

Overbooking in Planning Based Scheduling Systems *

Georg Birkenheuer, Kerstin Voss
Paderborn Center for Parallel Computing
University of Paderborn, Germany
{birke,kerstin}@upb.de

Matthias Hovestadt, Odej Kao
Technical University of Berlin, Germany
{battre,maho,okao}@cs.tu-berlin.de

September 24, 2008

Abstract

Nowadays cluster Grids encompass many cluster systems with possible thousands of nodes and processors, offering compute power that was inconceivable only a few years ago. For attracting commercial users to use these environments, the resource management systems (RMS) have to be able to negotiate on Service Level Agreements (SLAs), which are defining all service quality requirements of such a job, e.g. deadlines for job completion. Planning-based scheduling seems to be well suited to guarantee the SLA adherence of these jobs, since it builds up a schedule for the entire future resource usage. However, it demands the user to give runtime estimates for his job. Since many users are not able to give exact runtime estimates, it is common practice to overestimate, thus reducing the number of jobs that the system is able to accept. In this paper we describe the potential of overbooking mechanisms for coping with this effect.

*This work has been partially supported by the EU within the 6th Framework Programme under contract IST-031772 "Advanced Risk Assessment and Management for Trustable Grids" (AssessGrid).