

# 基于热备份的主备倒换在高端路由器中的应用

张玉峰<sup>1</sup>, 孙知信<sup>1,2</sup>

(1. 南京邮电大学 计算机技术研究所, 江苏 南京 210003;

2. 南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210003)

**摘要:**为了提高高端路由器的可靠性、减少故障持续时间,关键在于控制高端路由器主备倒换的时间。文中重点分析了热备份下的主备倒换技术,在出现故障或者人为主动触发主备倒换后,高端路由器的主控板会经过平滑函数阶段、热备份初始化阶段、批同步阶段和实时同步阶段,而备UPC会在极短的时间内变为主UPC,并承担前主UPC的所有业务,从而节省大量学习时间。实验证明,本文采用基于热备份的主备倒换技术并将其运用于高端路由器中,大大地提高了系统的可靠性,减少了故障持续时间。

**关键词:**热备份;主备倒换;主控板(UPC);高端路由器;高可靠性

**中图分类号:**TP39

**文献标识码:**A

**文章编号:**1673-629X(2010)03-0172-04

## Swap Based on Hot Standby and Its Application in High-end Router

ZHANG Yu-feng<sup>1</sup>, SUN Zhi-xin<sup>1,2</sup>

(1. College of Computer of Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. State Key Laboratory for Novel Computer Software Technology, Nanjing University, Nanjing 210003, China)

**Abstract:** In order to improve the reliability of high-end routers and reduce the time of its fault duration, the key lied in controlling swap time. Analyzed its swapping technology under hot standby, in the event of fault or taking the initiative to trigger the swap, the high-end router's UPC went through stages of the smoothing function, the hotbackup initialization stage, the batch synchronization and the real-time synchronization, and backup UPC became the main UPC in a very short time, and beard all the business from the former UPC, which saved lots of learning time. The experiments show that it is greatly increased reliability and reduced the duration of the fault.

**Key words:** hot standby; swap; universal processes card(UPC); high-end routers; high reliability

## 0 引言

随着IP技术的飞速发展,各种增值业务在互联网上得到了广泛的应用。新兴的NGN/3G/IPTV流媒体、大客户专线和VPN互联等重要的电信级业务,对IP电信网的可靠性提出了很高的要求。电信级业务对于IP网络的可靠性主要包括三个层面:设备可靠性、链路可靠性和网络可靠性。在承载网中,网络设备的可用性要求达到99.999%,大致相当于设备在一年连续运行中因各种原因造成停机维护的时间少于5分钟。高可靠性是电信级设备的基本要求,是电信运营商建设网络的基本出发点。提供路由冗余设

计,是提高网络可靠性的重要手段之一。文中通过主备单板复用,降低故障持续时间,从而达到最大的运营时间。降低故障持续时间主要就是尽可能地缩短主备倒换的时间。

## 1 相关技术的研究

传统的主备倒换是基于冷备份,主用板与备用板之间并不进行运行状态和运行数据的同步。路由器启动后,主用板和备用板都要进行程序加载,并且开始相关模块的初始化。但是,与主用板不同,备用板不需要完成所有的初始化(包括配置文件的执行),而是在完成之前的最后一步暂时阻塞。主用板正常执行,一旦主用板出现故障,备用板重新启动所有的业务板并完成最后的初始化,接替主用板的工作。

冷备份<sup>[1]</sup>节省了备用板的加载以及启动时间、备用板配置恢复的时间。但是,对于业务板和路由板的重启时间、业务板恢复配置时间、路由板路由学习聚合

收稿日期:2009-07-04;修回日期:2009-10-27

基金项目:国家自然科学基金(60973140);江苏省自然科学基金(BK2009425);江苏省高校自然科学研究项目(08KJB520005)

作者简介:张玉峰(1983-),男,江苏盐城人,硕士研究生,研究方向为计算机软件通信方面的研究;孙知信,博士,硕士生导师,教授,研究方向为计算机网络与安全、计算机仿真、软件工程。

