SCI/RT总线的扩展子优先级流量均衡

崔 剑 刘兴春 李 铮

(北京航空航天大学 电子信息工程学院, 北京 100191)

摘 要: 建立了基于 OPNET软件的 SCI/RT(Scalable Coherent Interface for Real Time) TRAIN协议仿真模型,对流量均衡问题进行仿真分析.结果显示,当环路负载增加时,每个节点流量的增量是不同的.在环路处于重负载时,远离管理者的节点得到的流量近似于零,而靠近管理者的节点基本上占用了所有的环路流量. SCI/RT TRAIN协议提出管理者迁移流量均衡方式解决这个问题.通过对 8节点环路进行仿真,得出管理者迁移方法的均衡效果依赖于管理者在环路的配置方式的结论,当每个节点均具有管理者功能时,环路均衡效果达到最好.提出了扩展子优先级的流量均衡新方法. 8节点环路的仿真结果显示,该方法具有更好的流量均衡能力和更高的节点流量.同时,通过对扩展子优先级均衡方法在 6节点, 10节点和 12节点环路的仿真,证明了该方法在不同规模的环路中均可使用.

关 键 词: 数据通信系统; 互联网络; 接口; 流量; 实时系统

中图分类号: TP 393. 02

文献标识码: A **文 章 编 号**: 1001-5965(2010)01-0091-04

Simulation of SCI/RT extending sub-priority loading balance protocol

Cui Jian Liu Xingchun Li Zheng

(School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: To analyze the loading balance problem, the scalable coherent interface for real time (SCI/RT) TRAIN protocol was simulated by using OPNET. The results show that the incensement of the nodes throughput is quite different from each other, along with the ringlet load increasing. In case of the heavy ringlet load, the down stream nodes far from ticketmaster gained almost zero load, while the nodes near ticketmaster gained almost all the ringlet load. The SCI/RT TRAIN protocol gave the ticketmaster migration method to solve it. According to the simulations of a ringlet of 8 nodes, the ticketmaster migration method can distribute load between ringlet nodes but the performance of its function is very dependent on its configuration. When every node has the ticketmaster module, the method gets its best function. In this work, a new comparative method named extending sub-priority was proposed. It was proven that the extending sub-priority method gets better loading balance and higher node throughput than the previous ticketmaster migration method. On the other hand, based on the simulations of 6 nodes, 10 nodes and 12 nodes ringlets, the proposed extending sub-priority method was proven to be functional in ringlets with different scales.

Key words: data communication systems; interconnection networks; interfaces(computer); throughput; real time systems

SCI/RT(Scalable Coherent Interface for Real Time,可扩展的一致性接口的实时性扩展)协议是一种面向高性能计算机和集群系统的总线互联形式,具有传输速度快,延迟低,并行性好的优点,而且具有很强的实时传输能力. SCI/RT总线可以应用于要求高实时性和低延迟、高传输速率的航

收稿日期: 2008-12-22

基金项目: 航空基金资助项目(03F51065)

作者简介: 崔 剑(1980 -),男,北京人,博士生, cuijianw@ee.buaa.edu.cn.

空航天器机载总线、工业控制和地面战车互联总线[1-3].

SCI/RT总线中流量均衡对网络性能的进一步提高具有重要作用. SCI/RT TRAIN协议采用环路仲裁的方式进行优先级调度,由于同优先级不能抢占的问题,造成同优先级数据包的流量分布在不同节点之间并不均衡. 为了解决该问题,协议提出管理者迁移(ticketmaster migration)的方式[3]. 本文在 TRAIN 协议的基础上对环路中ticketmaster节点的数量和放置位置对均衡效果的影响进行了研究,并进一步提出了扩展子优先级的环路均衡方式.

1 SCI/RT TRAIN协议模型建立

1.1 仿真模型[4-7]

依据 SCI/RT TRAIN 协议的主要功能,使用OPNET仿真环境,建立的 SCI/RT TRAIN 协议节点模型如图 1所示.

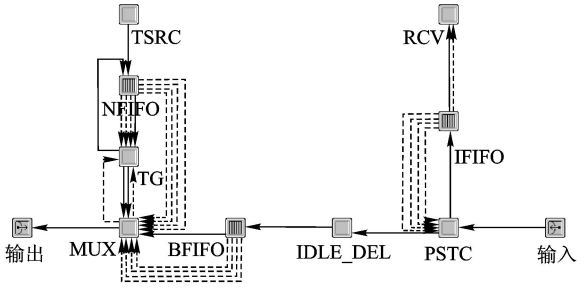


图 1 SCI/RT TRAIN协议的节点模型

图 1的模型中, PSTC(Packet Stripper /Ticket Changer)负责接收环路数据包,判断指向本节点的车票能否被接收. 将指向本节点数据包剥离环路,放入支持优先级操作的队列 IFIFO(Input FIFO). PSTC的进程模型如图 2所示.

