文章编号:1671-4598(2017)07-0257-03 DOI:10.16526/j.cnki.11—4762/tp.2017.07.064 中图分类号:TP3 文献标识码:A

基于国产 CPU 的并行冗余计算机系统研究

黄 晨,汪文明,张义超,岳 玮
(北京宇航系统工程研究所,北京 100076)

摘要:目前国家不断推进的国产自主可控信息系统建设,其核心国产计算机系统由于技术成熟度低、市场推广晚等原因,暴露出可靠性低、稳定性差的问题,直接导致系统功能无法成功应用;围绕国产化计算机系统的并行冗余架构开展研究,通过计算机系统架构的软硬件设计,以及高速缓存一致性架构、高速互联总线和三状态转换机制方法的应用,基于国产 CPU 并行冗余计算机系统,可以有效消除备份计算机系统进行当班切换时,存在的切换时间延时和切换过程数据丢失的问题;通过试验验证,该系统可以完成计算机系统中 CPU 处理器和功能桥片故障模式的容错处理,并保障信息数据的完整性和实时性,有效提高设备中计算机系统的工作可靠性与稳定性。

关键词:并行冗余计算机系统;HyperTransport 总线;高速缓存一致性协议

Research of Parallel Redundant Computer System Based on Domestic CPU

Huang Chen, Wang Wenming, Zhang Yichao, Yue Wei
(BeiJing Institute of Aerospace Systems Engineering, Beijing 100076, China)

Abstract: The construction of the domestic information system at present the country continues to advance, the core of the domestic computer system due to the low degree of technological maturity, market promotion and other reasons later exposed, low reliability, poor stability, led directly to the system function cannot be successfully applied. The parallel redundant architecture research on localization of computer system, the hardware and software design of computer system architecture, application and conversion mechanism method of cache coherence architecture, high-speed interconnection bus and three state, domestic CPU parallel computer system based on redundancy, can effectively eliminate the backup computer system on duty when switching the switch time delay the problem of data loss and switching process. Through the test, the system can complete the fault-tolerant computer system of CPU processor and the function of bridge chip fault modes, and ensure the completeness and timeliness of information data, effectively improving the working reliability and stability of the computer system in equipment.

Keywords: parallel redundant computer system; HyperTransport bus; cache coherent protocol

0 引言

近年来,国家在银行、政府、军队等众多领域加快推进国产自主可控替代计划,构建安全可控的信息技术体系。作为信息技术体系的核心,计算机系统通常基于龙芯、申威等国产处理器和中标麒麟等国产操作系统进行研制。然而国产计算机系统通常由于技术成熟度低、市场推广晚等原因,暴露出可靠性低、性能不足等问题。

然而计算机系统作为电气系统的控制中枢与数据中心,其地位和作用都是举足轻重的,计算机系统丝毫的差错与谬误的出现,轻则能够造成任务的延误,重则可能危及到全局成败。

为了提升国产计算机系统的可靠性,目前的电气系统在应用时通常都采用双计算机系统的设计,主从计算机系统采用相同的初始设置,一旦主计算机系统出现工作异常,经指挥决策后,关闭主计算机系统,启动从计算机系统来代替原计算机,继续完成相关工作,保证任务的继续完成。然而这种工作模式存在着一定的弊端,首先,该模式下,由主计算机系统到从计算机系统的切换需要一定的等待时间,然而在某些特定工作场合下,设备的工作停滞是难以承受的;其次,计算机系统切换后,必定存在一定程度的数据丢失,或者进程丢失,往往会带来不可挽回的损失。为了解决备份计算机系统切换所存在的隐患,本文对基于国产 CPU 处理器的并行冗余计算机系统进行针对性研究,旨在提升国产计算机系统的工作可靠性。

1 系统结构及原理

本并行冗余计算机系统结构组成如图 1 所示,包含主从两个计算机系统,两个计算机系统的硬件组成基本相同,每个计算机系统的结构均按照机箱插卡模块的样式完成设计,通过 VPX 接口插入设备机箱的插槽之中,两个模块之间的互连总线经 VPX 接口与机箱背板完成走线。

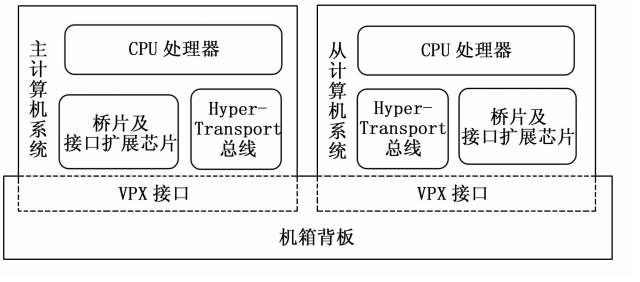

图 1 并行冗余计算机系统结构图

两个计算机系统构成分布式存储结构。多个存储单元与处理单元分布在整个系统之中,通过专门的互连网络结构连接在一起组成分布式的共享内存空间。每一个处理器可以访问自己的存储器,也可以访问其他处理器的存储器或共享的存储器。通过硬件维护的高速缓存一致性协议,使得各个处理器对于本地及远程存储单元的影响都是统一的[1]。

