This is a review submitted to Mathematical Reviews/MathSciNet.

Reviewer Name: Konstantopoulos, Takis

Mathematical Reviews/MathSciNet Reviewer Number: 68397

Address:

Department of Mathematics
Uppsala University
PO Box 480
SE-75106 Uppsala
SWEDEN
takis@math.uu.se

Author: Sigman, Karl; Whitt, Ward

Title: Heavy-traffic limits for nearly deterministic queues: stationary distributions.

MR Number: MR2836737

Primary classification:

Secondary classification(s):

Review text:

The paper under review is a sequel to the authors' previous one [Sigman and Whitt, *J. Appl. Probab.* **48**, 657-678, 2011] and deals with a queueing system in heavy traffic where all but every n-th customer are accepted for service. Specifically, consider first a sequence of queueing systems indexed by a positive integer n. Modify each queueing system by dropping all but every n-th arrival; also, divide the service time of each accepted customer by n. When n is large, by virtue of the strong law of large numbers, the modified queueing system is "nearly deterministic", with both inter-arrival and service times close to their mean. If, in addition, we assume that the traffic intensity (the ratio of the mean service by the mean inter-arrival time) ρ_n is strictly less than 1 (so that each queue is stable) but that $\rho_n \uparrow 1$, as $n \to \infty$, we have nearly deterministic queues in heavy traffic. The question addressed in this and the previous paper is in what sense do the stochastic processes describing the behavior of the system converge to a limit, as $n \to \infty$, under what scaling, and the properties of this limit.

Interestingly, the behavior depends on the rate of convergence of ρ_n. Two cases are distinguished. First, when $(1-\rho_n)\sqrt{n} \to \beta$, and second when $(1-\rho_n)n \to \beta$. Let $W^c_{n,k}$ be the waiting time experienced by the k-th customer in the modified system (where the superscript c is a reminder of the fact that the system is obtained by "cyclic thinning"). In the first case, the infinite vector $(\sqrt{n}W^c_{n,k}, k \in \mathbb{Z}_+)$ has a limit in distribution in the space (of probability measures on) $\mathbb{R}^{\mathbb{Z}_+}$ of infinite sequences of real numbers. In the second case, define the continuoustime stochastic process $W^c_n(t) := W^c_{n,[nt]}$, $t \geq 0$; then $(W^c_n(t), t \geq 0)$ has a limit in distribution in the space (of probability measures on) $D[0, \infty)$, the usual Skorokhod space. Both limits occur under natural assumptions.

In this paper, the behavior of the stationary distributions of waiting times (and other processes) is considered. In both cases, it turns out that the stationary distribution in the n-th system converges to the stationary distribution of the limit (with the appropriate scaling each time). In the first case, a formula for the mean stationary waiting time is obtained (as an infinite series). In the second case, the distribution of it shown to be exponential. The paper contains discussion of conclusions one can draw from these two limits. For example, in the first case the probability that a customer will find the system busy converges to a positive number strictly less than 1, whereas in the second case it converges to 1.

The paper also extends results of older papers, and contains, as special cases, several well-studied queueing systems.