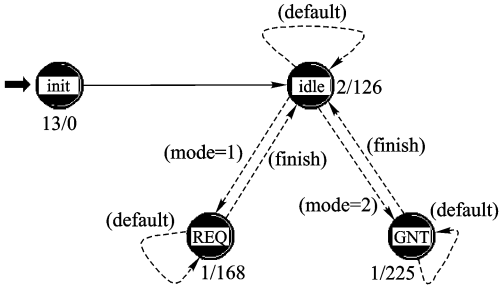


图 2 PSTC的进程模型

DEL_DEL 负责删除环路中多余的 idle符号,BFIFO(Bypass FIFO)负责节点输出数据包时暂存环路中的数据包.

节点产生数据包的过程由 TSRC(Transmit Source), NFIFO(Node FIFO), TG(Ticket Generator)和 MUX(Multiplexer)控制.

TSRC为数据源,根据不同仿真要求生成指定地址分布和优先级分布的 send数据包. NFIFO也是一个支持优先级操作的队列,负责将 TSRC产生的 send数据包缓冲. TG负责产生节点的 ticket数据包,并根据环路接收到的 ticket数据包决定是否发送对应的 send数据包. MUX负责维护 train在节点中的传输,负责控制 TG插入 ticket数据包,并计算 train的长度,在需要的时候截断 train. 同时 MUX也负责将 ticket数据包信息发送给 TG,并用 send数据包替换相应的 ticket数据包,图 3显示了 MUX的进程模型.

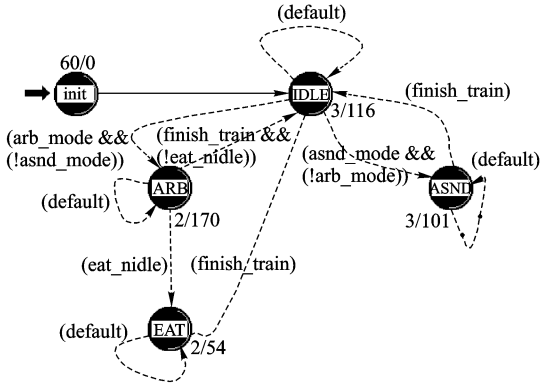


图 3 MUX模块的进程模型

1.2 扩展子优先级流量均衡方式

进行管理者迁移需要提高环路中节点的复杂程度,如不迁移管理者,需要在同优先级的数据包中引入抢占能力,且必须满足如下条件:

1) 能够周期性地改变同优先级的抢占方式,从而均衡流量.

2) 不能改变不同优先级数据包原有的数据包抢占模式.

3) 不能改变环路的工作方式和数据包格式.

基于以上考虑,引入子优先级:在每个节点内引入计数变量 lm_cycle,规定每次 request train到来节点时, lm_cycle增加 1, lm_cycle不超过环路中节点数 n. 则 lm_cycle计数值为 i时,节点内任一数据包的子优先级 P为:

$$P = M_{pri}(S)(i) \tag{1}$$

其中, S为数据包的源节点号. M_{pri}为优先级分配矩阵,对于一个 n节点环路, M_{pri}为 $n \times n$矩阵,且内容为:

$$M_{pri} = \begin{pmatrix} n-1 & n-2 & \cdots & 1 & 0 \\ n-2 & n-3 & \cdots & 0 & n-1 \\ \cdots & \cdots & & \cdots & \cdots \\ 1 & 0 & \cdots & 3 & 2 \\ 0 & n-1 & \cdots & 2 & 1 \end{pmatrix} \tag{2}$$

此后,节点判断数据包能否插入环路,不再以数据包本身的优先级作为判断依据,而是以数据包的优先级作为整数部分,扩展的子优先级作为小数部分,合并成新的优先级进行判断. 可见,相同优先级的数据包由于子优先级不同,依然能发生抢占,而且不影响不同优先级数据包的仲裁. 同时,由于具有相同 S 数据包的子优先级周期性变化,达到均衡流量的效果. 子优先级扩展过程只在节点内部进行,不对环路数据包和环路运行方式进行改变.

采用扩展子优先级的流量均衡方式,只需要对节点结构进行少量改进,即增加计数器及存储矩阵的存储器并对 MUX计算优先级的部分进行少量改进,而不需要在环路中增加新节点,也不需要对管理者进行改进. 然而, SCI/RT TRAIN协议建议的管理者迁移流量均衡方式却需要在环路中设置多个管理者. 本文后面的仿真指出,如果需要达到良好的流量均衡效果,需要为每个节点加入管理者功能,则当节点数量比较大时,设置多个管理者带来的硬件增加量要远大于为每个节点引入扩展子优先级流量均衡方式所带来的硬件增加量.