收稿日期:2017-02-13; **修回日期**:2017-03-31。

作者简介:黄 晨(1986-),男,山东龙口人,工程师,主要从事重点研究信息应用系统线路综合设计方向的研究。

正常工作时，由主计算机系统的芯片组完成系统控制与数据管理工作，从计算机系统的芯片组通过 16 位的 HT（HyperTransport）总线跟踪主计算机系统处理器的进程操作及工作状态，包括 CPU 进程信息、电子盘存储信息等，并在从计算机系统的电子盘之中同步备份主计算机系统的电子盘中的操作数据，利用高速缓存一致性协议，保证两个计算机系统的处理器核、内存以及电子盘之间的数据与缓存状态均保持一致。正常工作时信号通路如图 2 所示。

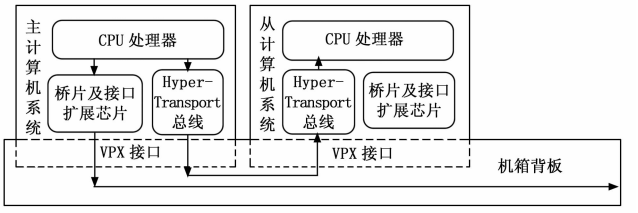

图 2 正常工作时信号通路示意图

当主计算机系统的桥片或者接口芯片出现问题，导致工作异常时，主计算机系统的处理器可以通过 HT 总线将信息或指令传输给从计算机系统的处理器，通过从计算机系统所具有的桥片或接口芯片继续完成系统工作或任务。桥片故障时信号通路如图 3 所示。

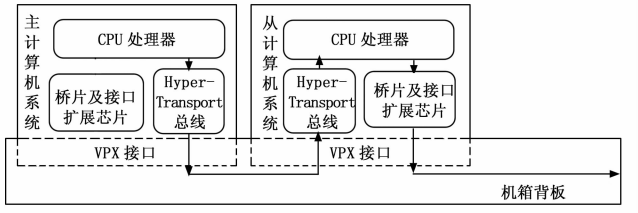

图 3 桥片故障时信号通路示意图

当主计算机系统的处理器出现异常时，从计算机系统的处理器将依靠 HT 总线接口的硬件所支持的系统高速缓存一致性维护，达到系统工作的近似无缝切换，保证系统工作的正常，CPU 故障时信号通路如图 4 所示。避免了由于计算机系统的工作异常，导致工作流程终止，甚至重要数据丢失等难以挽回的危害，从而有效提高系统可靠性，实现了计算机系统的热备冗余备份。

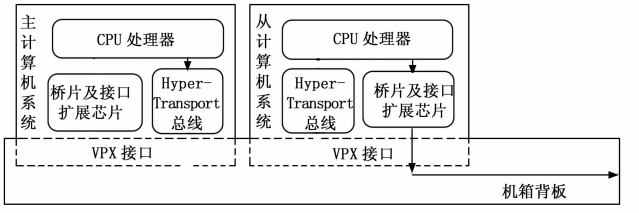

图 4 CPU 故障时信号通路示意图

2 系统硬件设计

出于安全性和自主性考虑，本计算机系统采用中国科学院计算所自主研发的龙芯 3A CPU 处理器作为核心处理器开展研究。

本并行冗余计算机系统包含主备两个计算机系统，构成双冗余模式，如图 5 所示。

两个计算机系统的硬件组成设计基本相同，均由 CPU 单元、南北桥单元、存储单元、接口单元 4 部分组成。CPU 单元主要包含龙芯 3A 四核处理器；南北桥单元主要由 RS780E 北桥芯片和 SB710 南桥芯片构成；存储单元则包括 DDR 内存、启动配置 ROM 和电子硬盘 3 大类；接口扩展芯片则依据系统需求具体设计，可以包含网络信号、串并行接口总线、视频信号、音频信号等等。两个芯片组之间通过 16 位 HyperTransport 总线接口实现两个 CPU 间互连，利用高速缓存一致性协议，保证两处理器核间数据及指令缓存的一致性。

