q_0,F\ )$$

式中　$K=\{q_0,q_1,q_2,q_3,q_4,q_5,q_6,q_7,q_8,q\}$；$\Sigma=\{0,1,2,3,4\}$，1/0表示操作成功/失败，2表示节点为当前主机，3表示节点为当前从机，4表示线路1恢复正常；$F=\{q\}$。

映射　d ：$K\times\Sigma\to K$ 为

$d(q_0,0)=q_1$　$d(q_0,1)=q_2$　$d(q_2,0)=q_3$　$d(q_1,1)=q_0$

$d(q_3,0)=q_7$　$d(q_2,1)=q_2$　$d(q_4,0)=q_7$　$d(q_3,1)=q_4$

$d(q_5,0)=q$　$d(q_4,1)=q_4$　$d(q_6,0)=q$　$d(q_4,3)=q_8$

$d(q_8,0)=q$　$d(q_5,1)=q_0$　$d(q_1,2)=q_5$　$d(q_8,1)=q_4$

$d(q_4,2)=q_6$　$d(q_8,1)=q_7$　$d(q_1,3)=q_0$　$d(q_6,1)=q_4$

$d(q_4,4)=q_2$　$d(q_7,1)=q_0$

该自动机的状态转换如图3所示，图中各状态的含义为：q 为结束状态；q_0 为初始状态；q_1 为进程与服务失效状态；q_2 为使用心跳线路1进行信息交换；q_3 为等待心跳线路2信息；q_4 为使用心跳线路2进行信息交换；q_5 为主动切换状态；q_6 为释放主机身份状态；q_7 为节点宕机处理状态；q_8 为被动切换状态。

双机启动后进入状态 q_0，在该状态中启动进程与服务监控、初始化心跳线路、创建用于心跳侦测的三个进程，用于心跳信息传输的读、写进程以及用于控制读写状态的主控进程，所有进程皆后台运行。当监测到进程与服务失效时，进入状态 q_1；当由心跳线路1接收到第一条心跳信息时，进入状态 q_2；用户退出系统时，M 转换为 q 状态。

状态 q_1 中重启失效进程或服务，重启成功返回状态 q_0，失败则转换为状态 q_5；如果当前节点为从机，返回状态 q_0。状态 q_5 中从机接管服务、释放系统工作资源，从机接管成功返回状态 q_0，否则转换为 q 状态。

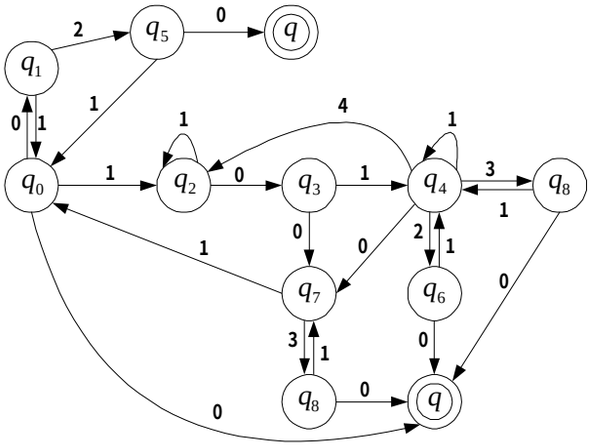

图3 双机系统自动机*M*状态转换图

在状态 q_2 中持续使用心跳线路1进行信息交换，读写进程使用用户数据报协议(User Datagram Protocol, UDP)**传递广播信息，信息内容包括当前进程扫描结果、当前服务扫描结果、主机状态值。通信双方根据接收的心跳信息分析对方节点的服务与主机状态，并进行必要的同步。当双方节点在指定的时间内无法从线路1中读取心跳信息时，** M **转换为状态 q_3。状态 q_3 中心跳线路切换至线路2，并等待线路2中第一条心跳信息的到达，如果在指定时间内接收到第一条心跳信息，则进入状态 q_4，否则进入状态 q_7。在状态 q_4 中持续使用心跳线路2进行信息交换，其信息交换的内容与状态 q_2 相同。在通过线路2接收到第一条心跳信息后，可以判断出是由于线路1出现故障(并非对方节点宕机)造成心跳信息在线路1上无法接收。此时派生网络参考点侦测进程判断发生网络连接故障的具体节点，当网络发生故障时，如果节点为当前主机，转换为状态 q_6，如果节点为当前从机，转换为状态 q_8。另外，状态 q_4 下派生监听进程用于对心跳线路1的监听，当线路1成功接收心跳信息后，心跳信息交换线路自动切换至线路1，即返回状态 q_2。如果使用线路2仍然没有接收到心跳信息，则判断对方节点宕机，进入状态 q_7。**

状态 q_6 中成功完成主机身份释放返回状态 q_4，否则转换为状态 q。状态 q_8 中成功完成被动切换返回状态 q_4，否则转换为状态 q。

状态 q_7 进行对方节点宕机处理，关闭读写进程，派生监听进程。如果本节点是从机则进入状态 q_8，接管当前服务。当监听进程接收监测到对方主机恢复正常时，返回状态0，完成心跳侦测的自动重建。状态 q_8 中成功完成被动切换，返回状态 q_4，否则转换为状态 q。

3 结 束 语

自动机是实际系统的抽象模型，它具有有限数目的内部状态，在不同的输入序列的作用下，系统内部的状态不断地转换，并且可能由此产生某种形式的输入序列。文中给出的双机系统的有限自动机，模拟了提供高可用性的双机系统的工作原理及内部各状态的相互转化，描述了服务切换、进程与服务监控、心跳侦测、网络参考点侦测等过程。

参 考 文 献

[1] Lewis P. A high-availability cluster for linux[EB/OL]. R. Network: http://linuxjournal.com/lj-issues/issue64/3247.html, 1999-08-01

[2] Robertson A. Linux-HA heartbeat system design[EB/OL]. http://lists.community.tummy.com/pipermail/linux-ha- dev/2000 -October/000992.html, 2000-10-1

[3] 何成武. 自动机理论及其应用[M]. 北京: 科学出版社, 1990. 133

编 辑 熊思亮