扩展子优先级流量均衡方式,只是根据 localmotive数据包进行流量均衡操作,不需要向环路输出数据包,相对管理者迁移方式需要发送管理者迁移命令数据包,不引入多余带宽占用,因此能够得到更高的有效流量,这在本文后面的仿真结果中得到证明.

2　SCI/RT TRAIN协议仿真

2.1　协议仿真参数

本文采用的仿真参数如表 1所示.

表 1　仿真主要参数

环路速度	链路延迟/ns	节点处理延迟/ns	TG缓存长度/包	localmotive的 size. size域	RCV处理速度(包)/ns
8Gbit/s	4	1	8	10×节点数	100

环路的基本仿真参数为:

环路中传递的 send包长度为 80 bit,每个节点的输出队列均为无限长. 输入队列为 8个 send包长,即 640 bit TSRC发送数据包的目的地址是随机均匀分布.

2.2　流量均衡方式对比

仿真中使用了一个 8节点环路,在各次仿真中分别对流量均衡方式与 ticketmaster节点的连接方式进行改变,环路中只运行最高优先级. 图 4a~图 4d的仿真参数如表 2所示.

结果表明,系统流量较小时,不加流量均衡系统各节点流量差别不大,管理者迁移协议和扩展子优先级协议两种均衡方式效果相同,均能改善环路的节点流量分布.

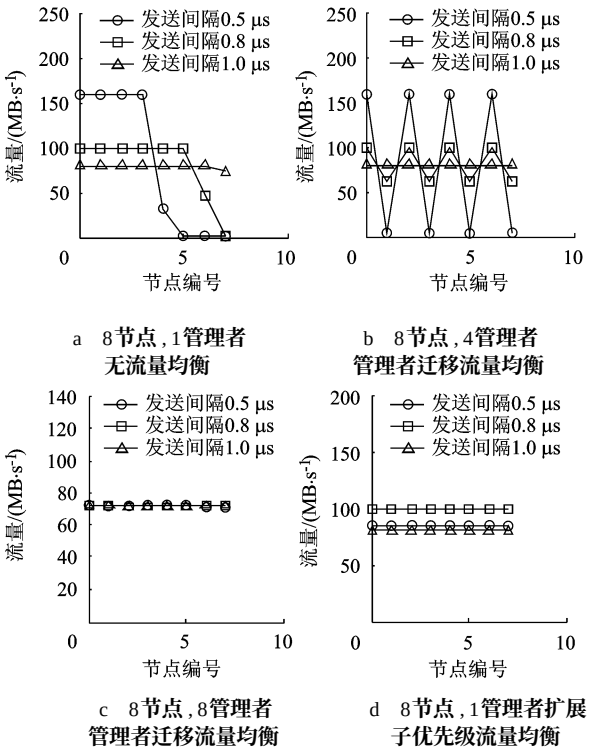


图 4　流量均衡效果对比

表 2　流量均衡仿真主要参数

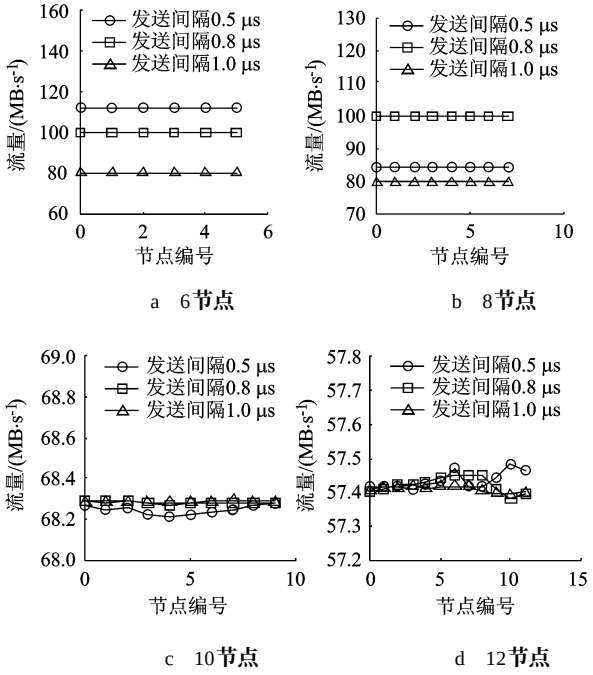
图号	管理者	流量均衡方式
图 4a	1	无
图 4b	4	管理者迁移流量均衡方式
图 4c	8	管理者迁移流量均衡方式
图 4d	1	扩展子优先级流量均衡方式

当系统流量较大,无任何流量均衡方法的环路,第 7号节点已基本丧失发送机会;对于 8管理者、使用管理者迁移流量均衡方法的环路,流量分布达到均匀分布,但微有起伏;而 4管理者,使用管理者迁移流量均衡方法的环路,节点流量分布为振荡方式,这是由于前一个节点插入的数据包影响随后节点的数据传输. 需要注意的是,使用扩展子优先级流量均衡方法的环路在这种情况下仍维持均匀的流量分布.