时间并没有改善,一般来说业务板和路由板的重启需要十几到几十秒的时间,再加上业务板恢复配置,路由板路由学习的时间甚至可能需要高达几分钟的中断,上线用户都得下线,所在业务都得中断,还需要客户重新拨号,这显然对于要求高可靠性的高端路由器来说是无法容忍的。

文献[2]主要分析了热备份下的主备倒换在节省业务板和路由板的重启时间、业务板恢复配置时间、路由板路由学习聚合时间等方面的优势。但是这种说法并没有通过实验得到验证,无法证明该方法的可行性。

文献[3~5]主要讲述了一种可自动实现主备倒换的双向光通信设备,此设备保证对端不支持主备倒换功能的条件下,仍可自动实现主备倒换以保证业务的正常运行,并且只需要一对光纤与对端设备连接,可大大降低成本。但是当本端出现故障的时候,整个设备有可能就不能正常运转;同时对端也不能自动实现主备倒换。如果考虑在本端出现故障时,对端仍能主备切换并保证业务正常运转,那就相对完善了。

综上所述,当代高端路由器在互联网拓扑中的特殊身份,其稳定性与可靠性是至关重要的,每一分钟的故障都会对企业 and 电信运营商带来巨大的损失,只有降低每次故障的持续时间,才能延长可持续运营时间。上述方法中可能不能全面地解决已有的问题,而能否合理利用冗余设计,提供基于热备份的主备倒换就是解决这一问题的关键。理论分析与实验表明,该方法具有较好的实验结果和所期望的可行性。

## 2 主备倒换

高端路由器实现了对 UPC 板 1:1 的保护,分为主备 UPC,在主板出现问题后,备板能及时替代主板工作,而不影响正常在线业务。主备倒换(见定义 1)需要确定主备倒换的状态、触发主备倒换的方式和主备倒换的性能等方面,才能够确保主备倒换按设计进行,具体分析如下。

**定义 1:**主备倒换 在主板出现故障或人为触发倒换之后,备板按照主板的运行,同时原主板自动复位,这个过程就是主备倒换。

### 2.1 主备状态确定

在双主控环境下,主控板的状态是主用还是备用,是在启动过程中由硬件决定的。一般情况,设备会首先选择小槽位号的主控板作为主用主控板(双主控启动环境下,硬件 FPGA<sup>[6]</sup>会在大槽位号主控板设置一个延时,使它稍后启动)。

初始启动时,两块主控板都处于备用状态,各自进行软件的启动。小槽位号主控板在启动到一定阶段时

将自己的单板状态位置设为正常,并且同时检测另一块主控板的状态是否正常;而大槽位号的主控此时会有一个延时,再检查另一块主控板是否正常,并且将自己的单板状态位置设为正常。这样,在小槽位号主控板的状态变为正常时,大槽位号主控板的状态还未正常,因此小槽位号主控板的状态变为主用;大槽位号主控板在延时过后检查,发现另一块主控板的状态为主用,便将自己的状态设为备用。

### 2.2 主备倒换触发方式

备 UPC 到达实时接收数据状态后,如果检测到倒换通知,则切换为主 UPC。检测通知是中断触发的,主备倒换的硬件切换时间在毫秒级。主备切换的触发因素主要有以下三类:(1)命令行执行主备倒换命令,强行切换;(2)主用主控板硬复位或者被手工拔出,引起主备倒换;(3)主 UPC 上发生软件异常重启,从而触发倒换。

对于以上三种触发方式,备 UPC 感应的都是一样的,都是在硬件中断触发的,状态切换时间在毫秒级。主控板和备用板都会周期发送握手机报,主控板或备用板在设定的周期内没有收到对方的握手机报,就会认为主备通信异常,重新复位备用板。

### 2.3 主备倒换性能

通过批量备份和实时备份过程,主用主控板的配置信息及时地备份到备用主控板。在主备倒换时机成熟时,备用主控板上已经保存有主用主控板完备的配置信息内容,并已经配置同步成功。因此,在主备倒换时,对于配置层,可以实现完全平滑的过渡,不需再额外收集或同步其他内容。

