Design of a Scalable Network Interface to Support Enhanced TCP and UDP Processing for High Speed Networks

A thesis submitted for the degree of
Doctor of Philosophy
by
Mohamed Elbeshti

Murdoch University
2014

Declaration

To the best of my knowledge, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for award of any other degree in any other university.

Signature:

Date: July 07, 2014

Dedicated to my Parents, Amira, Ahmed, Aseal, Yousef and Omar
for their support…

Abstract

Communication networks have advanced rapidly in providing additional services, with improvements made to their bandwidth and the integration of advanced technology. As the speed of networks exceeds 10 Gbps, the time frame for completing the processing of TCP and UDP packets has become extremely short. The design and implementation of high performance Network Interfaces (NIs) that can support offload protocol functions for current and next-generation networksis challenging. In this thesis two software approaches are presented to enhance protocol processing of TCP and UDP in the network interface. A novel software Large Receive Offload (LRO) approach for enhancing the receiving side has been proposed.The LRO works by aggregating the incoming TCP and UDP packets into larger packets inside the NI's buffer. The receiving side software has been improved to support out-of-order packets. The second proposed software solution is applied on the Large Send Offload (LSO). The proposed LSO function processing is implemented by segmenting TCP and UDP messages that are larger than the Maximum Transmission Unit to the Maximum Segment Size. New packet headers are generatedfor each new outgoing packet.

A scalable programmable NI based 32-bit RISC core is presented that can support 100 Gbps network speeds. Acceleration of the processing time frame required at the NI has been implemented to prevent hazards (such as Data Hazard and Control Hazard) during the execution of the LRO and the LSO functions. An R2000/3000 RISC has been used in order to test the LRO and LSO functions and to discover the instruction set that is most suitable. Following this the VHDL NI was implemented with three pipeline RISC cores, a simple DMA controller and Content Addressable Memory. An evaluation of the desired RISC clock rate that is required to process TCP and UDP streams at 100 Gbps was conducted. It was determined that aRISC core running at 752 MHz with a DMA clock of 3753 MHz was able to process packets 512 bytes or larger fast enough to support 100 Gbps network speeds.

Publications

Journal Articles:

- M. Elbeshti, M. Dixon, and T. KoziniecLarge Sending Offload: Design and Implementation for High-speed Communications Rate up to 100 Gbps, International Journal of New Computer Architectures and their Applications (IJNCAA). Vol 3, No. 2. 2013.

Conference Papers:

- M. Elbeshti, M. Dixon, and T. Koziniec, Design and Simulating Specialized Embedded Cores for UDP Network Interface Processing: *24th International Conference on Modeling and Simulation (MS 2013)* Canada, 2013.

- M. Elbeshti, M. Dixon, and T. Koziniec, Design consideration for efficient network interface supporting the Large Receive Offload with embedded RISC: *36th International Conference on Telecommunications and Signal processing (TSP)*, Italy, 2013.

- M. Elbeshti, M. Dixon, and T. Koziniec, Design and Simulating a Network Interface-based RISC Cores: *19th Asia-Pacific Conference on Communications*, Bali, 2013.

- M. Elbeshti, M. Dixon, and T. Koziniec, Design a Scalable Ethernet Network Interface Supporting the Large Receive Offload for 100 Gbps: *12thInternational Symposium on Communications and Information Technologies (ISCIT)*, Australia, 2012.

- M. Elbeshti, M. Dixon, and T. Koziniec, RISC core supporting the Large Sending Offload in 100 Gbps: *12thInternational Symposium on Communications and Information Technologies (ISCIT),* Australia, 2012.

- M. Elbeshti, M. Dixon, and T. Koziniec, An Evaluation of the TCP and UDP Processing Requirements Network Interface Design at 100 Gbps: *13thInternational Symposium on Advances of High Performance Computing and Networking (HPCN),* Canada, 2011.

- M. Elbeshti, M. Dixon, and T. Koziniec, TCP and UDP Processing Requirements for Network Interface Design at 100 Gbps:*3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Budapest, 2011.

- M. Elbeshti, M. Dixon, and T. Koziniec, Facing the Challenges of Designing a Network Interfaces for 100 Gbps: *2nd International Conference on Research Challenges in Computer Science,* China, 2010.

Posters and Demonstration:

- Mohamed Elbeshti, Designed and Implemented a Programmable Network Interface for High-speed Communications: *Hot Chips: A Symposium on High Performance Chips*, 2013.

Table of Contents

List of Figures

List of Tables

Acknowledgements

All praises are due to Almighty God “Allah”, Who provided me with the strength and willingness to undertake this work and the opportunity to contribute a drop in the sea of knowledge.

I am most grateful to my supervisors, Mike Dixon and Terry Koziniec. Their open doors and persistent encouragement was invaluable for the completion of this research. Also appreciated the time that Patrick, Lynette spent with me

Finally, these acknowledgements would not complete without appreciating the unwavering support of my family including my father, my wife and children, and the memory of my mother.

List of Abbreviation

ACK	ACK - Acknowledges received data
BOM	Beginning of Message
CAM	Content Addressable Memory
COM	Continuation of Messages
DMAC	Direct Memory Access Controller
EOM	End of Message
FIN	FIN - (Final) Cleanly terminates a connection
HI	Host Interface
HNIC	Host-NI Level of Communication Buffer
LI	Line Interface
LRO	Large Receive Offload
LSO	Large Send Offload
MAC	Media Access Control
RB	Receiving Buffer
RBI	Receiving Buffer Interface
REP	Receiving Embedded Processor
RISC	Reduce Instruct Set Computer
SB	Sending Buffer
SBI	Receiving Buffer interface
SEP	Sending Embedded Processor
SN	Sequence Number
SSM	Single Segment Message
SYN	SYN - (Synchronize) Initiates a connection
VHDL	Very High-Speed Description Language