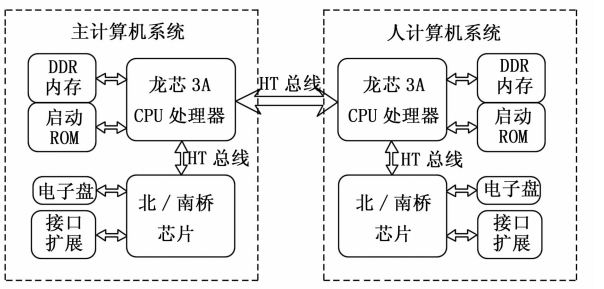

图 5 并行计算机系统架构组成

并行计算机系统利用龙芯 3A 处理器内部结构搭建并行互连体系结构，结构示意图如图 6 所示。

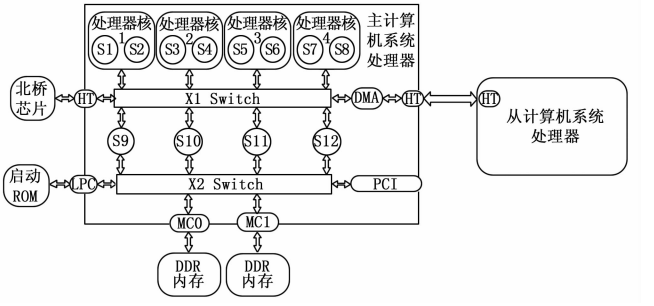

图 6 并行互连体系架构示意图

CPU 芯片内含有 4 个处理器核，每个处理器核内具有独立的一级数据缓存和一级指令缓存，一级缓存由各处理器核私有；芯片内的二级高速缓存被片内的所有处理器核共享，根据地址可以分成供并行访问的 4 个交错体。每个芯片具有两级 AXI 交叉开关，X1 和 X2，处理器核通过 X1 访问二级缓存块，二级缓存通过 X2 访问两个内存控制器，内存地址分布和二级缓存地址分布一致，以简化二级缓存和内存之间的通路并降低二级缓存访问失效的延迟[2]。

HT 控制器通过 DMA 控制器与 X1 相连，由 DMA 负责 IO 的 DMA 控制及片间、IO 访问和处理器访问之间的数据一致性维护，由于交叉开关不区分处理器端口和互连端口，HT 接口也可视为特殊的处理器处理。

互连结构体系中的存储单元与处理器节点可以分布在系统内不同位置，其中的存储单元可以通过互连网络被各个处理器所共享。互连体系结构中采用 X－Y 路由算法，点到点的路由是固定的，可以保证点对点数据包严格有序传输；每个模块均会被分配给一个与逻辑位置相关的全局 ID 号，以决定转发端口路由。

CPU 级间互连的 HyperTransport 总线是基于报文的、点对点串行链路结构，将芯片内部低频并行信号通过发送端物理层 DA 转换接口处理为高频串行信号，每个时钟沿传输并行信号中的一位，通过 LVDS 信号传输线，到达接收端，再由接

收端物理层 AD 转换接口还原为低速并行信号。HT 总线信号分为三类：

1）链路信号：32 位 CAD 传输信号、1 位 CTL 控制信号、4 位 CLK 时钟信号，用于控制、传输数据；

2）复位信号：PWROK 和 RESET 信号，用于初始化和复位链路；

3）管理信号：LDTSTOP 和 LDTREQ 信号。

HT 总线采用单向点对点传输技术，将 CAD 信号分成两组，按照不同方向单向传输，可达 GB/s，在处理器芯片上按照接收端和发送端分别处理，传输效率与有效带宽均远大于双向信号总线，简化板级设计工作。

HT 总线采用包交换方式进行信号传输，将总线操作分为控制包和数据包两大类，控制包还包括读命令包、写命令包、读响应包、写响应包，每组传输总线使用一位控制信号线来区分传输的为控制包还是数据包。

HT 总线采用虚通道技术，将 HT 协议划分为三种不同的数据流：无响应请求通道、需响应请求通道、响应通道，将一个物理链路划分为若干相互独立的逻辑通道。针对这三种通道，还设置了六种缓冲区类型，实现每个虚通道的缓冲流控自行维护，避免命令之间的互锁，提供了乱序执行的基础条件，实现不同数据流在同一信号线上的并发传输，提高总线传输的效率和性能[3]。

3　系统软件设计

并行冗余计算机系统采用基于目录的高速缓存一致性协议，系统内共享二级缓存与各个处理器核内私有指令缓存和数据缓存之间的数据一致性，由系统中共享存储层次的目录维护。目录与其数据的存储位置相关联，目录的管理由各个存储单元所在的存储节点目录决定，目录控制器存储的相关一致性信息包括存储单元的状态和拥有此存储单元备份的处理器号。

