Striped Replication from Multiple Sites in the Grid Environment

Marek Ciglan, Ondrej Habala, Ladislav Hluchy

Institute of Informatics, Slovak Academy of Science
Dubravska cesta 9
845 07 Bratislava, Slovakia
{marek.ciglan,ondrej.habala,hluchy.ui}@savba.sk

Abstract. Grid technology, as a highly distributed computing environment, requires an optimized access to the data resources to increase data availability. In this paper, we propose a replication technique which is based on parallel transfers from multiple sites containing replicas of the desired file. From each site, we transport in parallel only a portion of the given data source, obtaining the whole file at the end of the process. We describe the work related to the data replication; then we discuss two algorithms for striped replication optimization that aims at the minimization of the time necessary for data transfer. Finally, we present the results of the striped replication mechanism achieved by the prototype implementation of the striped replication algorithm. We compare them with the results of the standard replication tools and show interesting performance improvement.

1 Introduction

Grid computing is an important new concept for distributed processing, based on the idea of globally shared computer resources between organizations, such as disk space, information and computational power [**?**]. Grid computing helps users, who need to run computational and data intensive tasks, which could be too demanding and time consuming for a single supercomputer or computational cluster.

Such a task is distributed within the grid to several grid nodes, saving execution time. This brings immense computational power for relatively small cost. In such highly distributed environment, an efficient data management is needed to keep track of the data sources in the grid and to optimize the network traffic. Many grid applications request large data collections, which, possibly, have to

---