在平滑的过程中,为保证业务无间断转发,业务板上涉及各类转发表并不是删除后重新学习。在主控板进行的数据收集和同步过程中,基本维持原有业务板数据不变,只更新改变部分并进行刷新。控制软件与转发软件需要严格分开,控制软件负责整个设备的控制与管理(包括路由的发现和下刷等),而转发软件负责单纯的数据转发业务,两种软件由各自的处理器进行处理,功能完全独立。当控制软件重启(硬件或者软件故障)或者重新加载(软件升级)时转发业务可以不中断。在控制软件重启或者重新加载不影响转发软件的正常运行的情况下,只要保持网络拓扑稳定,正在重启的路由器在转发路线上进行数据转发,是完全可行、可靠的。

## 3 基于热备份的主备倒换在高端路由器中的应用

高端路由器<sup>[7]</sup>由于所处网络位置非常重要,不允

许出现单点故障,设备一般都配备两块主控板,分别称为主 UPC(Master)和备 UPC(Slave)。其中,主 UPC 作为控制平面的核心,与外部和业务板进行业务通信,完成系统内各模块的正常功能;而备 UPC 则只作为主 UPC 的备份,及时与主 UPC 同步进行热备份(见定义 2),与外部和业务板没有通信。当主 UPC 发生故障时,系统自动进行主备切换,由备 UPC 接替主 UPC 的工作,保证业务的正常运行。

定义 2:热备份<sup>[8,9]</sup> 上电时硬件根据两块 UPC 的插入顺序区分出主板和备板,完全启动之后,业务都运行在主板上,备板不参与业务控制,但主板会将自己保存的信息同步到备板上,这个过程称之为热备份。

主 UPC 出现故障后,发生主备倒换主 UPC 重启后变为备 UPC,并经过数据平滑阶段、热备份初始化阶段、批量备份阶段和实时备份阶段四个阶段,各阶段所发生的主备 UPC 间数据交互情况具体分析如下。

### 3.1 数据平滑阶段

主备倒换后,备用板升为新的主用板,会通知各个模块向业务板进行数据收集和同步,这个过程称之为数据平滑。在平滑过程中,各模块主动与业务板进行通信,在硬件状态、链路层状态、配置数据三个方面进行确认和同步,以保证整个系统维护的数据和状态是一致的,从而确保主备切换之后,系统能够正常运行。平滑阶段结束,新的主用板才称之为完全意义上的主用主控板。主控板包括主处理器单元(RPU)和协议处理单元(RPU)。平滑状态时备 UPC 的状态图如图 1 所示。

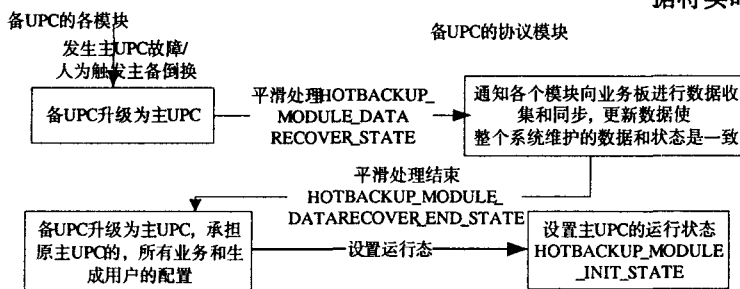


图 1 平滑状态时备 UPC 的状态图

### 3.2 热备份初始化阶段

主备 UPC 中 RPU 协议进程在上电后分别开始热备份控制模块和各个平台子模块的热备份初始化。平台初始化后,备 RPU 会立即向主 RPU 发起注册请求,主要目的是向主板告知备板在位,可以进行热备份。只有在主板的热备份状态处于初始化或空闲状态时才能接受备板注册,否则将复位备板重新开始热备份初

始化和注册的过程。在 Bras 子模块的热备份初始化阶段,主 RPU 会向主 MPU 和所有线卡发送初始化通告消息,备 RPU 会向备 MPU 发送初始化通告消息,都成功后表示初始化完成,可以开始同步。