每个共享存储单元的状态由这个共享存储单元自身维护，这个状态标识了当前这个共享存储单元在其它处理节点中私有缓存的使用方式，并且记录了哪些处理器的私有高速缓存中拥有该共享存储单元的备份。系统中任意一个处理器需要对共享存储单元进行操作时，都直接与这个共享存储单元本身交互，由这个共享存储单元的状态控制器再与其它处理器中的缓存备份交互，通过一致性协议控制整个系统中的缓存数据一致性。

龙芯 3A 处理器的一级缓存由各处理器私有，二级缓存和内存采用全局编制，由所有处理器共享。缓存块的目录信息在宿主二级缓存中维护，目录使用 32 位宽度的位向量来记录拥有每个缓存备份的一级缓存编号，因此硬件能自动维护各指令和数据缓存之间的一致性，同时也就维护了全系统各级存储结构间的数据一致性。

一级缓存块采用三状态转换机制，无效状态、共享状态和独占状态，无效表示这个缓存块中没有有效数据，独占表示这个缓存块中的数据有效且未经修改，共享表示这个缓存块中的数据已经被修改而且还未写回下级缓存。三种状态的相互转移图如图 7 所示。

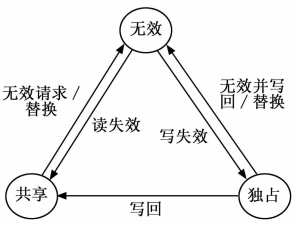

图 7　并行互连体系架构示意图

4　实验结果与分析

并行冗余计算机系统工作主频为 800 MHz，将系统工作周期设定为 125 μs，主副计算机系统同步周期设定每间隔 125 ms 主副计算机系统完成一次硬盘数据同步备份。通过在计算机系统二次供电母线增加功能开关的方式，进行故障注入，分别将主计算机系统的 CPU 处理器或接口桥片的供电关断，实现应急故障模拟。经测试验证，该并行冗余计算机系统可以通过加载 CPU 寄存器信息和内存信息，实现计算机系统运算及通信内容的无缝切换，而依靠 CPU 芯片内的 Cache 一致性设计和三状态转换机制，成功保障了寄存器信息与内存信息的高度同步性。

5　结束语

本文首先对于并行冗余计算机系统的功能应用进行了介绍与分析，然后又介绍了并行冗余计算机系统的构成，并对于该系统的硬件和软件设计的关键技术进行了细致全面的分析，为该系统的实现提供了基础与参考，最后通过故障注入方式，对于系统的容错备份功能进行了验证。并行冗余计算机系统在的应用，对于国产自主可控计算机设备的可靠性改进具有重要意义，为任务的成功完成提供有效支撑。

参考文献：

[1] Chen D，Su H，Yew P. The impact of synchronization and granularity on parallel systems [C]. Proceedings of 17th Annual International Symposium on Computer Architecture，1990，239－248.

[2] 王焕东，高　翔，陈云霁，等. 龙芯 3 号互联系统的设计与实现 [J]，计算机研究与发展，2008 (45)：2001－2010.

[3] HyperTranspon Technology Consortium. Hyper TransportTM1/O Link Specification Revision 1.03. http：//www. bypertransprot. org/default. cfm? page＝Hyper [EB/OL]. TransportSpecificationslx，2008－11－20.

(上接第 256 页)

[4] Patel Hasmukh S，Hoft Richard G. Generalized techniques of harmonic elimination and voltage control in thyristor inverters：part Ⅰ－Harmonic elimination [J]. IEEE Transactions on Industry Application，1973，9 (3)：310－317

[5] Patel Hasmukh S，Hoft Richard G. Generalized techniques of harmonic elimination and voltage control in thyristor inverters：part Ⅱ－Harmonic elimination [J]. IEEE Transactions on Industry Application，1974，10 (5)：666－673

[6] 李　伟，马志文，蔡华斌，等. 无二次滤波环节的单相四象限整流器输入电流控制研究 [J]. 铁道学报，2014，36 (5)：28－36.

[7] 王兆安，黄　俊. 电力电子技术（第四版）[M]. 北京：机械工业出版社，2002.

[8] 张永昌，赵争鸣，张颖超. 三电平逆变器 SHEPWM 多组解特性比较实验 [J]. 电工技术新学报，2007，22 (3)：60－65.

[9] 张永昌，赵争鸣. 三电平逆变器 SHEPWM [J]. 电工技术学报，2007，22 (1)：74－78.

[10] 程佩青. 数字信号处理教程 [M]. 第三版. 北京. 清华大学出版社，2010.