Chapter 11

Schedulability Analysis for Real Time On-Chip Communication with Wormhole Switching

Zheng Shi  
University of York, UK

Alan Burns  
University of York, UK

Leandro Soares Indrusiak  
University of York, UK

ABSTRACT

In this paper, the authors discuss a real-time on-chip communication service with a priority-based wormhole switching policy. The authors present a novel off-line schedulability analysis approach, worst case network latency analysis. By evaluating diverse inter-relationships and service attributes among the traffic flows, this approach can predict the packet network latency for all practical situations. The simulation results provide evidence that communication latency calculated using the real time analysis approach is safe, closely matching the figures obtained from simulation.

INTRODUCTION

On-chip multiprocessing has reached mainstream in server, desktop and laptop computers, and is being increasingly applied to embedded systems. As 45nm and 32nm process technologies mature, we approach the mark of 2 billion transistors on a single silicon die, allowing for the on-chip integration of dozens of processing cores. This is an increase by one order of magnitude, if compared with processors that are commercially available. As the number of cores per chip increases, on-chip communication plays an increasingly important role on system performance, power dissipation, chip area, design effort and total cost. Since early solutions based on point-to-point interconnects or on-chip busses don’t scale and quickly become a communication bottleneck (Furber & Bainbridge, 2005; Henkel, Wolf, & Chakradhar, 2004), Networks-on-Chip (NoCs) have emerged
as a suitable architectural template for on-chip communication (Benini & De Micheli, 2002; Dally & Towles, 2001). Such template offers a generic communication platform which can be parameterised and reused for a large number of SoC designs.

As an integration platform, the NoC should have the capability to provide different level of services for various applications on the same network (Benini & De Micheli, 2001). Some applications have very stringent communication service requirements; the correctness relies on not only the communication result but also the completion time bound. A data packet received by a destination too late could be useless. These critical communications are called real-time communications. For a packet transmitted over the network, this time bound is denoted by the packet network latency. The worst case acceptable time metric is defined to be the deadline of the packet. A traffic-flow is a packet stream which traverses the same route from the source to the destination and requires the same grade of service along the path. For hard real-time traffic-flows, it is necessary that all the packets generated by the traffic-flow must be delivered before their deadlines even under worst case scenarios. A set of real-time traffic flows over the network are termed schedulable if all the packets belonging to these traffic flows meet their deadlines under any arrival order of the packet set.

As a popular switching control technique, wormhole switching (Ni & McKinley, 1993) has been widely applied for on-chip networks due to its high throughput and smaller buffering requirements (Kavaldjieva & Smit, 2003). However, the analysis of the real-time packet schedulability for wormhole switching networks is still an open problem. Predictable behaviour of the network services is essential to support real time requirements – particularly satisfying deadline bounds. But the situation for on-chip wormhole networks is partially non-deterministic due to the contentions in communication. In on-chip networks, several tasks running on different nodes exchange information periodically. During a transmission period, each transmitted packet shares the resources, such as buffers or physical links, with other packets. When several packets try to access the same resource at the same time, contention occurs and the network can only serve one packet and must suspend the others based on some arbitration policy. Once a packet becomes blocked, it can block other packets, which can in turn block other packets, and so on. The exact analysis of congestion in this situation is hard due to the possibility of a packet becoming blocked at several routers during its journey from source to destination. The contention problem leads to packet delays and even missed deadlines. Therefore, it is necessary to give an arbitration strategy and analysis approach to predict whether all the real-time packets can meet their timing requirements.

In this paper we explore real-time communication in wormhole switching for on-chip networks. Utilizing the priority based preemption arbitration (Balakrishnan & Ozguner, 1998; Hary & Ozguner, 1997; Kim, Kim, Hong, & Lee, 1998; Lu, Jantsch, & Sander, 2005), a novel real-time schedulability analysis approach is presented which successfully emulate wormhole switching as classic single processor scheduling model (Liu & Layland, 1973). By evaluating diverse inter-relationships and service attributes among the traffic-flows, this approach can predict the packet transmission latency for a given traffic-flow based on three quantifiable different interferences: direct interference from higher priority traffic-flows, indirect interference from higher priority traffic-flows and self-blocking produced by the previous packet from same traffic-flow. To further validate our approach, we performed extensive simulations aiming to demonstrate that the analysis is always safe, and to show situations when it is tight or not.

The rest of this paper is organized as follows: The second and third sections cover the major features of NoC architectures based on wormhole
Related Content

Handover Algorithm to Avoid Duplication Authentication Within or Between Proxy Mobile IPv6 Domains
[www.igi-global.com/article/handover-algorithm-to-avoid-duplication-authentication-within-or-between-proxy-mobile-ipv6-domains/228193?camid=4v1a](www.igi-global.com/article/handover-algorithm-to-avoid-duplication-authentication-within-or-between-proxy-mobile-ipv6-domains/228193?camid=4v1a)

A Multi-Dimensional Mean Failure Cost Model to Enhance Security of Cloud Computing Systems
[www.igi-global.com/article/a-multi-dimensional-mean-failure-cost-model-to-enhance-security-of-cloud-computing-systems/180291?camid=4v1a](www.igi-global.com/article/a-multi-dimensional-mean-failure-cost-model-to-enhance-security-of-cloud-computing-systems/180291?camid=4v1a)

External Pressures for Adoption of ICT Services Among SMEs
[www.igi-global.com/chapter/external-pressure-adoptions-ic/s-20460?camid=4v1a](www.igi-global.com/chapter/external-pressure-adoptions-ic/s-20460?camid=4v1a)

A Novel Dynamic Noise-Dependent Probabilistic Algorithm for Route Discovery in MANETs
[www.igi-global.com/article/novel-dynamic-noise-dependent-probabilistic-route/50481?camid=4v1a](www.igi-global.com/article/novel-dynamic-noise-dependent-probabilistic-route/50481?camid=4v1a)