Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

SPECIFICATION, SIMULATION AND ANALYSIS OF ALTERNATIVES FOR ON-LINE SCHEDULING OF INDEPENDENT JOBS IN DIFFERENT SERVERS

Jaume Figueras Jové
Pau Fonseca Casas

inLab FIB
Universitat Politècnica de Catalunya (UPC)
Campus Nord, Edifici B5, C/ Jordi Girona 1-3
Barcelona, E-08034, SPAIN

ABSTRACT

Service companies have the challenge to analyze a large number of documents in order to extract relevant information for decision making. Such analysis can be made automatically reducing drastically the time amount and human effort needed. However, the computer system must ensure that the analysis of each document will be completed within a specified period of time which depends on the type of the document. A real case study is presented in this paper where the objective is to propose a new scheduling model for a computer system with 6 servers with a total of 384 logical cores. The arrival of documents is aperiodic and the processing time stochastic though processing time estimation can be done based on the number of pages and the type of the document. A simulation model has been developed to analyze the quality of each algorithm. A delay maximum time (DMT) algorithm is also proposed.

1 INTRODUCTION

This abstract presents an heterogeneous computing application made of a set of servers that process legal documents with the goal of extracting relevant information for decision making. Each job (document) can be characterized by its type and the number of pages.

There is a wide variety of job scheduling algorithms due to the wide spectrum of needs and requirements of different real-time systems. A scheduling algorithm can be described as a set of rules that tell the scheduler how to manage these systems, through queuing jobs and allocating processing time. Therefore, the choice of a specific algorithm to carry out the scheduling can have a great influence on the behavior of the system (Lindh 2010). Among the possible criteria to evaluate a planning algorithm we find (Krallmann 1999). In addition, the proposed algorithms to optimize job processing planning are divided into 2 groups, depending on whether the jobs have a strict deadline (hard deadline) or a more permissive deadline (soft deadline). Thus, a hard deadline job is defined as one in which failure to comply with said restriction could result in a dramatic destabilization of the system (Kopetz 2000; Juvva 1998). The rest of the conditions would be considered as soft deadline, that is, systems where the failure of a given time limit reduces the efficiency of the system but does not entail a significant economic loss.

Heterogeneous computing refers to systems that use more than one kind of processor or core. These systems gain performance or energy efficiency not just by adding the same type of processors, but by adding dissimilar coprocessors, usually incorporating specialized processing capabilities to handle particular tasks (Gregg 2011). The dynamic approach to scheduling is based on the ability to predict the processing time when the job is assigned to a particular server and the expected server hourly utilization.

2 LENGTH AND CONTENTS OF EXTENDED ABSTRACTS

During the day different jobs arrive. They will be distributed by the different servers so that they meet or try to meet the established service levels in the most optimal way. It is important to keep in mind that timeout time starts counting from the moment the work reaches the main queue.

Once a job is assigned to one server, it can no longer be moved to another. The goal of this study has been to analyze and propose a flexible scheduling algorithms. The scheduling approach used previously by the service company was based in using an hourly configuration file that is repeated for all days of the week. This configuration file specified, for each hour of the day, the server and number of cores assigned to each type of job.

3 SIMULATION AND MODEL VALIDATION

The quality of the different scheduling algorithms will be evaluated by means of a discrete event simulation model. A stage prior to the analysis of the different algorithms consists in the validation of the simulation model. For this, the model has been fed with a real sequence of works to verify that the degree of utilization of each server obtained with the simulator is similar to the utilization observed in reality.

The dynamic approach to on-line scheduling is based on the ability to predict the processing time of a job for a specific server and hourly utilization load for each job type.

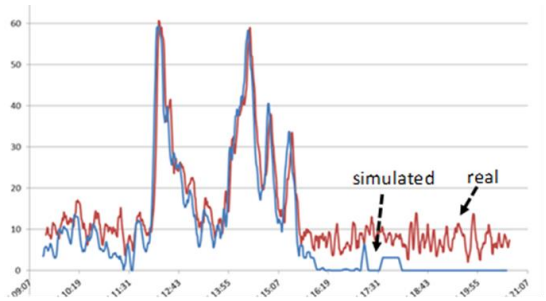

Figure 1: Comparison of Server #1 Utilization.

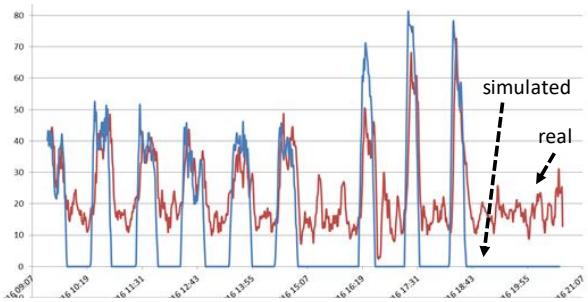

Figure 2: Comparison of Server #3 Utilization.

4 RESULTS AND CONCLUSION

The results have been obtained in a simulation of a 10 days period in which the total number of job orders received has been 773.000. The arrival time, type of the job and processing time has been taken from a data base of jobs processed in the 10 days period. In different scenarios changing the numbers of cores the FCFS algorithm has the worst results whilst EDF and LST algorithms have similar results, so it is preferable to use the EDF since it has a lower computational cost.

REFERENCES

G.M., Amdahl. 1967. "Validity of the Single-Processor Approach to Achieving Large-Scale Computing Capabilities." In *Proc. Am. Federation of Information Processing Societies Conf.*, AFIPS Press 483–485 .

Gregg C., M. Boyer, K. Hazelwood, and K. Skadron. 2011. "Dynamic Heterogeneous Scheduling Decisions Using Historical Runtime Data." In Work. Appl. Multi-and Many-Core Process. https://api.semanticscholar.org/CorpusID:15305172

Juvva, K. 1998. "Real-time systems." https://users.ece.cmu.edu/~koopman/des_s99/real_time/

Kopetz, H. 2000. "Software Engineering for Real-time: A Roadmap." In *Proc. of the Conf. on The Future of Soft. Eng.*, 201–211.

Krallmann J., U. Schwiegelshohn, and R. Yahyapour. 1999. "On the Design and Evaluation of Job Scheduling Algorithms." In: *Job Scheduling Strategies for Parallel Processing*. Lecture Notes in Computer Science, vol 1659, 17–42.

Lindh F., T. Otnes, and J. Wennerström. 2010. "Scheduling Algorithms for Real-Time Systems." Department of Computer Engineering, Malardalens University, Sweden.