Applications of a Queueing Network Model for a Computer System

LESTER LIPSKY

Department of Computer Science, University of Nebraska, Lincoln, Nebraska 68508

J. D. CHURCH

Department of Mathematics, University of Kansas, Lawrence, Kansas 66045

Constructing a detailed "analytic" model for a real-world computer system is a complicated task, and interpreting results of applications may be difficult and risky. This tutorial article presents a case study of a modeling project for a specific IBM 360/65J computer system and describes some practical applications of the resulting central server model. A sketch of the underlying theory of such models is included as an aid to interpreting results.

The model was used to obtain the relationship of throughput, as measured by CPU activity, to characteristics of the job-stream and to properties of various hardware and software components of the system. This relationship was used to estimate changes in performance resulting from various alternative modifications of the system. The utility of these estimates was confirmed by system measurements taken after some of the recommended modifications were implemented.

Keywords and Phrases: Computer performance prediction, analytic models, queueing, system modifications

CR Categories: 8.1, 6.20

BIBLIOTHEQUE DU CERIST

INTRODUCTION

This paper demonstrates the utility of queueing network models for computer systems by describing a modeling project for a specific system. This "case study" is presented in detail, with a discussion of the theoretical background, some of the practical problems involved, validation of the model, and interpretation of the results of studies made using the model.

Throughout the paper, emphasis is placed on system productivity (throughput) as determined by the fraction of time the central processing unit (CPU) is actively processing the normal job-stream. Thus, throughput is measured relative to a given CPU, and is not an absolute measure. Different CPUs can be compared by using this model, but this requires a measure of productivity that includes CPU speed as a factor. Another measure of system performance, turn-around time, is ignored. Since throughput and turn-around time tend to be inversely related, most strategies for increasing throughput will decrease average turn-around time as well unless an attempt is made to use knowledge of individual jobs (i.e., shortest jobs first).

After some general theoretical results for queueing networks are described and a hypothetical example of their application given, a detailed model is developed for the IBM 360/65J system at the University of Nebraska-Lincoln Computing Facility

Copyright © 1977, Association for Computing Machinery, Inc. General permission to republish, but not for profit, all or part of this material is granted provided that ACM's copyright notice is given and that reference is made to the publication, to its date of issue, and to the fact that reprinting privileges were granted by permission of the Association for Computing Machinery.