NTNU
Norwegian University of
Science and Technology

TDT4900 Computer Science, Master's Thesis

Optimization of Seed Selection for Information Diffusion with High Level Synthesis

Julian Lam

Fall 2016

Department of Computer and Information Science
Norwegian University of Science and Technology

Supervisor 1: Donn Alexander Morrison
Supervisor 2: Yaman Umuroglu

0.1 Assignment

Information diffusion is a field of network research where a message, starting at a set of seed nodes, is propagated through the edges in a graph according to a simple model. Simulations are used to measure the coverage and speed of the diffusion and are useful in modelling a variety of phenomena such as the spread of disease, memes on the Internet, viral marketing and emergency messages in disaster scenarios.

The effectiveness of a given spreading model is dependent on the initially infected nodes, or seeds. Seed selection for an optimal spread is an NP hard problem and is normally approximated by selecting high-degree nodes or using heuristic methods such as discount-degree or choosing nodes at different levels of the k-core.

High-level synthesis (HLS) is becoming an important tool in the optimization/acceleration of algorithms in hardware. Starting with an algorithm written in a high-level language such as C or C++, HLS aids with hardware design by providing a methodology and tools that guide the developer through the design process.

This project should employ HLS as a design methodology for hardware accelerated seed selection in large graphs. The student will study seed selection for a given diffusion model, write a high-level model, and use HLS to implement a hardware design that exploits parallelism in the seed selection algorithm in order to improve performance over a GPCPU implementation. –

Abstract

Information Diffusion are often used for different simulations in network research because it simulates how information propagetes thorough a network, from memes on the Internett, spreading of disease in populations, to viral marketing. Measuring spread and speed, we can find influential targets in the network, such targets are optimal targets to pass message during disaster scenario, vaccinate to prevent spreading of a disease, or even targets for viral marketing.

High Level Synthesis have in recent years matured greatly. With HLS, designing custom architectures is no longer a

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Information diffusion is a field of network research where a message, or data, is propagated through a *network* or a *graph*. The message originates from a chosen set of nodes, known as *seed nodes*. These seed nodes passes the message to it's neighbour through the edges and thus propagate the message over theentire network. There are different models used in Information diffusion, *Independent Cascade Model*, and *Linear Threshold Model*. Information Diffusion can be used to model different phenomena such as the spread of disease, viral marketing, or even spread of viral videos and *"memes"*[2]. The effectiveness of the simulation is measured in the spread and the speed of propagation. The effectiveness of the simulation is dependent on the chosen seed nodes. By finding the most optimal set of seed node, we can potentially stop an epidemic by vaccinating influential nodes, we can find important target for viral marketing by giving free sample, and use this information to quickly spread message during disaster scenarios.

There are multiple studies done regarding information diffusion, [3], [2], [4], [5]. There are few that focus on optimizing the seed selection, especially in hardware. The seed selection algorithm is an greedy solution, where every set of node is tested and the set with best coverage and time is chosen. This is a time consuming process and highly parallelizable. This makes it a good candidate for *Field-programmable gate arrays*(FPGAs).

High Level Syntesis(HLS) synthesis high level behaviour and constrains to lower level design.[6]. It allows users to implement an algorithm in high level language, C or C++, and generate an optimal design in *verilog* or *VHDL*. Verilog and VHDL are hardware descriptive language designed to describe digital systems [7]. In recent years, High Level Synthesis have gotten more attention and more support, the xilinx forums are anwsered quickly by the developers and highly populated with seasoned hardware designers and novices.

Unlike traditional hardware design, HLS allows programmer with limited knowledge to desgn an optimal custom *Intellectual property core*(IP-core). In

HLS, programmers can test out multiple different optimization schemes in short period of time. Thus allowing the programmer to quickly test out different optimization schemes.

For our implementation, we focused mainly on the ICM. The ICM is a special case of the common graph traversal algorithm *Breadth First Search*(bfs). For our implementation, we chose to implement the ICM as a custom *sparse matrix vector multiplication*(spmv). By performing ICM as spmv, we can utilize the parallelism options that spmv uncovers.

1.2 Assignment Interpretation

From the assignment text, these task were chosen as the main focus of this thesis:

Task 1 ***(mandatory)*** Implement Information Diffusion as Sparse matrix vector multiplication, with high level language C.

Task 2 ***(mandatory)*** Tailor the implementation of Information Diffusion for synthesise with Vivado HLS.

Task 3 ***(optional)*** Implement said design on a Zynq FPGA board.

Task 4 ***(optional)*** Extend the system to be able to handle graph in the size of toy graphs(containing 2^{26}) nodes)

1.3 Report Structure

We have here the basic outline for this report and a short overview of the remainder of this report:

Chapter 2: Background contains the information regarding network, Information diffusion, matrix vector multiplication and High level synthesis. Most of the background information regarding this report can be found in this chapter.

Chapter 3: Related Work shows what the related works and state of the art regarding information diffusion.

Chapter 4: Architecture

Chapter 5:.

Chapter 6: Future Work

Chapter 7: Conclusion

somehting

Works

Chapter 2

Result

as we can see, the algorithm was able to finish a

Chapter 3

Discussion

3.1 problem that was encountered

One problemt that I encoutered during this project was that the output signal from the synthesiser was not in the correct direction. The output signal was often set as input signal. The HLS would automaticlly set the values as output signal or input signal.The reurn value from a funciton would be set as a the output signal, while the variable that the function takes, would be set as the input signal. Another way wo specify that something is the output signal would be to explicitly set them as pointer arguments. This will in set the signal to be output signal.

3.2 AXI4

To use a standard protocol to transport data, Xilinx have created what is known as AXI4, there are different standards to implement, axi4 lite, stream, master, etc: each have a different usage and standards. The axi4 lite is more suitable for smaller and easier IP-core[need to explain what IP core is]

2.0000e+09 2000 500 222.2222 125 80 55.5556 40.8163 31.2500 24.6914

2000000000.00000, 2000, 500, 22.222222222222, 125, 80, 55.5555555556,40.8163265306122,31.2500000000000
24.6913580246914

3.3 implementaion

The vivado implementation on the Zedboard was problemtatic.

Work

Chapter 4

Conclusion

Bibliography

[1] Deepayan Chakrabarti, Yiping Zhan, and Christos Faloutsos. R-mat: A recursive model for graph mining. 4:442–446, 2004.

[2] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 491–501, New York, NY, USA, 2004. ACM.

[3] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 10(10-17):30, 2010.

[4] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *2010 IEEE International Conference on Data Mining*, pages 599–608, Dec 2010.

[5] Seth A. Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 33–41, New York, NY, USA, 2012. ACM.

[6] M. C. McFarland, A. C. Parker, and R. Camposano. The high-level synthesis of digital systems. *Proceedings of the IEEE*, 78(2):301–318, Feb 1990.

[7] Donald Thomas and Philip Moorby. *The Verilog® Hardware Description Language*. Springer Science & Business Media, 2008.