在系统流量进一步增大,无任何流量均衡方法的环路中,约半数的节点被阻塞,无法发送数据;在 8管理者、使用管理者迁移流量均衡方法的环路中,节点流量分布仍为均匀,只有微小起伏;在 4管理者,使用管理者迁移流量均衡方法的环路中,节点流量分布大幅振荡,其中管理者节点下游第 2个节点基本上被阻塞,无法得到环路使用

权;而在使用扩展子优先级流量均衡方法的环路中,虽然流量均衡方式和 8管理者、使用管理者迁移流量均衡方法的环路相同,但是其节点流量高.这主要因为扩展子优先级均衡方式不需要在环路中传输额外的数据包,而迁移管理者需要周期发送迁移命令,占用了一定的环路带宽.由此可以看出,使用扩展子优先级的流量均衡方式不仅可以在有较少网络管理者的情况下维持网络的正常运行,并且可以降低流量不稳定性,提高网络流量.

为了验证扩展子优先级流量均衡方式对节点数不同的 SCI/RT网络的有效性,分别在 6, 8, 10, 12节点的条件下进行了验证仿真,结果如图 5a～图 5d所示.结果显示扩展子优先级的均衡方法在不同规模的网络中均可以使用.



a 6节点 b 8节点

c 10节点 d 12节点

图 5 使用扩展子优先级流量均衡方式的环路仿真

3 结 论

从仿真结果可以得到:

在没有加入流量均衡的方式下,节点流量分布受总流量的变化的影响很大.当流量增加时,由于最下游的节点在传输数据包时不能抢占位于前面节点插入的同优先级数据包,使得大流量的环路中,后面的节点永远不能传送数据.

采用 TRAIN协议推荐的管理者迁移方式,其均衡效果受环路中管理者节点数量和位置的影响.如果分布是均匀的,则大流量情况下,节点流量分布是振荡形式的.只有当每个节点都具有管理者功能时,环路才能达到均匀的流量分布.

采用扩展子优先级方式,流量均衡能力好于管理者迁移方式,且有更高的节点流量.通过在不同规模环路中进行验证,证明扩展子优先级流量均衡方式可应用在不同规模的环路中.

综上所述,采用扩展子优先级方式进行节点流量均衡,能够得到优于迁移管理者均衡方式的均衡效果.并且,由于扩展子优先级方式均衡操作只限制在节点内部,不会影响数据包结构和 train的运行,不需多个管理者,简单可靠.另外,扩展子优先级方式不需要周期发送迁移命令,不占用环路带宽,节点流量大于管理者迁移方式.当然,扩展子优先级方式需要得到环路的节点号等参数,在灵活性上较管理者迁移方式差一些.

扩展子优先级流量均衡方式还可应用于其它包含多级优先级的网络中.适当选择优先级分配矩阵,也可人为控制节点流量的分布.

参考文献 (References)

[1] Chuck Pinney. Joint advanced strike technology program, avionics architecture definition, appendices[DB/OL]. 1994 - 08 - 09. http: ftp. jast mil

[2] 崔剑,郑铮,李铮. SCI协议分析 [J]. 光通信技术, 2005(9): 37 - 40
Cui Jian, Zheng Zheng, Li Zheng SCI standard analysis [J]. Optical Communication Technology, 2005(9): 37 - 40 (in Chinese)

[3] 余强力,陆达,王明芬. SCI数据通信技术的研究和实现 [J]. 计算机技术与发展, 2007(11): 190 - 192
Yu Qiangli, Lu Da, Wang Mingfen. Research and implementation on SCI data transmission[J]. Computer Technology and Development, 2007(11): 190 - 192 (in Chinese)

[4] IEEE1596-1992 IEEE standard for scalable coherent interface [S]

[5] IEEE P1596. 6 Draft TRAIN protocol For SCI/RT draft version 1. 3[S]

[6] 刘兴春,刘阳,崔剑,等. 基于 Opnet的航空电子 SCI链路节点建模 [J]. 光通信技术, 2007(5): 30 - 34
Liu Xingchun, Liu Yang, Cui Jian, et al An OPNET-based simulation model of SCI-node linc [J]. Optical Communication Technology, 2007(5): 30 - 34 (in Chinse)

[7] Hagen M, Varki E. Fibre channel arbitrated loop-performance analysis of loop tenancy overhead [C]// Proceedings of the 2008 International Conference on Networking, Architecture, and Storage. Washington: IEEE Computer Society, 2008: 225 - 232

(编 辑:娄 嘉)