TJHSST Senior Research Project
Decentralized Distributed Procsesing
2006-2007

Michael Tao

November 2, 2006

1   I. Problem Statement

With the enormous amount of data being collected every day, a single computer's CPU's computational ability to analyze the data and to utilize meaning behind the data is less than satisfactory. In order to mine through of the data within certain time constraints, a collection of computers is needed. The purpose of this project is to produce a medium for distributing the load of enormous tasks to networked peers with varying computing power in an efficient manner.

2   II. Purpose

This will distribute the work load from one computer to other computers within a network of peer computers by sending portions of the data and the proper analytical tools to all of the specified peers while also computing various peer's tasks. Peers can be running on multiple computer platforms such as Windows and Linux.

3   III. Scope of Study

Research on the TCP/IP protocol, packets, and distributing CPU power.

4 IV. Background and Review of current literature/research in this area

Though distributed servers and clusters have existed for a while, there is a lack of sharing, most distributing acts rely on a single taskgiver, and the peers being enslaved to the server, with little / no reciprocation. As the quantity of data and complexity of analysis from individual groups becomes greater, the efficiency current distributed processing units will certainly become less than satisfactory.

5 V. Procedure and Methodology

This project will be completed by gradually adding separate components until the whole is completed. Initially an application which will allow two test nodes to communicate will be used to test TCP/IP, then several at the same time. Then it will pass textual data to other nodes, for a simple analysis. The next step would be to send files, and add a modular analysis interface. With a test analysis and data, a load heuristic and a difficulty heuristic will be developed, and finally a balancing system which uses those heuristics to decide where to distribute work.

6 VI. Expected Results and Value to Others

The result should be an efficient multi-platform peer-to-peer load distributer. The equality of load distribution will be a measure of how well the application is working. This application would be a means for organizations who need to analyze large amounts of data to do so more efficiently.