Effect of Model Uncertainty on Some Optimal Routing Problems

B. Mohanty and C. G. Cassandras

Communicated by Y. C. Ho

Abstract. We study the effect of model uncertainties on optimal routing in a system of parallel queues. The uncertainty arises in modeling the service time distribution for the customers (jobs, packets) to be served. For a Poisson arrival process and Bernoulli routing, the optimal mean system delay generally depends on the variance of this distribution. However, as the input traffic load approaches the system capacity, the optimal routing assignment and corresponding mean system delay are shown to converge to a variance-invariant point. The implications of these results are examined in the context of gradient-based routing algorithms. An example of a model-independent algorithm using on-line gradient estimation is also included and its performance compared with that of model-based algorithms.

Key Words. Routing, optimization, queueing systems, robustness, distributed algorithms.

1. Introduction

Consider the basic resource sharing problem frequently encountered in computer systems and communication networks (see Fig. 1), where a stream of customers (e.g., jobs, packets) is distributed among several parallel
service facilities (e.g., processors, links) modeled as queueing systems. A simple objective is to determine the fraction of customers routed to each facility so as to minimize some specified cost function. When the arrival and service distributions involved in the model are given, an analytical solution to this problem may be obtained using standard queueing theory. In this paper, we address the following fundamental robustness question related to this problem: what is the effect of model uncertainty on the optimal routing solution? In particular, if the arrival and/or service distributions are inaccurately modeled, one generally expects the cost attained by the optimal solution under this model to differ from the actual minimum cost.

In addressing this issue, we are motivated by two observations. First, to optimize routing in many systems, one can use sophisticated algorithms (e.g., Refs. 1-3) which rely on specific models for the arrival and service distributions. It is therefore of interest to investigate whether the computational effort involved in executing such algorithms is worthwhile if the resulting solutions are sensitive to the assumed models, or whether one can occasionally rely on some basic robustness properties of the optimization process in order to get by with possibly inaccurate models. Second, one may question whether any optimal routing can be done if no model at all is available. In this case, using gradient-based routing algorithms in conjunction with several recently developed on-line gradient estimators (e.g., Refs. 4-6) presents the potential for robust optimal routing in the presence of little or no a priori information about the operating environment.

A large number of complex and sophisticated routing algorithms can be found in the context of communication networks. Most of the algorithms employed in practice, such as ARPANET (Ref. 7), DNA (Ref. 8), TYMNET (Ref. 9), TRANSPAC (Ref. 10), and CODEX (Refs. 11 and 12) employ some form of shortest path routing, where each link is assigned a length

---

**Fig. 1.** M-parallel queue model.