Abstract

Improving response time is considered a fundamental objective in interactive environments. CPU scheduling aimed mainly to optimize the response time by minimizing its average in order to attain faster responses to users’ requests. However, for interactive systems, reasonable and predictable services are more preferred than faster responses but highly variable. Delivering service in a timely manner at less variable response time is an issue that has been addressed in this paper. A goal programming (GP) model is proposed to perform CPU scheduling at minimum variance and low response time. The GP method determines the optimal process in the ready queue that best minimizes the variance to be executed first. A simulation system that can generate varied scheduling situations was developed and several tests were conducted. The performance of the proposed GP scheduling method is measured and compared to the other related scheduling methods. The evaluation results show that the GP scheduling method can provide predictable and reasonable service and it performs scheduling at minimum variance and lower response time. The GP method outperforms the other related methods with varying degrees.
References


18. G. You, X. Wang, and Y. Zhao, A Dynamic Requests Scheduling Model Based on Prediction in Multi-core Web Server. In: Internet of Vehicles – Technologies and Services,


43. P. M. Broadwell, Response time as a performability metric for online services. Computer Science Division, University of California. 2004.


Index Terms

Computer Science  Parallel Computing

Keywords

CPU scheduling, Goal programming, Interactive systems, Response time, Variance in response time