Decentralized replication strategies for P2P based scientific data grid

ABSTRACT

Scientific Data Grid provides geographically distributed resources for large-scale data-intensive applications that generate large scientific data sets and it mostly deals with large computational problems. Research in the area of grid has given various ideas and solutions to address these requirements. However, since the number of participants (scientists and institutes) that involve in this kind of environment is increasing tremendously, scalability, availability and reliability have been the core problem for such system. Peer-to-peer (P2P) is one of the architecture that promising scale and dynamism environment. In this paper, we present a P2P model for Scientific Data Grid that utilizes the P2P services to address those problems. For the purpose of this study, we have developed and used our own data grid simulation written using PARSEC. In this paper, we illustrate our P2P Scientific Data Grid model, our data grid simulation and the design of proposed data replication strategies. We then analyze the performance of data discovery service with and without the existence of replication strategies relative to their success rates, response time, average number of hop and bandwidth consumption. The results from simulation study that show how the proposed replication strategies promote high data availability in the proposed Scientific Data Grid model and how these strategies improve the discovery process are presented.