

Torus 网络中基于标志位的容错路由

李 银¹, 梁家荣¹, 伍华健²

(1. 广西大学计算机与电子信息学院, 南宁 530004; 2. 玉林师范学院数学与计算机科学系, 玉林 537000)

摘要: 针对 Torus 结构的多处理机系统中容错路由的问题, 提出标志位概念, 给出一个基于标志位的容错路由算法。存储于 Torus 网络中各节点的标志位记录系统中的故障信息, 用于判定消息的源节点和目的节点之间是否存在最优通路。标志位的赋值可以通过与邻节点间的信息交换完成。

关键词: Torus 网络; 标志位; 容错路由

Fault-tolerant Routing Based on Flag Bit in Torus Networks

LI Yin¹, LIANG Jia-rong¹, WU Hua-jian²

(1. School of Computer, Electronics and Information, Guangxi University, Nanning 530004;

2. Department of Mathematics and Computer Science, Yulin Normal University, Yulin 537000)

【Abstract】 This paper presents a new concept, flag bit, for fault-tolerant routing on Torus multicomputers. It gives a fault-tolerant routing algorithm based on flag bit. Flag bit stored on each node of a Torus keeps faulty information and indicates whether there is an optimal path from the source to the destination. The values of flag bit of nodes can be determined by information exchange between neighbors.

【Key words】 Torus networks; flag bit; fault-tolerant routing

1 概述

目前很多商用或实验性多处理机系统都采用直接互连网络系统中各处理机节点的方式, 其中常用的一种网络拓扑结构是 Torus。它具有很多优点, 如结构简单、直径小、结构对称、可扩展性好、路径多样性等。随着 VLSI 和硬件技术的发展, 多处理机系统规模也越来越大, 系统中出现处理机故障或处理机间链路故障的可能性也随之增加, 因此, 容错路由策略的设计对多处理机系统也越来越重要。

多处理机系统中的故障可划分为节点故障模型和链路故障模型。在节点故障模型中, 所有消息都不能通过故障节点传递; 在链路故障模型中, 一条链路出现故障仅使以此链路通信的 2 个端点不能再使用此链路通信。本文同时考虑节点故障与链路故障, 提出的容错路由算法不仅适用于只出现 1 种故障模型的情况, 对 2 种故障模型同时出现这一更真实的情况也适用。

点对点通信是多处理机系统中一种非常重要的通信方式, 根据节点记录系统中故障信息的多少, 可将算法分为基于全局故障信息的容错路由算法和基于局部故障信息的容错路由算法。基于局部故障信息的容错路由算法需要的存储量和通信开销少, 但所能找到的最优通路与系统中实际存在的最佳通路有很大差距; 而基于全局故障信息的容错路由算法则相反, 虽能记录更多的最佳通路, 但存储和维护全局的故障信息需要大量的存储和通信开销。一个折中的方法是基于对故障信息的预处理, 通过与邻接点的信息交换来记录尽量多的关于最佳通路的信息, 它的代价是只记录所存在通路中的一部分, 因此, 这类方法的核心问题就是如何减少所舍去的最佳通路。文献[1]提出不安全节点的概念, 每个节点记录邻接点的状态(安全、不安全或故障), 由此得到一个简单的

路由算法。文献[2]修改了不安全节点的概念, 扩大了算法的容错能力。文献[2-3]分别提出了路由能力和安全向量的概念, 用来表示系统中各节点沿最优通路传递消息的能力, 所选通路长度在源节点处即可确定。文献[4]在文献[3]的基础上修改了安全向量的定义, 提出扩展安全向量的概念; 文献[5]在文献[4]的基础上, 改用矩阵来表示各邻节点的每条链路是否存在链路故障, 重新设计了对最优通路的计算方法, 提出了最优通路矩阵的概念。文献[6]在文献[5]的基础上修改了最优通路矩阵的定义, 提出了扩展最优通路矩阵的概念。文献[7-9]分别提出安全通路向量、极大安全通路向量、极大安全通路矩阵的概念。从不安全节点到极大安全通路矩阵, 这些算法的容错能力和沿最优通路传递消息的能力都有了一定提高, 但是它们都是针对超立方体网络提出的。

2 预备知识

n 维 Torus 网络 $T_{k_0 \times k_1 \times \dots \times k_{n-1}}$ 有 $k_0 \times k_1 \times \dots \times k_{n-1}$ 个节点, 它们的地址可表示为如下形式 $(a_0, a_1, \dots, a_{n-1})$, 其中, $0 \leq a_i \leq k_i - 1$, $0 \leq i \leq n-1$, 如果 2 个节点 $A(a_0, a_1, \dots, a_{n-1})$ 与 $B(b_0, b_1, \dots, b_{n-1})$ 满足如下条件: $a_i = (b_i \pm 1) \bmod k_i$ 且 $\forall j \neq i, a_j = b_j$, 则节点 A 与 B 互为邻节点。

