

ABSTRACT of the thesis presented by **Adán Hiraes Carbajal** as a partial requirement to obtain the DOCTOR OF SCIENCE degree in COMPUTER SCIENCE. Ensenada, Baja California, México, June, 2012.

WORKFLOW SCHEDULING OF MULTIPLE WORKFLOWS USING USER GIVEN RUNTIME ESTIMATES

Summary approved by:

Dr. Andrey Chernykh
Director de Tesis

In this thesis, we present an experimental study of deterministic non-preemptive multiple workflow scheduling strategies on a grid. We distinguish twenty five strategies depending on the type and amount of information they require. We analyze scheduling strategies that consist of two and four stages: labeling, adaptive allocation, prioritization, and parallel machine scheduling. We apply these strategies in the context of executing the Cybershake, Epigenomics, Genome, Inspiral, LIGO, Montage, and SIPHT workflows applications. In order to provide performance comparison, a joint analysis considering three metrics is performed. A case study is given and corresponding results indicate that well known DAG scheduling algorithms designed for single DAG and single machine settings are not well suited for grid scheduling scenarios, where user run time estimates are available. We show that the proposed new strategies outperform other strategies in terms of approximation factor, mean critical path waiting time, and mean critical path slowdown. The robustness of these strategies is also discussed.

Keywords: Grid computing, scheduling, workflow, resource administration