### 3.3 批同步阶段

备用 UPC 启动后,由于此时主 UPC 和备 UPC 的差异比较大,主 UPC 会将当前需要备份的数据和状态一次性同步到备 UPC,而后当主板上的数据发生变化时就采用实时同步机制将其同步到备板上,因此批同步是一次性的,批同步完成之后不可能再一次进行批同步,除非发生主备倒换或者重启,这个过程称之为批量备份,批量备份时间的长短取决于需要备份数据量的大小。批量同步过程如图 2 所示。

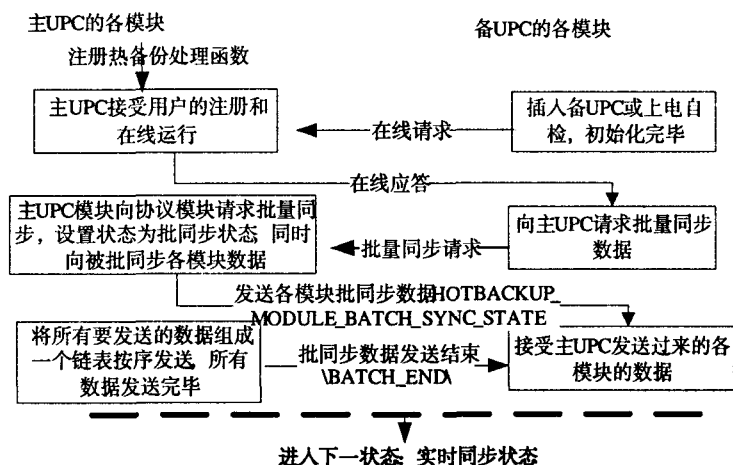


图 2 批同步阶段的主备 UPC 交互情况

### 3.4 实时同步阶段

批量备份过程结束后,系统进入实时备份过程,在此过程中,当主用主控板备份数据发生变化时,备份数据将实时同步到备用板,这个过程通常很快。实时同步就相当于变化同步,由备板向主板发起请求,主板建立定时器,定时向备板同步数据,同时通告 MP 及 NP。平台的同步操作都是在主备板对应 CPU 上的协议进程之间进行的。

综上所述,高端路由器的主备 UPC 经过平滑阶段、初始化阶段、批同步阶段和实施同步阶段等四个阶段,主 UPC 经过上述过程成为备 UPC,并实时/定时的备份数据到备 UPC,原备 UPC 变成主 UPC,承担原主 UPC 与用户的数据和业务往来。在保证故障持续时间最小化的同时,提供了业务不损失的特性。同时系统采用分布式结构,将业务口只存在于业务板上。在系统单板发生故障并倒换后,业务板继续处理业务不会重新启动。热备份保证主备系统单板之间的数据和状态始终一致,因而

业务单板也感觉不到系统单板发生倒换,再加上协议状态的一致,因此从业务板侧保证业务不会丢失。主备倒换发生时,路由板不发生变化,允许报文继续转发,从而在路由板侧保证了业务不损失。

4 实验验证

文中测试需要的测试环境<sup>[10]</sup>:中兴通讯 ZXUAS10800E 高端路由器一台;PC 机三台;串口线一根(连接);网线若干。高端路由器的基本配置为:两块 UPC3 主控板(采用主备方式);两块 SFC3 交换板(采用主备方式);两块 NPCIX 线卡,一块 BIC 接口卡,一块 10 端口的千兆 SFP 接口板(BT-10x1GE-SFP)。

文中的网络拓扑结构如下图 3 所示。在超级终端上安装上中兴数据综合测试平台 ZXDTTP 和 TFTP/FTP 服务器软件。通过数据综合测试平台 ZXDTTP 在高端路由器上配置 PPOE 上线拨号命令,PC1 通过 PPPOE 拨号方式和 PC2 建立通信。