定义 1 用 $Dist(A, B)$ 表示 A, B 间的 Hamming 距离, 若 2 个节点 A, B 间的路径长度等于 $Dist(A, B)$, 则称该路径为两者的最佳通路。

基金项目: 国家自然科学基金资助项目(60564001); 教育部“新世纪优秀人才支持计划”专项基金资助项目(NCET-06-0756)

作者简介: 李 银(1984-), 女, 硕士研究生, 主研方向: 计算机网络, 容错计算; 梁家荣、伍华健, 教授

收稿日期: 2008-10-22 **E-mail:** yinli020608@126.com

3 标志位赋值及性质

3.1 标志位的赋值算法

故障部件会导致消息传递过程中的延迟,为了在系统存在链路和(或)节点故障的情况下确定节点间是否存在最优通路,本文提出标志位的概念。

定义 2(标志位) 对于 Torus 网络中的源节点 A 与目的节点 B , 节点 B 的标志位记为 F_B , 其赋值如下:

$$F_B = \begin{cases} 1 & A \text{ 与 } B \text{ 之间存在最优通路} \\ 0 & A \text{ 与 } B \text{ 之间不存在最优通路} \end{cases}$$

从以上定义可以看出, $F_B=1$ 说明 A 与 B 间存在最优通路, 只要 A 与 B 存在至少一条最优通路 $F_B=1$, 根据标志位的值就可以在源节点 A 确定与目的节点是否存在最优通路, 如果存在则选择此路径以减少消息延迟。

为确定任意 2 个节点间是否存在最优通路, 进而对标志位进行赋值, 下文给出标志位的赋值算法。

为了便于叙述, 以二维 Torus 网络为例, 该方法很容易扩展到任意维 Torus 网络。如图 1 所示为二维 Torus 网络 $T_{5 \times 5}$, 给定源节点 $A(2, 2)$ 与目的节点 $B(4, 4)$, 可以计算其汉明距离: $H(A, B)=4$, 因此, 最优通路长度为 4。节点中的数据为当前节点与 A 节点的汉明距离, 节点左下的数据为节点的地址。

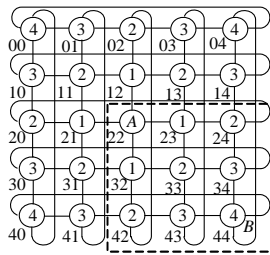


图 1 规模为 5×5 的 Torus 网络

定理 给定节点 A 与 B , 与节点 B 邻接且与节点 A 的汉明距离为 $H(A, B)-1$ 的节点 C , 在节点 B 与节点 C 之间链路正常的条件下, 节点 A 与节点 B 之间存在最优通路, 由此可推导节点 A 与节点 C 之间存在最优通路, 反之亦然。

证明 根据最优通路定义可容易证明: 节点 A 与节点 C 之间存在最优通路, 且 B 与 C 之间链路正常 $\Leftrightarrow A$ 与 B 之间存在通路; 又 $H(A, C)=H(A, B)-1$, $H(B, C)=1 \Leftrightarrow$ 节点 A 与 B 之间存在最优通路。即得证。

因此, 可以通过与满足条件的邻节点交换信息的方法对标志位进行赋值, 通过 $H(A, B)-1$ 轮同步信息交换就可以确定 F_B 的值, 假定每个节点能够获得其邻节点的状态。其赋值算法如下:

Algorithm F_B Values ()

```
{
    if (B 正常)
    {
        if(H(A,B)=1 && A, B 之间的链路正常)
            return 1;
        for(int i=0; i<4; i++)
        {
            //  $T_{k_0 \times k_1 \times \dots \times k_{n-1}}$  网络,  $k_i (0 \leq i \leq n-1)$  为奇数时最多
            // 有 2 个满足条件的邻节点, 为偶数时最多有 4 个满足条件的邻节点
            Search 与 B 邻接且汉明距离为 H(A,B)-1 的点 C;
            if(Algorithm  $F_C$  Values ()=1 && B, C 之间的链路正常)
                return 1;
        }
    }
}
```

```
}
return 0;
}
else return 0;
}
```

