

## **A cluster-based data replication technique for preserving data consistency in data grid**

### **Abstract**

Data grid has been developed to provide a scalable infrastructure for managing and storing data files and support data intensive applications. However, managing the huge and widely distributed data has raised some issues such as data consistency, data availability and communication costs. To address the issues, one of the commonly used techniques is data replication which can provide high availability and increase the performance of the system. Many replica control protocols have been proposed in distributed database and grid which achieved both high performance and availability. However, most of the previously proposed protocols perform well in small size systems and have a small number of replicas. As the network size grows, a larger number of replicas are required to be accessed in order to maintain data consistency, which is not suitable for a large scale system such as data grid. Thus, in this paper, we propose a new replica control protocol named Cluster-Based Replication (CBR) protocol for managing the data in data grid. We analyze the communication cost of the operations and compare CBR protocol with previously proposed tree-based replica control protocols namely Logarithmic protocol and Dynamic Hybrid protocol. A simulation model was developed using Java to evaluate CBR protocol. Our results show that for the read and write operations, CBR provides lower communication cost as well as maintains data consistency.

**Keywords:** Data Grid; Data replication; Data availability; Communication cost