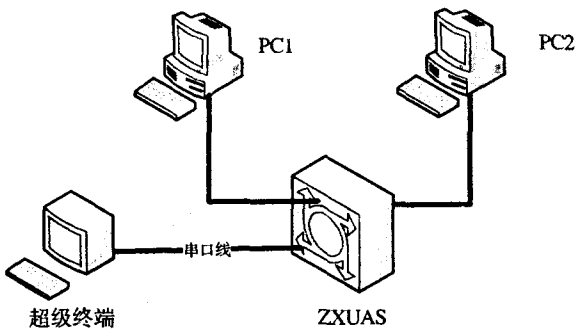


图 3 文中所采用的网络拓扑结构

通过 TFTP/FTP 服务器将 ZXUAS10800E 软件部分导入高端路由器中,让整个设备运转起来。输入用户名和密码后,开始正式配置路由器。

ZXUAS10800E 的主备倒换的命令为:redundancy force-switchover;在高端路由器 ZXUAS 上存在通信流量的时候,实施主备倒换。当出现如下情况的时候,说明备板还没有准备好:(1)没有备控板 UPC3;(2)主备倒换太频繁的时候,备板还没有完成上一次主备倒换。当主备倒换完成后,数据综合测试平台会出现如图 4 所示情况,表明主备倒换实施成功。

```
2822_K6#
2822_K6# redundancy force-switchover
proceed with switchover to standby UPC? [yes/no]:yes
synchronizing...../OK!
*****
Welcome to ZXR10 Carrier-Class High-end Router of ZTE Corporation
*****
2822_K6>
```

图 4 主备倒换实施成功

文中将基于热备份的主备倒换技术和基于冷备份的主备倒换技术分别应用于高端路由器中,并得出相应阶段的运行时间,所得时间具体如表 1 所示。

表 1 高端路由器中主备倒换技术的比例图

Item	File Time(s)		Mp 的配置恢复	RP 路由学习聚合时间	系统高可靠性
	备用板的加载和启动时间				
方案 1	17		5	3	63%
方案 2	1.7		2.3	1.9	99.9%

注:方案 1:基于冷备份的主备倒换技术

方案 2:基于热备份的主备倒换技术

从上述表格中,这些数据正说明了运用基于热备份的主备倒换技术和基于冷备份的主备倒换技术的软件进行测试的结果符合实际要求,经过热备份的主备倒换后的高端路由器,其各个阶段的主备倒换时间相对较短。当主 UPC 发生故障时,系统自动进行主备切换,由备 UPC 接替主 UPC 的工作,保证上线用户不掉线,接入的业务正常运行。因此,实验结果表明文中基于热备份的主备倒换技术在高端路由器中的应用是非常有效的。

5 结束语

文中采用将基于热备份下的主备倒换技术运用在高端路由器中,大大的缩短了系统的故障持续时间;同时又提供了业务不损失等特性,大大提高了系统的可靠性,它能够给用户带来实实在在的商业价值,包括提高用户的工作效率,节省生产成本,增强竞争力。通过与基于冷备份的主备倒换技术的比较,显示出基于热备份的主备倒换技术的可靠性、稳定性、及时性。由于交换网板的主备倒换并没有涉及到模块的数据备份的情况,在此就没有做更多的解释。实践证明,基于热备份的主备倒换在高端路由器中的应用,对运营商和企业激烈的竞争环境中取得成功都会有很大帮助。

参考文献:

[1] SystemMaster. 华为高端路由器设备级可靠性[EB/OL]. 2006-11-23. www.huawei.com.cn.  
[2] 李晓云,胡之惠. 基于热备份的主备倒换在高端路由器中的应用[J]. 现代电子技术,2005(11):71-72.  
[3] 黄澄,周恩松,李阳贤. 一种可自动实现主备倒换的双向光通信设备[M]. 北京:国家知识产权出版社,2002.  
[4] Feng Qingfeng, Jia Xiaozhu, Liu Shizhong, et al. A Secure Standby Route Mechanism for Ad Hoc Networks[C]// IC-CT2008. 11th IEEE International Conference. [s.l.]: IEEE Communications Magazine,2008:77-80.

