Parallel Computer Workload Modeling with Markov Chains

Baiyi Song, Carsten Ernemann*, and Ramin Yahyapour

Computer Engineering Institute, University Dortmund, 44221 Dortmund, Germany
{song.baiyi, carsten.ernemann, ramin.yahyapour}@udo.edu

Abstract. In order to evaluate different scheduling strategies for parallel computers, simulations are often executed. As the scheduling quality highly depends on the workload that is served on the parallel machine, a representative workload model is required. Common approaches such as using a probability distribution model can capture the static feature of real workloads, but they do not consider the temporal relation in the traces. In this paper, a workload model is presented which uses Markov chains for modeling job parameters. In order to consider the interdependence of individual parameters without requiring large scale Markov chains, a novel method for transforming the states in different Markov chains is presented. The results show that the model yields closer results to the real workloads than other common approaches.

1 Introduction

The use of parallel computers and workstation clusters has become a common approach for solving many problems. The efficient allocation of processing nodes to jobs is the task of the scheduling system. Here, the quality of the scheduling system has a high impact on the overall performance of the parallel computer. To this end, many researchers have developed various job scheduling subsystems for such parallel computers [28, 15, 17]. As already pointed out in [16, 7], the performance of a scheduling algorithm highly depends on the workload it is applied to. There is no single scheduling algorithm that is best for all scenarios. To this end, the evaluation of scheduling algorithms for different workloads is an important step in designing a scheduling system. Therefore, much effort has been put in the characterization and modeling of the workload of parallel computers [4, 1, 6, 24].

A typical approach for the performance evaluation of a scheduling system is the application of an existing workload trace which has been recorded on an existing machine [29, 25, 13]. However, while this represents a realistic user behavior on a real machine, there are several drawbacks. For instance, such a workload trace cannot directly be applied to configurations different from the

* Carsten Ernemann is a member of the Collaborative Research Center 531, "Computational Intelligence", at the University of Dortmund with financial support of the Deutsche Forschungsgemeinschaft (DFG).

D. Feitelson, L. Rudolph, and U. Schwiegelshohn (Eds.): JSSPP 2004, LNCS 3277, pp. 47–62, 2005.

original machine. In addition, the size of the workload, that is the number of jobs in the trace, cannot be scaled easily.

Therefore, often a statistical workload model is adopted as an alternative. The most common approach is the use of a probability distribution function model (PDF) [16]. However the PDF model often omits the dynamic characteristics of workloads. That is, the sequential correlation of different jobs is not taken into account. In this paper, we propose an extended job model based on Markov chains which uses information from the previous job to consider the sequential dependencies for the next job submission. After a discussion of the necessary background in Section 2, we discuss in Section 3 the relevant model parameters. In Section 4, the model is constructed. The quality of the model is evaluated by comparing its outcome with real workload data in Section 5. The paper ends with a short conclusion.

2 Background

Many parallel computers or supercomputers use a space-sharing strategy for efficient execution of parallel computational jobs. This means that a job runs exclusively on the allocated processor set. Moreover, jobs are executed until completion without any preemption. The scheduling problem is an online scenario in which the jobs are not known in advance and are continuously submitted to the scheduling systems by the users.

A workload model is an abstract description of the parameters of the jobs in the workload. A job consists of several parameters, for instance the number of required processing nodes, the job runtime, or memory requirements. In this paper, we concentrate on the modeling of the required number of nodes and the corresponding runtime. However, our approach is general and can easily be extended to consider other parameters as well. Note, that we do not model the submission time or inter-arrival time of jobs. For this task several other adequate models are available [3, 6].

As mentioned before, often a probability distribution function model is chosen for modeling workload parameters. Thereby, the parameters are typically considered independently and, consequently, individual distributions are created for each parameter. For example, Jann et al. used a hyper-Erlang distribution to match the first 3 moments of an observed distribution [20]. Alternatively, Uri Lublin and Dror Feitelson used a three-stage hyper-gamma distribution to fit the original data [24].

Besides the isolated modeling of each attribute, the correlations between different attributes are also very important. Lo et al. [23] demonstrated how the different degrees of correlation between job size and job runtime might lead to discrepant conclusions about the evaluation of scheduling performance. To consider such correlations, Jann et al. [20] divided the job sizes into subranges and then created a separate model for the inter-arrival time and the service time in each range, which may have a risk of over-fitting and too many unknown parameters. Furthermore, Lublin and Feitelson in [24] considered the runtime