

## **Data discovery algorithm for scientific data grid environment**

### **ABSTRACT**

In modern scientific computing communities, scientists are involved in managing massive amounts of very large data collections in a geographically distributed environment. Research in the area of grid computing has given us various ideas and solutions to address these requirements. Data grid mostly deals with large computational problems and provides geographically distributed resources for large-scale data-intensive applications that generate large data sets. Peer-to-peer (P2P) networks have also become a major research topic over the last few years. In a distributed P2P system, a discovery algorithm is required to locate specific information, applications, or users within the system. In this research work, we present our scientific data grid as a large P2P-based distributed system model. By using this model, we study various discovery algorithms for locating data sets in a data grid system. The algorithms we studied are based on the P2P architecture. We investigate these algorithms using our Grid Simulator developed using PARSEC. In this paper, we illustrate our scientific data grid model and our Grid Simulator. We then analyze the performance of the discovery algorithms relative to their average number of hop, success rates and bandwidth consumption. © 2005 Elsevier Inc. All rights reserved.

**Keyword:** Data grid; Discovery algorithm; Grid computing