CCIT Report # 537 June 2005

Queueing systems with many servers: null controllability in heavy traffic*

Rami Atar† Avi Mandelbaum‡ Gennady Shaikhet§

June 21, 2005

Abstract

A queueing model has $J \geq 2$ heterogeneous service stations, each consisting of many independent servers with identical capabilities. Customers of $I \geq 2$ classes can be served at these stations at different rates, that depend on both the class and the station. A system administrator dynamically controls scheduling and routing. We study this model in the Central Limit Theorem (or heavy traffic) regime proposed by Halfin and Whitt. We derive a diffusion model on $\mathbb{R}^I$ with a singular control term, that describes the scaling limit of the queueing model. The singular term may be used to constrain the diffusion to lie in certain subsets of $\mathbb{R}^I$ at all times $t > 0$. We say that the diffusion is *null-controllable* if it can be constrained to $\mathbb{X}_-$, the minimal closed subset of $\mathbb{R}^I$ containing all states of the prelimit queueing model for which all queues are empty. We give sufficient conditions for null controllability of the diffusion. Under these conditions we also show that an analogous, asymptotic result holds for the queueing model, by constructing control policies under which, for any given $0 < \varepsilon < T < \infty$, all queues in the system are kept empty on the time interval $[\varepsilon, T]$, with probability approaching one. This introduces a new, unusual heavy traffic 'behavior': On one hand the system is critically loaded, in the sense that an increase in any of the external arrival rates at the 'fluid level' results with an overloaded system. On the other hand, as far as queue lengths are concerned, the system behaves as if it is underloaded.

1 Introduction

We consider a multiclass queueing model with heterogeneous service stations, each consisting of many independent servers with identical capabilities. The servers offer service to different classes of customers at rates that may depend on both the station and the class. A system administrator dynamically controls all scheduling and routing in the system. The model is considered in the heavy

*Research supported in part by the Israel Science Foundation (grant no. 126/02) and the Technion funds for the promotion of research and sponsored research

†Department of Electrical Engineering, Technion - I.I.T., Haifa 32000 Israel

‡Department of Industrial Engineering and Management, Technion - I.I.T., Haifa 32000 Israel

§Department of Industrial Engineering and Management, Technion - I.I.T., Haifa 32000 Israel