Predicting Performance of the GLU System

Juan Ley, KLA Inc. and Ed Ashcroft, ASU

Abstract

The GLU system allows the revitalization of existing code through the use of a distributed network of workstations. The aim is get speedup, and it has long been clear that the speedup depends on the number of workstations used and the size of the tasks the workstations perform, and the number of such tasks for the whole job. Choosing the task-size, and hence the number of tasks, for a given job, when the number of workstations is known, has been something of mysterious art. Even harder is the prediction of how many workstations GLU could profitably employ for given problem. What is needed is a mathematical model of the GLU system that reliably corresponds to the system’s actual behaviour.

Such a model has been developed for situations where certain simplifying assumptions hold, such as uniformity of the workstations and time-division multiplexing (so that access to the network is actually sequential). This model describes the relationships between such things as task granularity, network latency, problem parallelism, and communication overhead.

The model has been validated by the comparison of its theoretical predictions with actual measurements of GLU’s performance, at SRI, for such applications as Matrix Multiplication and Computer Tomography. One interesting prediction that has been confirmed is the way that adding workstations sometimes doesn’t increase performance. In fact, the model sometimes predicts performance varying in a step-function manner, and this has actually been observed.