Study on Hardware Acceleration on Algorithmic Trading

Hiroyuki Kobayashi (1410752)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2018

Keywords: SoC FPGA, Algorithmic Trading, Low Latency, TCP/IP, FIX Protocol.

1 Introduction

Developing domain specific hardware to react instantaneously to real-time streaming data has become active in the field of network processing, embedded computing, and controlling self-driving vehicles [1].

Various applications can take benefits from hardware acceleration. Among them, financial transaction activity is one of the fields which especially require low latency processing. Today, the major market participants are severely competing with each other in transmitting packets as quickly as possible, locating their trading servers in the same data centers in which the matching engines of the trading venues exist.

This study aims to reduce the latency between receiving price packets and sending order packets, by designing specific digital circuit in FPGA without expensive instruments. The trading algorithm to be targeted is the Stop Order, which is frequently used by many retail investors. The network protocols to be processed in hardware are FIX [2], TCP/IP and Ethernet.

The Stop Order is such a simple logic that accelerating the computing of algorithm is not intended. Instead, the primary advantage of applying

FPGA is bypassing time-consuming processes which inevitably come up both in the OS kernel and in the application software.

2 Related Work

Several research works have been engaged for low latency trading by exploiting the rapidity from hardware accelerators.

Gareth W. Morris, David B. Thomas and Wayne Luk (2009) [3] show that the latency can be shortened to less than 26μs by using FPGA as a stream processor to aggregate price information from various sources.

John W. Lockwood, Michaela Blott et al. (2012) [4] report that their NIC with FPGA can transact messages in less than 200ns. The first three chapters of this paper can also be used as a survey material on the current FPGA usage for low latency trading.

The research report by ARGON DESIGN, ARISTA and FTTC(2013) [5] mentions that their L7 switch starts sending messages before receiving all bits of incoming messages, which enables to reduce the latency to about 150ns.

3 Exchange Simulator and Client Software

We first develop server and client software which communicate with each other via FIX protocol. The server software is the exchange simulator to emulate the functionality of the matching engines at securities/derivatives exchanges such as TSE, NASDAQ or CME. In these exchanges, during their continuous trading sessions, orders on both sides of buy and sell are aggregated. When a pair of buy and sell orders satisfies the condition that the price of the sell order is equal or less than the price of the buy order, these orders are promptly matched and executed.

Thus the exchange simulator we develop accepts orders from clients, matches the orders, notifies the clients which have sent the orders that their orders have been accepted or executed, and distributes the updated price information to all connected clients.

The client program is able to send new orders to the exchange simulator, either manually or automatically with the pre-defined trading algorithm.

The trading algorithm in this study is the Stop Order, which sends orders when the last price becomes less than (or more than) the pre-determined trigger price.

4 Hardware Acceleration

We use PYNQ-Z1 board [6] which includes SoC FPGA (Xilinx Zynq 7020) to develop the system to expedite the network processing. Since the Zynq chip has Ethernet controllers (hard macro), as well as dual-core CPU (Cortex-A9), one of the controllers can be used to receive and transmit Ethernet frames.

The accelerator designed in the FPGA segment in Zynq contains BRAM and the FIX Processing unit. BRAM is the block memory which stores every data being exchanged between FPGA and software (a device driver and the FIX client application). The FIX processing unit is the logic to read Ethernet frames, to judge whether orders should be sent, and to build the Ethernet frame to be sent.

Next, we modify the network device driver to deliver the incoming Ethernet frames to BRAM instead of DRAM, as well as to send the outgoing frames in BRAM. Finally, the client application is also adjusted to be able to set the Stop Orders in BRAM.

5 Evaluation

We evaluate two types of latency. One is the latency arising in the digital circuit. The other is the total latency recognized at the server directly connected to the client via Ethernet.

The clock frequency given to the FIX processing system is 50MHz, namely, a cycle is 20ns. The FIX processing system writes outgoing frames in 32 cycles on average (at most 39 cycles) after the incoming frames are written in BRAM. Therefore, the latency coming up within FPGA fabric is 640ns on average and 780ns at a maximum, respectively.

As for the total latency, we measure it with packet capture tool Wireshark running in the server machine. The total latency is 148μs on average with this hardware accelerator, while the total latency of software execution is

684μs on average with the process running in a Linux PC, and 1573μs with the process running in the same PYNQ-Z1 board.

6 Conclusion

We designed a hardware accelerator to process real-time streaming data, with focusing on developing the low latency trading system. One of the most simple trading algorithms, the Stop Order was implemented on FPGA. Only the latency-sensitive modules, such as TCP/IP protocol stack and FIX engine, were allocated in the digital circuit on FPGA so as to make use of the existing software.

The maximum latency arising in digital circuit was under 1μs, while the total latency measured in the server side was 1/10 compared with the software execution in the same FPGA board, or 1/5 of the latency of the software running in a Linux PC.

References

[1] Data Flow Processor. https://www.denso.com/jp/ja/news/media-center/press-kits/tms2017/TMS_DFP_JPN.pdf. December 27 2017.

[2] http://www.fixtradingcommunity.org/ FIX Trading Community. December 27 2017.

[3] Gareth W. Morris, David B. Thomas and Wayne Luk. FPGA accelerated low-latency market data feed processing. In 17th IEEE Symposium on High Performance Interconnects 2009.

[4] John W. Lockwood et al. A Low-Latency Library in FPGA Hardware for High-Frequency Trading(HFT) In 2012 IEEE 20th Annual Symposium on High-Performance Interconnects

[5] ARGON DESIGN, ARISTA, finteligent. Research Report: The Arista 7124FX Switch as a High Performance Trade Execution Platform. 2013.

[6] Xilinx Python productivity for Zynq(Pynq) Documentation Release 1.01. 2017.