Abstract

Aspects of Workload Characterization for Distributed System Models
Andria L. Hunter
Master of Science
Graduate Department of Computer Science
University of Toronto
1996

Distributed systems have been steadily increasing in number and scale over the past two decades. This increasing trend intensifies the need for workload characterization and system modelling in this area. The increasing size of distributed systems creates the need for compact models and techniques to efficiently analyze massive amounts of data.

In this thesis, data are collected from an academic distributed computing system that was operating at peak user load. We provide a detailed workload characterization of the system, and then proceed to design a model that could be used to drive a simulation of this distributed system. The particular model that we have designed is static, as it is based on a period that had little variance in workload activity. The model is distribution-based; it uses stochastic sampling from representative distributions. We use a nonparametric density clustering method to classify the data and to identify the major components of the model.

The major contribution of this thesis is its furthering of the understanding of the tools and methodology that can be used in workload characterization and model design for distributed systems. We demonstrate how cluster analysis and other techniques can be used to provide a meaningful analysis of a massive set of "live" workload data. In addition to the specific results that we present, we feel that the collective wisdom provided by the discussion of the methodology throughout this thesis provides valuable insight for others in the same area of research.