3.2 标志位的性质

标志位用来判定节点对之间是否存在最优通路, 因此, 每当给定一对源节点与目的节点, 都要计算其相应的标志位, 但是标志位的赋值是简单高效的, 只须与有限节点进行信息交换即可, 如图 1 所示, 对 B 节点标志位的赋值只须与虚线框内节点进行交换信息即可。事实上, 由于 Torus 网络的对称性, 可以把网络进行变形使其与图 1 相似, 进而确定其交换信息的范围, 如图 2 所示。

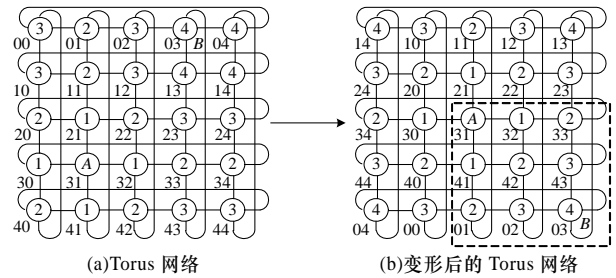


图 2 Torus 网络的变形

4 路由算法和实验结果

4.1 路由算法

基于标志位, 本文给出一个简单的容错路由算法, 在节点间存在最优通路的情况下, 将信息沿最优通路传递, 具体描述如下:

Algorithm F_Route ()

```
{
    //源节点 A, 目的节点 B
    if( $F_B=1$ ) //A 与 B 之间存在最优通路
    {
        send message with this minilength link;
        return (SUCCESS);
    }
    if( $F_B=0$ ) //找次优通路
    for(int i=0; i<4; i++)
    {
        寻找通过正常链路与 B 节点邻接的节点 C 且  $H(A,C) \geq H(A,B)$ ;
        Algorithm  $F_C$  Values ();
        if( $F_C=1$ )
        {
            send message with this link;
            return (SUCCESS);
        }
    }
    return (FALSE);
}
```

基于以上描述, 可知算法具有如下特性:

- (1)能在 $O(n)$ 时间内判定源节点与目的节点间是否存在最优通路。
- (2)若它们之间存在最优通路, 则消息沿此最优通路传递, 路径长度为 $H(A, B)$ 。
- (3)若它们之间不存在最优通路, 则消息传到合适的邻节点, 沿次优通路传递, 路径长度为 $H(A, B)+1$ (当 $T_{n \times n}$ 中 n 为

奇数时)或 $H(A, B)+2$ (当 $T_{n \times n}$ 中 n 为偶数时)。

4.2 性能分析与实验结果

对于给定源、目的节点,标志位记录了它们之间最优通路的信息,只要存在最优通路,路由算法选择最优通路进行路由,当不存在最优通路时,提出的路由算法所构造的路径长度也不超过最优通路路径长度+2。算法的不足在于,对于给定的每一对源、目的节点,都要重新计算其标志位,但从前面标志位的赋值分析得知,这一过程非常快,只须知道邻节点的状态,通过一小部分节点间的信息交换,即最多经过 $H(A, B)-1$ 轮信息交换即可完成对标志位的赋值。

为了验证所提的路由算法,本文模拟二维 Torus 网络中存在节点故障的情况,给定节点出错概率,随机构造一具有出错节点的这样规模的二维 Torus 网络,并随机选取 200 对源、目的节点对进行消息传递,以下是测试程序所使用的输入参数和输出参数介绍。

输入参数: $n \times n$ 表示二维 Torus 网络规模; p 表示节点出错概率。

输出参数: *PathFound* 表示算法找到路径的数量与总的路径数量的比值; *MinPathFound* 表示算法找到最优通路数量与源、目的节点间存在最优通路数量的比值。

算法 *F_Route()* 的容错性效率模拟结果见表 1。

表 1 算法 *F_Route()* 容错性效率模拟结果

$m \times n$	p (%)	<i>PathFound</i> (%)	<i>MinPathFound</i> (%)
50×50	1.41	99.19	100
100×100	1.04	99.09	100
200×200	0.75	99.06	100
400×400	0.54	99.03	100
500×500	0.48	99.04	100

可以看出,算法找到的最优通路数量与源、目的节点间存在最优通路数量相等,说明标志位记录了所有的最优通路,随着故障数目的增多,在节点出错概率不超过时 0.48% 时,

提出的路由算法找到源、目的节点间通路的概率至少为 99%。

5 结束语

理论分析与实验结果表明,本文给出的标志位记录了所有最优通路,即使不存在最优通路,也可选择次优通路进行路由,从而保证了系统出现故障时仍能高效地传递信息。为了提高网络的容错能力,如何结合概率的方法来研究 Torus 网络的容错性问题是下一步的研究方向。

