

PROPOSAL REPORT

CENG491 Computer Engineering Design

October 16, 2008

Team

Team Name: YAKUT

Team Members:

Yunus Basagalar

Yigit Cagri Akkaya

Sefa Kilic

Isil Doga Yakut

Project

Project Title:

Cluster library for high performance computing : Converting an existing library to parallel library.

Cluster Computing

A computer cluster is a group of computers linked to each other with fast local area networks to work as a single computer. The main purpose is to reach high speed computation with less cost. To reach this goal, the process is splitted into smaller and independent parts and these parts are assigned to processors to run concurrently. To split the process, first bottlenecks -most CPU time consuming parts- of the sequential algorithm are detected, by using some profiler tools. Then either with existing techniques or newly developed techniques the process is parallelized.

Genetic Algorithms

A genetic algorithm is a search procedure to find the optimal solution. Inspired by genetics , it operates on an initial population to generate new populations at each step by producing new individuals by crossover and mutation. By evaluating each individual according to its fitness to the population ,at each step it replaces worst ranked part with offspring.

Parallelization of Genetic Algorithms

Generating an offspring from the existing population is done by mutation and crossover. Mutation is changing an arbitrary bit of an individual's genetic sequence. And crossover is interchange of some sections of genetic sequence between two individuals. Crossover and mutation operations are appropriate for parallelization, since mutation requires only one individual and crossover requires a pair of individuals and their execution does not effect other ones. On the phase of selection of new population, a fitness function is applied. This function is applied to each individual independent from others. By parallelization of fitness function, selection phase can be sped up.

However, processing jobs in parallel has the disadvantage of communication overhead. To determine if there is a loss in efficiency, analysis of the trade-off between communication overhead and parallelization time is performed. According to the result of this analysis, it is decided whether a change is required such as reducing the number of parallel processes which deals with that part of the job, modifying the algorithm to make it prone to parallelization, or even decide not to parallelize the related part.

Tools to Develop a Parallelized Genetic Algorithm Library

Some of existing genetic algorithms libraries are GALib, GPC++ , Beagle , EO Evolutionary Computation Framework , EGJ - Evolutionary Computation/Genetic Programming Research System. We are planning to use GALib¹ - Genetic Algorithms Library which is an open-source library. We will be coding in C++ since GALib is in C++.

In our project we plan to use NAR, the High Performance Computing Facility (HPCF) ² of computer engineering department to develop a generic library. HPCF is a computer cluster, consisting of blade servers. To develop a generic library on only one type of a cluster we will use OpenMPI, since it is installed on HPCF. OpenMPI which is an implementation of MPI (Message Passing Interface).

¹<http://lancet.mit.edu/ga/>

²<http://hpc.ceng.metu.edu.tr>