(下转第 179 页)

表5 实例部分仿真结果

编号	p	q	r	组数	$\lambda$	$\mu$	繁忙率 p	队长 L
1	9	3	6	3	0.8	0.4	90.8%	2.87
2	10	5	8	3	0.9	0.35	87.7%	2.79
3	11	4	8	3	0.9	0.4	93.1%	3.91
4	9	5	7	3	0.8	0.5	83.5%	3.20
5	12	4	8	4	0.95	0.4	82.8%	3.49
6	11	7	6	3	0.95	0.4	93.4%	8.39

相关变量定义如下:

```
color INT= int;
val p,q,r :INT;
colset E = with e;
colset worker = index w with 1...p declare ran;
colset facility = index f with 1...q declare ran;
colset module = index m with 1...r declare ran;
fun repair(group)=3*p+ +1*f+ +2*m;
fun judge(P,L)
    if (P>=80% and P<=90% and L<=4)
then 1'e else empty;
```

仿真结果如下:

```
Statistics
-----
Occurrence Graph
Nodes:14
Arcs:13
Secs:1
Status:Full
Boundedness Properties
-----
Home Markings: All
Liveness Properties
-----
Dead Markings: None
Dead Transitions Instances:None
Live Transitions Instances: All
```

由仿真结果可知,模型中不存在死锁,具有安全性和可达性。通过调整  $\lambda$  和  $\mu$  的大小进行仿真后,第2

组、第4组、第5组数据符合任务要求,仿真通过。仿真结果也表明,如果供应要求更加频繁或者平均运输时间变长,则需要增加维修力量(见第5组数据)。

4 结束语

文中分析了战场装备抢修流程、petri 网在战场抢修流程建模中的优势,建立了基于 petri 网的战场抢修模型,并通过 CPN-Tools 实例仿真,验证了模型的可行性。但模型还不够细致,不少决策还需要人工干预,对于判断条件的细化等工作还有待于下一步的研究。

参考文献:

[1] 齐胜利,武 昌,杨 懿.基于分层着色 Petri 网的装备战场抢修系统建模[J].系统仿真学报,2005,17(10):2327-2329.

[2] Donateli S, Sereno M. On the Product Form Solution for Stochastic Petri Nets[C]//In: Proc of IEEE In FOCOM 2004. San Francisco, CA, USA: [s. n. ],2004.

[3] 陈春良,王岩磊,孙盛坤. HTCPN 在装备保障业务流程建模与优化中的应用[J]. 系统仿真学报,2008,20(10):2746-2749.

[4] Peter J H. Estimation Methods for Nonregenerative Stochastic Petri Nets[J]. IEEE Trans Software Engineering, 1999, 25(3):218-236.

[5] 杨美玲,朱建冲.基于 petri 网的装备维修流程通用模型研究[J].系统仿真学报,2005,17(S1):128-130.

[6] 曹继平,宋建社,王正元.战时装备维修保障资源优化调度方法研究[J].系统工程与电子技术,2007,29(6):915-919.

[7] 封利民,龚传信,高晓利,等.装备保障链中的信息流分析[J].海军航空工程学院学报,2006,21(6):675-679.

[8] 王建军,花兴来.基于 petri 网的装备物流管理模型[J].空军雷达学院学报,2002,16(1):27-30.

[9] 谭雪花,王华伟.飞机维修方案优化支持系统框架研究[J].计算机技术与发展,2008,18(11):183-186.

[10] 原菊梅,侯朝桢,王小艺,等.基于随机 Petri 网的可修系统可用性模糊评价[J].计算机工程,2007,33(8):17-19.

(上接第 175 页)

[5] 李 水,武成宾.一种实现主备倒换的装置和方法[M].北京:国家知识产权出版社,2005.

[6] 刘玉倩.高端流量管理中流量整形和标签处理算法的 FP-GA 实现[D].成都:电子科技大学,2005.

[7] Gottlieb Y, Peterson L. A Comparative Study of Extensible Routers[C]// 2002 IEEE. [s. l.]: IEEE Communications Magazine,2002:51-62.

[8] 曹 琼.高端路由交换机中的路由热备份技术应用研究[D].成都:西南交通大学,2005.

[9] Lundberg J. Routing Security in Ad Hoc Networks. URL: [EB/OL]. 2000. <http://citeseer.ist.psu.edu/lundberg00routing.html>.

[10] 宋 芳. ZXUAS 10800E(V2.8.01)电信级宽带接入服务器用户手册[M].深圳:中兴通讯,2008.