

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 propagates thorough a network, from memes on the Internet, 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

0.1	Assignment	i
1	Introduction	2
1.1	Motivation	2
1.2	Assignment Interpretation	2

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 pass the message to its neighbour through the edges and thus propagate the message over the 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"[?]. 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 influential nodes we can find target

Both Independent cascade model and the Linear Threshold model have different uses. W

High level synthesis takes algorithms implemented in high level programming language such as C and C++ and *synthesise* the design to low level designs[?].

1.2 Assignment Interpretation

From the assignment text, these tasks 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 synthesis 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(2^{26})