参考文献

- [1] Lee T C, Hayes J P. A Fault-tolerant Communication Scheme for Hypercube Computers[J]. IEEE Transactions on Computers, 1992, 41(10): 1242-1256.
- [2] Chiu Geming, Chen Kaishung. Use of Routing Capability for Fault-tolerant Routing in Hypercube Multicomputers[J]. IEEE Trans. on Computers, 1997, 46(8): 953-958.
- [3] Wu Jie. Adaptive Fault-tolerant Routing in Cube-base Multi Computers Using Safety Vectors[J]. IEEE Transactions on Parallel and Distributed Systems, 1998, 9(4): 321-334.
- [4] 高峰,李忠诚,闵应骅,等. 超立方体多处理机系统中基于扩展安全向量的容错路由[J]. 计算机学报, 2000, 23(3): 248-254.
- [5] 高峰,李忠诚. 用最优通路矩阵实现超立方体多处理机系统的容错路由[J]. 计算机学报, 2000, 23(3): 242-247.
- [6] 田绍槐. 超立方体多处理机系统中基于扩展最优通路矩阵的容错路由[J]. 计算机学报, 2002, 25(1): 87-92.
- [7] 王雷,林亚平,陈治平,等. 超立方体系统中基于安全通路向量的容错路由[J]. 软件学报, 2004, 15(5): 783-790.
- [8] 王雷,林亚平,陈治平,等. 基于极大安全通路向量的超立方体容错路由[J]. 通信学报, 2004, 16(4): 130-137.
- [9] 王雷,林亚平,陈治平,等. 超立方体中基于极大安全通路矩阵的容错路由[J]. 软件学报, 2004, 15(7): 994-1004.

编辑 顾姣健

(上接第 79 页)

控制管理电路由 PowerPC 处理器实现,可完成静态配置、监控功能和交换机终端等功能;交换电路实现端口滤波管制、大容量端口缓存和虚链路交换等功能;辅助电路包括存储器、端口 PHY 等,实现数据和程序的存储以及数据的收发功能;电源及控制电路为 AFDX 接口板提供电源和控制功能。

交换电路是 AFDX 网络交换机的核心,为实现协议规定的延时、流控、带宽等性能,拟采用 ALTERA 公司高性能的 STRATIX III 的 EP3SL340 实现交换逻辑。它与片外 8 端口 PHY 结合,实现高速数据存储、校验和转发,达到全端口的线速数据交换。

交换 FPGA 采用存储转发方式,交换结构采用共享总线实现。全部端口的帧转发由统一的交换机构控制,共享统一的 VL 解析机构。当一帧数据进入交换端口后,交换端口根据 VL 识别号对 VL 进行地址解析,并在解析过程中得到该 VL 相关控制信息,对帧实施对应过滤策略。如果该帧无效,则该帧丢弃;否则在交换端口内部缓存。帧转发过程包括帧从接收 FIFO 传输到 Buffer 和帧从 Buffer 传输到发送 FIFO,前者采用端口请求、VL 解析机构响应方式,而后者则采用 VL 解析机构查询、端口响应方式。帧通过接收 FIFO 传输到缓存的同时进行交换转发控制,最后由接收端口、交换控制器、VL 解析机构三者共同决定当前帧是否有效以及转发到哪

个端口并交给存储管理模块进行相应的处理。

6 结束语

本文针对航空电子系统“综合化”发展的需求,提出了航空全双工交换式以太网网络技术,描述了 AFDX 通信网络的特点、结构、通信机制和协议栈,并给出了 AFDX 终端系统和网络交换机的设计思路。此网络能够满足新一代航电数据通信的高速、实时和可靠性传输,对 AFDX 通信网络的研究具有一定的实用参考性和推广性。

参考文献

- [1] 石改辉,张原. 下一代航空数据体系结构研究[J]. 电子工程, 2006, (4): 29-33.
- [2] ARINC Specification 664-7-2005 ARINC Aircraft Data Network Part7: Avionics Full Duplex Switched Ethernet(AFDX) Network[S]. Airlines Electronic Engineering Community, 2005-07-27.
- [3] AFDX Protocol Tutorial[Z]. Condor Engineering, Inc., 2005-05.
- [4] CES White Paper on AFDX[Z]. Creative Electronic Systems, Inc., 2003-11-25.
- [5] AFDX Test Responder Test Equipment User Manual[Z]. TechSAT GmbH, Inc., 2007.
- [6] Developing AFDX Solution[Z]. Actel Corporation, 2005-05.

编辑 顾逸斐

