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
takiskonst@gmail.com

Author: Atar, Rami; Biswas, Anup; Kaspi, Haya

Title: Fluid limits of $G/G/1+G$ queues under the nonpreemptive earliest-deadline-first discipline.

MR Number: MR3378178

Primary classification:

Secondary classification(s):

Review text:

This paper studies a single-server queue under a service discipline that makes its macroscopic behavior non-obvious. By macroscopic behavior we mean a functional law of large number for a process associated with the queueing system, such as the queue length or the departure process, under suitable scaling. The limits are usually called fluid limits. There is a single renewal stream of arriving customers. Each customer comes with a service time requirement and a deadline. Both service times and deadlines are i.i.d. random variables. Arrivals, services and deadlines are independent from one another. A customer arriving at time a having deadline u will abandon the system at time $a+u$ if he has not received service by this time. The lead time of the customer at time $t \geq a$ is $L(t) = a+u-t$. The queue is always arranged in such a way that the customer with the least positive lead time is at the head of the queue. If at some point of time the lead time of a customer in the queue becomes zero, the customer abandons the system at once. Upon a service completion, the server immediately selects the customer from the head of the queue and serves him. At this point, the lead time is not important, that is, if the customer starts service at time τ, he will depart the system at time $\tau+\sigma$ if his service time is σ regardless of his $L(\tau)$ (nonpreemptive discipline).

To obtain a fluid limit, one needs a sequence of such systems, parametrized by, say, an integer N in such a way that the arrival and service rates λ^N, μ^N, respectively, satisfy $\lambda^N/N \to \lambda$, $\mu^N/N \to \mu$, as $N \to \infty$, while the distribution function G^N of the deadline converges weakly to a distribution function G.

Initial conditions also scale with N, for example, the queue length $Q^N(0)$ at time 0 must satisfy $Q^N(0)/N$ converges to a constant $Q(0)$ as $N \to \infty$. These are the essential assumptions; more are needed for technical reasons. What is interesting here is that the system can be described by means of a Skorokhod reflection problem on a time-varying semi-infinite interval. This translates into a neat description of the limit. For example, if $\lambda \geq \mu$, the limiting queue-length process is the reflection of $Q(0) + (\lambda - \mu)t$ downwards at a function $h(t)$ that can be expressed explicitly in terms of the deadline distribution function (and the initial conditions); e.g., if the system is empty at time 0 then the limiting queue length is the reflection of $(\lambda - \mu)t$ downwards at $h(t) = \lambda \int_0^t (1 - G(s))ds$.

An essential ingredient of the analysis is keeping track of the lead times $L_i^N(t)$ of those customers i that have arrived before t, have not started service by time t and who will not abandon the system. This is done by creating a (random) point measure $\mathcal{Q}_t^N$ placing a unit mass at each such $L_i^N(t)$. The measure-valued random process $\{\mathcal{Q}_t^N, t \geq 0\}$ is shown to converge weakly to an explicitly defined deterministic measure-valued process $\{\mathcal{Q}_t, t \geq 0\}$, provided that $\lambda \geq \mu$. The fluid limits also describe a certain process called the frontier that is well-known to play a key role in systems operating under the above policy.