Project proposal bachelor thesis

Erik Leonards

April 5, 2023

Bachelor thesis Mathematics and Computer Science

Supervisor: prof. dr. (Rude)sindo Núñez Queija, prof. dr. Anuj Pathania

Informatics Institute
Korteweg-de Vries Institute for Mathematics
Faculty of Sciences
University of Amsterdam

Context

The thesis will be done in the field of parallel computing. The thesis concerns the parallel execution of multiple independent tasks in the CPU. Parallel computing greatly increases the throughput of a CPU. However, executing n independent tasks in parallel won't give a speedup of n, even when the overhead is ignored. This is caused due to the fact that all parallel tasks may create memory requests. Since each core has own local memory, the service rate of those requests is not effected by the other tasks. However, the memory requests to the shared memory of the computing system are slowing down the executiong due to the finite memory bandwidth of the shared memory.

This means that the execution time of a task executed in parallel is dependent on its own memory dependency and the memory dependencies of the other tasks. How exactly the execution time of a task in parallel execution compares to the execution time in isolation is not clear. This can only be determined by careful analyse of the dynamic in the shared memory.

This is done by modelling the memory requests as entities travelling from the processor core to the shared memory and back. The memory requests don't leave the system and the model corresponding to this situation is called a closed migration process. The analysis of the model will be done using the theory, definitions and theorems from queueing theory and Markov chains.

Relevant literature

The relevant literature from queueing theory concerns the analysis of closed queueing networks. This includes the equilibrium distribution of a closed migration process. There seems to be a lack in the literature of theory about queueing network models used for CPU's. This could mean that CPU's contain characteristics and properties very hard to incorporate in queueing network, but this is hopefully not the case.

The CPU simulation software named HotSniper will be used to compare the model to the correct execution time. This software is considered to be a highly accurate in simulating a real processor.

Research question

The thesis will investigate the relation between the isolated execution of tasks and their execution time when run in parallel. The final goal is to determine the parallel execution time from the isolated execution times. It could be the case that the model is not a good representation of the CPU, which will result in inaccurate results. It would then be interesting to determine which property or characteristic of the CPU are unable to be incorporated in the closed migration process.

Methods

The research question will be answered by defining a base model, which is equal to a simple closed migration process in our case. Then the model can be extended and adjusted to hopefully better capture the dynamic of the memory requests. The model will ultimately contain many abstract parameters. The value of these parameters can be determined by fitting the model to data from the HotSniper simulation program. Whether the model is correct with the fitted data can then be determined by running additional simulations using HotSniper. The model can then be extended or adjusted when the error is too big. This process can be done until a correct model is found.

The planning

The planning can be found below but is just a rough estimation.